Vol. 94 issue 2, 2023

RIDP

Juliette Lelieur *(Ed.)*

Artificial Intelligence and Administration of Criminal Justice

(International Colloquium, Buenos Aires, Argentina, 28-31 March 2023)

Revue Internationale de Droit Pénal
International Review of Penal Law
Revista internacional de Derecho Penal
Международное обозрение уголовного права
刑事法律国际评论
المجلة الدولية للقانون الجنائي
Revista Internacional de Direito Penal
Rivista internazionale di diritto penale
Internationale Revue für Strafrecht

MAKLU

Artificial Intelligence and Administration of Criminal Justice

International Colloqium, Buenos Aires, Argentina

28th-31st March 2023

Artificial Intelligence and Administration of Criminal Justice

International Colloqium, Buenos Aires, Argentina

28th-31st March 2023

Edited by

Juliette Lelieur

RIDP
Revue Internationale de Droit Pénal
International Review of Penal Law
Revista internacional de Derecho Penal
Международное обозрение уголовного права
国际刑事法律评论
المجلة الدولية للقانون الجنائي
Revista Internacional de Direito Penal
Rivista internazionale di diritto penale
Internationale Revue für Strafrecht

Maklu

Antwerpen | Apeldoorn | Portland

AIDP – Association Internationale de Droit Pénal | The International Association of Penal Law is the oldest association of specialists in penal law in the world. Since 1924, it is dedicated to the scientific study of criminal law and covers: (1) criminal policy and codification of penal law, (2) comparative criminal law, (3) international criminal law (incl. specialization in international criminal justice) and (4) human rights in the administration of criminal justice. The Association's website provides further information (http://www.penal.org).

RIDP – Revue Internationale de Droit Pénal | The International Review of Penal Law is the primary publication medium and core scientific output of the Association. It seeks to contribute to the development of ideas, knowledge, and practices in the field of penal sciences. Combining international and comparative perspectives, the RIDP covers criminal law theory and philosophy, general principles of criminal law, special criminal law, criminal procedure, and international criminal law. The RIDP is published twice a year. Typically, issues are linked to the Association's core scientific activities, ie the AIDP conferences, Young Penalist conferences, world conferences or, every five years, the International Congress of Penal Law. Occasionally, issues will be dedicated to a single, topical scientific theme, validated by the Scientific Committee of the Association, comprising high-quality papers which have been either presented and discussed in small-scale expert colloquia or selected following an open call for papers. The RIDP is published in English only.

Peer review: All contributions are subject to double-layered peer review. The primary scientific and peer review responsibility for all issues lies with the designated Scientific Editor(s). The additional scientific quality control is carried out by the Executive Committee of the Editorial Board, which may turn to the Committee of Reviewers for supplementary peer review.

Disclaimer: The statements and opinions made in the RIDP contributions are solely those of the respective authors and not of the Association or MAKLU Publishers. Neither of them accepts legal responsibility or liability for any errors or omissions in the contributions nor makes any representation, express or implied, with respect to the accuracy of the material.

ISSN – 0223-5404
ISBN 978-90-466-1224-8
D/2023/1997/40
NUR 824
BISAC LAW026000
Theme: LNF, LAR

Maklu Publishers
Somersstraat 13/15, 2018 Antwerpen, Belgium, info@maklu.be
Koninginnelaan 96, 7315 EB Apeldoorn, The Netherlands, info@maklu.nl
www.maklu.eu

USA & Canada
International Specialized Book Services
920 NE 58th Ave., Suite 300, Portland, OR 97213-3786, orders@isbs.com, www.isbs.com

Editorial Board

Summary

PREFACE

By Juliette Lelieur[1]

This issue of the RIDP presents the results of collective research begun in 2020. Based on a long questionnaire, this research addresses emerging questions regarding the use of artificial intelligence (AI) – mainly machine learning – in the criminal justice context: predictive policing, predictive justice, and AI-based evidence. It provides a comparative study of the laws and academic opinions from various European, American, and Asian countries.

In 2022, building on the findings of national reports written in 2021 as well as on additional scientific literature, the research group prepared for the International Colloquium to be held in Buenos Aires. There, from 28 to 31 March 2023, the national rapporteurs agreed on the 32 resolutions reproduced at the end of this issue.

Parts of the national reports on AI and the administration of criminal justice are included in this issue; the entire reports as well as additional national reports are published online in the e-RIDP (www.penal.org/de/2023-2). The final results of our research will be presented at the XXIst International Congress of Penal Law that will take place in Paris from 26 to 28 June 2024 (Section 3 of the Congress).

Warm thanks are due to the President and Vice-president of the *International Association of Penal Law (AIDP)*, Prof. John Vervaele and Prof. Katalin Ligeti, and the members of the scientific committee of the AIDP for launching and supporting this project. My gratitude goes as well to the Argentinian national group of the AIDP, particularly to Prof. Javier Augusto De Luca and Prof. Francisco Figueroa, for organizing the inspiring International Colloquium of Buenos Aires.

Moreover, I am deeply grateful to the authors of the national and special reports as well as the colleagues who participated in the International Colloquium. Without their enthusiasm and their dedication to the scientific activities of the AIDP, this research project could not have been completed successfully.

Finally, I would like to express my thanks to the dynamic team of the RIDP, Prof. Gert Vermeulen, Prof. Nina Peršak, and Ass. Prof. Beatriz García-Moreno, who made the publication of the findings of this long-term collective work possible. I also thank the mixed research unit DRES of the University of Strasbourg for its financial support and Ms. Catherine Zimmerlin for her invaluable contribution to the layout of this issue.

Strasbourg, 30 November 2023

[1] Professor of Criminal Law at University of Strasbourg, France (juliette.lelieur@unistra.fr)

GENERAL REPORT

GENERAL REPORT

By Juliette Lelieur, Kelly Blount, Sarah Cherqaoui and Eftychia Bampasika [*]

1 Introduction

The present general report is a cross-analysis of about twenty national reports[1] written in response to a questionnaire established at the end of 2020.[2] This questionnaire scrutinizes three domains of artificial intelligence (AI) used in criminal justice: Predictive policing, predictive justice, and evidence acquired through AI. It contains 116 questions, which concern the practices observable in the criminal justice systems of the represented countries as well as the national legal frameworks – either existing laws or ongoing legal projects. Additionally, national rapporteurs were invited to assess the rise of 'AI solutions' in their criminal justice systems in light of the well-established principles of criminal procedural law and human rights. National reports were mostly written in 2021 (or 2022), and some of them were presented at the International Colloquium of Buenos Aires (28th – 31st March 2023). Their cross-analysis is reflected in this general report.

According to the combined reports, most countries only began to use AI in the context of criminal justice in 2021-2022. Only some had already been utilizing AI systems over several previous years.[3] A cumulative assessment indicates that many national rapporteurs faced a lack of information about concrete practices in their countries.[4] It was frequently reported that at times law enforcement authorities declined to provide them with precise information concerning the forms and methods by which they carry out their tasks; but for affirming that they use state-of-the-art systems. Some authorities outright denied using any program that relies on AI.[5] Moreover, even in countries that already have full experience in using AI in the field of criminal justice, there is very little legislation on the use of AI systems by law enforcement authorities. This suggests that

---

[*] Juliette Lelieur is a professor of Criminal Law at University of Strasbourg; Kelly Blount is Dr. in Law of the University of Luxembourg; Sarah Cherquaoui is a PhD Student at University of Bordeaux; Eftychia Bamapsika is a PhD Student at University of Würzburg.

[1] The reports mostly are from European countries (Belgium, Finland, France, Italy, Germany, Greece, the Netherlands, Poland, Portugal, Spain, Turkey and the UK). The Americas are represented (Argentina, Brazil, Canada, Chile and the USA) and a few additional countries also participated: Australia, China and Russia. Some reports only deal with one part of the questionnaire (Australia for predictive policing and the UK for predictive justice) or two parts of it (Belgium and the USA for predictive policing and predictive justice).

[2] The questionnaire can be consulted in English, French and Spanish on: https//www.penal.org/en/information

[3] Canada, Germany, the Netherlands, the UK and the USA.

[4] Belgium, Chile, Finland, Poland, Spain, Turkey.

[5] Chile and Poland. According to the Chilean national police forces, none of the systems they used is based on AI, however, according to researchers the system used by Carabineros de Chile is a combined approach of expert systems and machine learning.

almost no comprehensive democratic debate has taken place on the use of this new technology in an area where human rights are strongly at stake. Assessments of the results provided by AI systems are rare – and not always conclusive. While the media frequently discusses and speculates upon AI performance, there is little transparency about its actual use in criminal justice and little realistic information as to its purported benefits. The failure of AI systems is regularly highlighted in the media, however, this mostly occurs in areas other than criminal justice. Nevertheless, well-documented instances of harm due to AI tools escape public attention. It may be surprising to the astute observer that the Dutch childcare benefit scandal, for example, received very little attention abroad. In this case, false and xenophobic allegations of fraud emanating from AI systems[6] put thousands of innocent families into dire poverty due to erroneous treatment of reimbursement claims. Though this eventually led to the resignation of the Government in early 2021, it is still an underreported example of potential AI harm in the social sphere.

Before going deeper into this analysis, it is necessary to define AI and outline its most relevant, core characteristics about its use in the field of criminal justice. A general definition of AI refers to it as a set of theories and techniques used to create machines (robots or software) capable of simulating human intelligence. According to the High-Level Expert Group of the European Commission,

> Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans[7] that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.[8]

The research on which this report is built references this as its working definition.

AI was founded as a scientific discipline in 1956.[9] It is situated at the crossroads of statistical and algorithmic mathematics, computer sciences, and cognitive sciences. It unfolds in many different techniques that can be classified into two main categories. The first, symbolic AI, is based on high-level symbolic (human-readable) representations of problems and rules of logic. Since the mid-1950s, it made it possible to develop so-called expert systems, which are knowledge-based systems. The AI most people refer to today is, however, another more powerful – and disconcerting – technique. Connectionist AI uses interconnected networks, such as artificial neural networks. The well-known 'machine learning' technique derives from connectionist AI. It can build correlations between data

[6] See the Amnesty international report "Xenophobic machines", 25 October 2021: <https://www.amnesty.org/en/documents/eur35/4686/2021/en/> accessed 30 November 2023.
[7] Humans design AI systems directly, but they may also use AI techniques to optimize their design.
[8] <https://www.aepd.es/sites/default/files/2019-12/ai-definition.pdf> accessed 30 November 2023.
[9] Dartmouth Summer Research Project on Artificial Intelligence.

to adapt without following explicit instructions. To this purpose, it uses so-called 'self-learning algorithms', which may be supervised by humans or not. Supervised learning is also known as 'human-in-the-loop' machine learning. Machine learning expanded in the mid-1990s because of the extreme rapidness of computers' calculations and it subsequently flourished with the big data boom. Deep learning is one of the approaches of machine learning that developed in the 2000s and started to offer stunning results in the 2010s – like in the field of autonomous driving. It uses deep neural networks (many layers of interconnected artificial neurons) to learn patterns from massive amounts of data. Many of its applications like facial recognition and voice recognition have ramifications for criminal justice.

It is important to notice that machine-learning calculations are neither completely explainable by humans nor entirely traceable. The counterpart of their expansive capabilities is that their functioning comprises a part of mystery that even AI experts are not able to solve (the 'black box'). This leads to difficulties in criminal courts when AI systems are used for evidence purposes and explains why case law is starting to emerge regarding its use and effects on fair trial principles.[10] Furthermore, the reliability and impartiality of systems based on machine learning is a practicably unsolvable question, since it largely depends on the quality of the data they are processing. The bigger the volume of data they use, the more powerful they are, but it is hardly possible to check whether all these data, open data found on the web for many of them, are complete, up-to-date, accurate, and truthful. In addition, the lawfulness of these data is a serious problem as they may infringe on the right to privacy or data protection laws. We decided nevertheless not to include this topic into this research since it is not directly applicable to criminal justice concerns.

Following the structure of the questionnaire, this general report will first discuss 1) predictive policing such as the AI techniques that have reshaped it, then turn to the two facets of 2) predictive justice – actuarial justice and quantitative legal analysis – and finally examine 3) the incidence of AI on evidence questions in criminal matters. Its objective is not to provide a systematic comparison between the different national practices and legal approaches, but rather to highlight the most interesting emerging phenomena, as well as to point at some remarkable uses – or abandonment of uses – of AI systems, and some legal developments standing out from national reports and deserving special attention.

2 Predictive policing[11]

As a preliminary remark, it is necessary to discuss the term 'predictive'. It comes from the Latin praedictio (prediction), which refers to speech that announces the future. The

[10] Report on the USA, <https://www.penal.org/de/2023-2> accessed 30 november 2023, A-08, p. 14-16. See also the report on predictive justice in the USA, in this volume, p. 232-237.
[11] This part of the general report has been written by Kelly Blount and Juliette Lelieur.

word was often used in connection with religious discourses, for instance when describing prophecies or oracles, like by Pythia of Delphi who served as an interpreter of Apollo's voice. Still, in the context of policing, no one imagines that a machine is telling the future. AI systems only calculate the probability that an event happens, for instance, that a crime is committed as meteorologists forecast or foretell the weather. It would therefore be more adequate to speak of previsions for policing purposes or forecasting, rather than of predictions (but the adjective 'predictive' seems to have no equivalent in these verbal roots: neither do the words 'previsive' nor 'forecastive' exist).

This comment does not only have linguistic relevance. As a scientific discipline, AI holds a 'scientific aura' and most users follow – more or less blindly – the calculations provided by AI systems. The phenomenon of 'automation bias', which designates the propensity of humans to favor suggestions from automated decision-making systems and to ignore contradictory information made without automation, even if it is correct, has been mentioned in some national reports.[12] It is important to recall that results provided by AI systems are statistical calculations; as they are only probabilities, they should not be used to directly infer human behavior. They simply belong to the information the human decision-maker has and may be taken into account by the decision-maker if she estimates it appropriate to do so.

2.1 Definition

All national rapporteurs notice the lack of a legal definition of the term 'predictive policing' in their country, except in the USA where some local ordinances include such a definition.[13] Several reports refer to a doctrinal definition,[14] sometimes provided by foreign authors.[15] According to a research team of the Australian Institute of Criminology, predictive policing is

> the use of dynamic prediction models that apply spatio-temporal algorithms to core business data supplemented by secondary data sources, including internal corporate data and external environmental and socio-economic data, with the purpose of fore-

[12] Belgium, Canada, Greece.

[13] While there is no legal definition of predictive policing in federal or state legislation in the USA, a number of local ordinances provide such a definition. As an example, the City of Pittsburgh, Pennsylvania, defines predictive policing technology as 'Any fully or partially-automated computational application of programs, devices, hardware, or software based on machine learning or artificial intelligence that is, independent of a user, used to predict information or trends of crime or criminality that has or has yet to occur, including, but not limited to, the characteristics or profile of any individual(s) likely to commit a crime, the identity of any individuals likely to commit a crime, the locations or frequency of crime, or the individuals affected by crime or criminality'.

[14] Canada, Germany, the Netherlands, Spain, Turkey.

[15] Belgium, Finland.

> casting areas and times of increased crime risk, which could be targeted by law enforcement agencies with associated prevention strategies designated to mitigate to risks.[16]

This definition reflects the main understanding and encompasses the most common AI systems used in the world for predictive policing purposes.

However, in several reports, the doctrinal definition referred to is larger. It does not only focus on crime prevention but also includes solving past crimes.[17] Moreover, as will appear in the cross-analysis, several rapporteurs include the use of AI for wide surveillance purposes in their survey. On the one side, the surveillance programs concern digital transactions and are used to detect fraud, among others in customs, money laundering, and financial crimes.[18] It is clear that this kind of surveillance does not only serve prevention purposes: while detecting suspicious financial flows it builds a bridge between prevention and investigation. On the other side, surveillance through video cameras equipped with AI systems applies to the public space in different countries and it sometimes includes the identification of persons. Several reports – but neither the China nor the USA report – discuss intelligent video surveillance and facial recognition in public areas – the so-called biometric city surveillance.[19]

The variation in reporting first illustrates that the concept of predictive policing should be construed as a broad scope of policing measures rather than the lack of a universal definition of predictive policing. Second, it shows that the need for the scientific literature to work beyond the traditional definition of predictive policing is not isolated. It implicitly asks the question of whether non-police authorities – for instance, the cities using biometric city surveillance as administrative entities – include policing in their common activities. It also recalls that surveillance is a key element in crime-reducing strategies. More precisely, surveillance provides for a continuum between different aspects of crime control. In this context, AI systems not only indicate to the police where and when they should surveil, thus improving the quality of surveillance. Instead, they additionally provide surveillance means that apply with a very wide scope, even when no prior suspicion was detected. AI also has a huge impact on the quantity of surveillance by enabling mass surveillance. This general report, therefore, needs to consider the role of AI systems in the different surveillance facets, though it exceeds the traditional boundaries of the above-mentioned definition of predictive policing.

---

[16] Daniel Birks, Michael Townsley and Timothy Hart, Predictive policing in an Australian context: Assessing viability and utility, Trends & issues in crime and criminal justice, Australian Institute of Criminology, no. 666, 2023, p. 2.

[17] Finland, Spain, Turkey, the USA.

[18] Germany, Italy, the Netherlands, Poland.

[19] Argentina, Brazil, France, Germany, Russia.

2.2 Uses and perceptions of AI in predictive policing

2.2.1 National practices of using AI for predictive policing and surveillance purposes

Eleven of eighteen countries are reported as using some form of AI predictive policing software.[20] However, this alone is fairly inconclusive as a benchmark, as predictive policing represents a broad category of policing techniques and many programs included in the reports are highly specific in purpose and scope. The most unified example of predictive technology is that used to anticipate crime according to geographical location, often referred to as crime mapping or hot spot analysis. Several national reports mention the use of such programs in specific areas – mostly cities – and for targeting different categories of crimes, mostly property crimes such as burglaries and car-related crimes, less often violent crimes – including gun violence.[21] Besides the police, also other law enforcement agencies consider developing AI tools to assist them in their duties. In Finland, the Border Guard's 2020 annual report mentions a project that ims to use AI systems to better surveil the land borders and sea areas.[22] Moreover, some country reports reference highly specific programs that are based on geospatial policing but include other criteria. In the Netherlands, for instance, two programs precisely target young offenders.[23] In Canada, the Edmonton Police Service has participated in innovative projects aiming to identify links between criminality and the consumption of drugs and alcohol – for instance thefts in liquor stores.

Besides geospatial policing, AI systems serve person-based policing, like in the context of combatting terrorism. In Germany, a risk assessment tool serves to evaluate whether individuals who have already been identified by the police authorities as potentially dangerous are likely to commit Islamic terrorist attacks.[24] In Italy, the Ministry of Defense and Carabinieri adhere to a program financed by the European Commission to identify terrorism-related web content.[25] Furthermore, in the European Union, Directive 2016/681[26] on the use of passenger name record (PNR) data for the prevention, detection, investigation and prosecution of terrorist offences and serious crime, obliges air carriers to transfer PNR data to the Passenger Information Unit of each Member State. Several

[20] The countries reported as using predictive policing in some form are Argentina, Canada, Chile, China, Finland, Germany, Italy, the Netherlands, Russia, Spain, and the USA.
[21] Brazil, Canada, China, Germany, Italy, Netherlands, and the USA.
[22] Report on Finnland, <https://www.penal.org/de/2023-2>, A-15, p. 5.
[23] The Amsterdam municipality uses the 'Top600' program that calculates the risk of committing a crime for young individuals under the age of 16, while the national police works with the actuarial semi-automated risk assessment instrument 'ProKid 12-SI System', which concerns youths ages 12-18 years.
[24] RADAR-iTE was developed by the Federal Criminal Police Office in cooperation with a research group on forensic psychology.
[25] DANTE (Detecting and ANalysing TErrorist-related Online Contents and Financing Activities).
[26] Directive (EU) 2016/681 of the European Parliament and of the Council of 27 April 2016.

national reports righty classify analyzing PNR data to identify people who were not previously or otherwise suspect as a form of predictive policing.[27]

Another use of AI against serious crime is the tracking of people implicated in child exploitation, via means such as analyzing financial data that are associated with child sexual abuse material, as mentioned in the Australian report. Moreover, Australia, Canada, and Spain are reporting programs that are designed to flag high-risk family and domestic violence offenders. It is of note that not only potential offenders, but also potential victims are the target of these analyses.[28]

A few very specific programs finally deserve attention. In Spain, the 'Veripol' system, which seems to be unique in the world, estimates the probability that reports of robbery – possibly with violence or intimidation – are fake. The goal here is to help the police sort out false complaints and dissuade abusive whistle-blowers. In Australia, a machine-learning program was utilized to forecast serious police misconduct.

In some national reports, there is a lack of predictive policing as a common, formal practice, though certain, targeted programs are used. The apparent reason for the seeming lack of predictive policing stems from the classification of policing practices and mandates. In other words, how the specific policing function is defined and regulated depends also on its classification within policing, criminal investigation, or prosecutorial competencies. In Poland, it is reported that the overlap between prevention and crime control may mean that operationally predictive policing is indeed a function of crime control, but technically outside the auspices of prevention and policing, and therefore what may be considered predictive policing elsewhere is not categorized as such in Poland. Similarly, the Belgian report cites a distinction in the use of police 'mapping' prior to a crime versus post-crime, when the development of suspicion is traditionally initiated. Therefore, in addition to varying definitions of what constitutes predictive policing, its place in legal classification schemes further complicates the comparison of applicable legal regimes.

Turning to general surveillance through AI, many of the programs referenced in the reports are used to collect and sort biometric or vehicle plate information, to be stored and used in conjunction with other data.[29] Surveillance measures solely directed at persons are also referred to under the term predictive policing. This is particularly the case in Argentina and Russia where video surveillance enhanced by facial recognition is applied

[27] Finland, France, Germany.

[28] In Canada, for instance, the Saskatchewan police predictive analytics lab uses programs to identify children and young people at risk to be kidnapped. In Spain, the VioGen program aims the prevention against gender violence. In Australia, it is said that some technologies are suitable to examine the relationship between abuse types and victim injuries as well as the risk of escalation for victims of domestic violence.

[29] Various software are reported to be used in conjunction but for distinct purposes, as in Greece where both biometric and vehicle registration information is collected by the Hellenic Police. See also Brazil, Canada, Germany and the Netherlands.

in the cities of Buenos Aires[30] and Moscow.[31] Also in France and Germany, experimentation with biometric surveillance in public areas is reported,[32] though reportedly oriented toward so-called intelligent video surveillance, designated to detect dangerous situations and behaviors. For instance, the French Government recently decided to use it during the Olympic Games of Paris in 2024. Additionally, it is reported as at times used by private actors to perform biometric surveillance of persons. The Spanish report provides the example of the Mercadona company, using facial recognition to identify already condemned thieves who were prohibited from entering its supermarkets until it was condemned by the Spanish data protection authority for illegal processing of sensitive data.

Law enforcement authorities in an increasing number of countries appear interested in using AI to detect fraud and other economic crimes based on the surveillance and analysis of financial transactions. Several reports of European countries mention the use of AI systems in the context of tackling fraud in customs, tax, or social matters, and it is clear that the detection of money laundering by private companies as well as by Financial Investigation Units is much easier and quicker with the help of AI.[33] The same is true for the detection of illegal content circulating on the web. The Canadian report notes the use of AI systems by the police for the surveillance of chats on social media.

2.2.2 National practices of not using AI (anymore) for predictive policing

Those countries reported as not known to be officially using predictive policing, namely France, Greece, and Portugal, cite a variety of reasons. In France, it is most notably issues of data protection and fundamental rights that are seen as obstacles to the development of crime mapping programs. Additionally, experimentation of geospatial systems that took place in France did not convince the French government to engage more deeply in predictive policing technologies. By contrast, French investigators use software to produce crime analysis diagrams to solve past crimes.[34] Greece and Portugal are reported as exploring the development of predictive policing as a part of wider AI research and development. The Portuguese rapporteur explicitly mentions that cooperation with the European Union institutions will be decisive for further commitment to AI-based projects

[30] The AI facial recognition surveillance system of the city of Buenos Aires is part of a comprehensive video surveillance system. It is used to identify the faces of wanted persons like defendants, convicted defaulters and fugitives. During the covid-19 pandemic, an infrared temperature detection system was added to the facial recognition surveillance cameras.

[31] The Safe City program of Moscow utilizes a complex network of computer systems and 178,000 cameras across the city, linked with FindFace Security that allows for facial recognition scanning. This was well used during the COVID lock-downs, though its official purpose is to identify known offenders and missing people, as well ensure the security of public places.

[32] In France, experimentation took place during the Nice Carnival and, in Germany, it was used in big train stations as part of a broader experimentation also including intelligent video surveillance.

[33] Germany, Italy, the Netherlands, Poland. The Finnish report mentions projects in tackling money laundering and financing of terrorism, grey economy and economic crimes.

[34] Anacrim and SALVAC, which are working based on machine reasoning (rather than machine learning).

and the Italian report relates that predictive policing software was co-funded by the European Union.

It is remarkable that in several countries the use of geospatial systems was abandoned after a few years. In the USA, where the tool PredPol (now Geolitica) was first widely used,[35] several police departments or city councils decided to either ban or suspend the use of predictive policing technology. This was the case in Santa Cruz (California), Pittsburgh, New Orleans, and Oakland (California) in the years 2020 and 2021, and even the Los Angeles Police Department, which developed PredPol, announced in April 2020 that it would stop using it. The reasons given for these decisions were that AI systems led to the over-policing of neighborhoods most heavily populated by people of color and the poor, and also that the benefit of AI-based predictive systems was low: the system did not offer much more information than what police authorities already knew. In Germany, the State Baden-Württemberg stopped further authorization of PreCobs in 2019, and Bavaria ended its use by police in 2021; both claimed that not enough data are available to use the program. Additionally, in Low Saxony PreMAP is no longer used based on a cost/benefit analysis. In Spain, local police entities had considered the possibility of using the EuroCop Pred'Crime software. They finally did not fulfill the project but it is not known whether it was simply not implemented or whether it was abandoned after implementation due to its impact on fundamental rights or the lack of sufficient regulation.

Turning to AI systems that are not dedicated to crime mapping, the Canadian report describes two specific systems that were abandoned due to concerns over violations of fundamental rights. First, the Ottawa Police Service stopped using the facial recognition system NeoFace Reveal after tests showed that data protection was not being honored, as explicit consent was not given by subjects. Second, the Toronto Police Service stopped using an automatic gunshot detection system (ShotSpotter) as it was considered to potentially violate the right to privacy. In Belgium, the cessation of any trial predictive policing seems to be based on legal issues, such as lack of bases or by nature of being a pilot project. Finally, in the Netherlands, the AI system used to detect fraud in the child allowance program was discontinued after the 2020 childcare benefits scandal.

2.2.3 Incentives, assessment, and perception

Incentives for using AI predictive policing systems are similar or nearly identical in all reported countries: On the one side, preventing crime and avoiding victims; on the other side, reducing costs in the context of scarce human resources in the police, through a more strategic allocation of police resources. For instance, high probabilities of crime in certain areas allow restructuring patrol routes, which facilitates more efficiency in crime control. By contrast, the aim of better understanding the causes of crime is rarely mentioned and there are no concrete indications that it is a stated goal. The Chilean report interestingly notes that the incentive for the use of AI in policing seems to be political in

[35] In 2018, more than 60 police departments around the country used PredPol.

nature, mainly based on promises about public safety. Though not explicitly stated, there is no reason to assume it is not also true in other countries.

Although the questionnaire explicitly asked about the potential existence of assessments on the reliability, impartiality, and effectiveness of AI systems in predictive policing, many national reports did not address the theme.[36] Some of them explicitly express that the reason for any lack of transparency is a lack of public information.[37] In countries where assessments took place, the findings concerning crime-mapping systems differ. In the USA, where such tools are or were relatively most used, evidence regarding their accuracy, reliability, and overall utility is, at best, mixed. First, it should be noticed that it is primarily their vendors who, while prohibiting independent, third-party review of their systems, provide information concerning their reliability.[38] An independent study published in 2021 evaluated PredPol predictions in 38 cities and countries and found that its algorithm 'disproportionately targeted vulnerable populations, including low-income communities and residents of public housing' as well as 'neighborhoods with proportionately more Black and Latino residents'. Finally, concerning the effectiveness of AI crime-mapping tools, the few conducted evaluations did not show conclusive positive results. It is not surprising, in this context, that many cities or police departments terminated their contract with the companies that developed the systems. Similar findings appear from the Dutch report: Though the Dutch police claim that the use of AI systems causes a decrease in crime, studies show no correlation. In Italy, though, some AI software was evaluated both by their users and third parties, and the results are positive. Delia, for instance, is reported as self-finding that it has produced an 89% reduction in retail robberies in Milan from 2008-2017. A single academic study of the program found that robberies in the same sector are about 8% more likely to be solved. Other Italian crime mapping tools are reported as having very high accuracy results; however, most evaluations are not independently conducted.

Concerning non-geospatial tools, evaluations seem to be positive too. In Spain, the reliability of Veripol, tracking fake robbery reports, was carried out by the police and reached 90% accuracy. Moreover, the VioGen system is used against gender violence and externally evaluated by the Ministry of the Interior as well as a non-profit organization, 'Eticas Foundation' autonomously.' These findings indicate that the use of VioGen has coincided with a decrease in recidivism by 25% over a decade.

Finally, concerning facial recognition, a 2019 study about the biometric surveillance system of the City of Buenos Aires, done by the authority that put the system into operation, reported an accuracy rate of more than 93%. However, after errors later appeared in the identification of individuals with facial similarities, the publishing of accuracy results ceased to be published.

---

[36] Belgium, Chile, China, Russia, Turkey. In countries where predictive policing tools are declared as not used, assessment logically cannot take place (Finland, France, Greece, Portugal).

[37] Canada, Chile, Poland.

[38] Report on the USA, <https://www.penal.org/de/2023-2>, A-08, p. 13.

Turning to the perception of AI systems by the population, in most countries there is no significant public debate for now or public discussion is just beginning.[39] Not surprisingly, where assessments are positive like in Italy and Spain, a favorable public opinion is perceivable. However, even in those countries, as well as in the USA where public opinion seems to have changed over the years, attention to the dangers of bias and discrimination has increased. The same is true in the Netherlands, where the public discussion is based on a risk-averse approach and AI in all areas of society is considered to be a way to avoid risk. Despite this, a more critical debate has emerged since the childcare benefit scandal. In Germany, the reception of 'predictive policing' in the media and general public is very diverse and includes awareness around the excessive use of personal data, blind trust in technology, direct and indirect discriminatory effects, and the 'chilling effects' of automated policing. Also in France, where the media largely discuss the abuses of mass surveillance in foreign countries, especially China, critical voices on facial recognition in public areas have increased.

It has principally been for NGOs and legal scholars to raise the strongest concerns and challenges to predictive policing programs.[40] In 2019, some 400 academics in the USA discredited the PredPol program in an open letter sent to the Los Angeles Police Commission. The Turkish report shows as well that much concern exists toward the risks of AI predictive policing systems.[41] In general, the main concerns are discrimination and violations of individual privacy. Furthermore, a lack of transparency and the difficulty of verifying the accuracy of systems due to corporate secrecy – including by foreign firms – are seen as serious difficulties. In several countries, the need for stricter regulation appears more or less explicitly.[42]

A last but important remark concerns the acceptance of AI systems by the police itself. Whereas in Italy a positive perception seems to exist among police officers, in the Australian and Belgian country reports, the authors point to the problem of over-estimated promises of AI, and consequently to the risk of discounting the value of individual and collective human knowledge in police forces. The German report interestingly indicates that there is a better acceptance of AI predictive policing systems in high ranks of the police hierarchy than among patrol officers because patrol officers can better justify their actions based on statistical evaluation, making requests to secure additional resources more successful.

[39] Argentina, Belgium, Chile, Finland, Italy, Poland, Portugal, Turkey. The Chinese and Greek reports did not address the question.

[40] Australia, France, Belgium, Canada, Germany, Italy, the Netherlands, Russia, Spain, Turkey and the USA.

[41] One study from the University of Szeged, University of Konstanz, and Istanbul University found that the use of data by predictive technologies risks increasing or creating biases for marginalized groups. The Istanbul Bar Association Informatics Law Commission finds that one of the main problems with predictive policing algorithms is a lack of transparency (2020).

[42] Canada, Germany, Russia, Spain.

2.3 Normative Framework: law & policy

2.3.1 Legislation and soft law

No reporting country has a legal framework specific to predictive policing. Instead, most cite their fundamental rights and data protection regimes as well as constitutional rights frameworks as generally applicable.

Soft law is therefore making its way, sometimes even at the national level, although predictive policing is still not the direct subject. The Portuguese *Charter on Human Rights in the Digital Age* adopted in 2021[43] encompasses a provision, Article 9, which is dedicated to rights as regards AI. It calls for the application of fundamental principles of the Portuguese legal order, as well as for precautions that are adapted to AI technology. Similarly, in Spain, the *Digital Rights Charter* was adopted in 2021 to strengthen existing legal frameworks in tandem with the development of AI systems to ensure a human-centric approach (Article XXIII of the Charter).[44]

Other countries have proposed AI policy frameworks that will apply to predictive policing as part of a larger legislative package that focuses on nationwide assessments of fairness, accountability, transparency, and efficiency. Even here, predictive policing is not the direct subject of policy but may be categorized within public administration or criminal justice. In the Netherlands, the report notes the 'Guidelines for the use of algorithms by public authorities' for the development by authorities, as well as to inform the public about the use. The report carefully indicates that these guidelines are for development and explicitly not meant to provide legal guarantees.[45] The Portugal report refers to the implementation of the Horizon 2020 projects and indicates a recognition of the utility of AI-based systems, but predictive policing per se has not been proposed, according to the report.

Some countries highlight specific challenges to prospective measures, as in Russia where AI-based predictive policing legislation is being considered but issues such as tension between judicial and police uses of AI, as well as a perception that it will cause the loss of police jobs, provide hurdles to concrete law. Most of the countries, however, report existing, non-binding frameworks on public authorities' uses of AI moving forward, but still, legislation seems to largely be based on concerns of privacy and data protection, namely the processing of data.

An exception in the USA, the tentative Draft N° 3 of the American Law Institute (ALI) 'Principles of the Law, Policing,' agreed upon in 2021, provides guidance and suggests a

[43] Law no. 27/2021, 17th may 2021. See in this volume p. 338.

[44] Government of Spain, 'The Government adopts the Digital Rights Charter to articulate a reference framework to guarantee citizens' rights in the new digital age,' (14 July 2021) <https://www.lamoncloa.gob.es/lang/en/gobierno/news/Paginas/2021/20210713_rights-charter.aspx> accessed 30 November 2023.

[45] <https://www.dataguidance.com/news/netherlands-council-state-publishes-guideline-use> accessed 30 November2023.

comprehensive set of best practices to courts, legislatures, and police.[46] These include that an agency 'should not rely on an algorithm or profile to direct police resources to a particular location, to identify potential targets for further investigation or surveillance, or to assess the risk of harm that individuals may pose to others' without meeting requirements set forth therein. In addition, predictive policing is expressly mentioned as one of the automated systems that should be covered by the 'Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People', a white paper published by the White House Office of Science and Technology Policy in October 2022. The lack of explanation and transparency regarding the placing by a predictive policing system of individuals on a watch list is used as an example of a problem that the principle of notice and explanation was designed to address and protect against.[47]

Many of the countries with some form of legislative framework, namely in the EU, also cite international legal regulation, such as the General Data Protection Regulation (GDPR).[48] Other countries, such as Russia, reported legal instruments include guidance from the United Nations Office on Drugs and the Interpol Innovation Center.[49] In a few cases, transnational agreements dictate the trajectory of predictive policing development. For instance, the Franco-Canadian Declaration on AI aims to develop national compliance standards according to the OECD and International Group of Experts in AI (G21A). Nearly every country reported that it follows or accepts the guidance of international laws or regulations, with the exception of the USA, whose federal criminal justice system does not refer to international or regional laws applicable to predictive policing.

2.3.2 Case law

As regards case law, there have been very few determinative cases on the use of predictive policing anywhere, mostly due to the lack of its formal use or discrepancies in its categorization or the legal framework applied. However, in one criminal case in the US, the federal court of appeal held that the finding of a weapon in the course of a suspicionless search was 'unreasonable' and evidence based on it had to be suppressed. Concur-

[46] American Law Institute, Principles of the Law, Policing § 3-2.06 (Tent. Draft No. 3, 2021).

[47] The White House, Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People, 2 (October 2022) <www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf> accessed 5 November 2023. The Blueprint identifies five principles that should be used to guide the design, use, and deployment of automated systems to order to protect the public in the age of artificial intelligence: safe and effective systems; algorithmic discrimination protections; data privacy; notice and explanation; and human alternatives, consideration, and fallback.

[48] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.

[49] Interpol: <https://www.interpol.int/How-we-work/Innovation/Artificial-Intelligence-Toolkit> accessed 30 November 2023; The Ethical Use of AI, UNODC, <https://www.unodc.org/ji/en/knowledge-products/artificial-intelligence.html> accessed 23 October 2023.

ring judges argued that the use of predictive policing methods could lessen constitutional protections of people who live in high-crime areas and could contribute to the perpetuation of racial bias and profiling in the criminal justice system.[50]

In addition, a few non-criminal cases that deal with AI systems used for predictive policing have been litigated. The first theme is transparency. A case against the New York City Police Department for failing to comply with a Freedom of Information Act request as to its use of predictive policing software was decided in 2017. The trial court held that a non-disclosure agreement with a vendor could not, without more, withstand the request for public information.[51] In the Netherlands, an important decision related to the use of AI for securing public benefits or identifying the fraud thereof was issued by the District Court of The Hague, which found that the use of the program was not transparent and therefore not verifiable and so unlawful.[52] Finally, in Argentina, there was a request for access to information related to the 2019 Order that allowed the use of the facial recognition system in Buenos Aires and a request for access to public information. At trial, however, the Court held that the following inquiries were not, and must be answered: '1. Security and reliability protocols for facial image capture, 2. Data erasure audit, 3. Identification of individuals not included in the databases of Co.Na.R.C. and the National Registry of Recidivism, 4. Determination of the percentage of false positives, and 5. Appointment of police force agents who are provided with confidential information.'[53]

In Spain, the Military Chamber of the Supreme Court convicted a Civil Guard because he refused to use VioGen although this is mandatory for all State security forces and bodies.[54] In another case concerning VioGen, the Spanish State engaged in civil liability after a woman died because of gender violence. The police officer who was in charge of using VioGen failed to correct the automated evaluation, which did not include a criminal record completed outside Spain. As the tool did not consider the content of this report, it indicated a "not appreciated" level of risk, while the risk was very high.[55]

Some cases concern the right to data protection, such as in France, where few cases have been raised, but challenges to predictive policing come from administrative avenues, such as through the authority of the CNIL, which adjudges the use of personal data.[56] In

[50] United States v. Curry, 965 F.3d 313 (4th Cir. 2020)(en banc). See the report on the USA, <https://www.penal.org/de/2023-2>, A-08, p. 19-24.
[51] Brennan Ctr. for Justice at N.Y.U. Sch. of Law v. New York City Police Dep't, 2017 N.Y. Misc. LEXIS 5138 (N.Y. Sup. Ct. Dec. 27, 2017).
[52] Judgment of 5 February 2020, ECLI:NL:RBDHA:2020:1878 (case nr. C-09-550982-HA ZA 18-388).
[53] Case file: 9480/19-0, 'Observatorio de Derecho Informático Argentino O.D.I.A. c/GCBA s/Acceso a la Información' [Argentine Computer Law Observatory (O.D.I.A.) v. Government of the City of Buenos Aires on Access to Information], judicial decision rendered on 20 May 2020.
[54] STS 73/2020, of October 28, (Fifth Chamber, Military), Rec. 26/2020.
[55] Spanish National Court, (Audiencia Nacional) specifically the Contentious-Administrative Chamber, in the Judgment of September 30, 2020
[56] For example, CNIL 12 janvier 2021, délib. N°SAN-2021-003 sur l'utilisation des drones, notamment lors du confinement et la mise en œuvre des mesures dans le cadre de la crise sanitaire.

Germany, the Federal Constitutional Court ruled on the right of personality in its manifestation as a right to informational self-determination first in 2016, 2018, and again in 2023. In its February 2023 decision, it found that the legislation of two States, Hesse and Hamburg, regarding automated data analysis for the prevention of criminal acts, was unconstitutional because they did not provide a sufficient threshold for interference.[57] In contrast, a court case was brought in Russia, which challenged the use of facial recognition and biometric data as a violation of privacy; however, the court held that its use in public places and for the aims of security and public safety legitimated its use.[58] National data protection agencies play an important role too. In Spain and Italy, private actors have been severely sanctioned for using sensitive personal data illegally in the context of facial recognition. In July 2021, the Spanish data protection agency fined Mercadona S.A. 2,5 million Euros for the illegal biometric surveillance of its supermarkets, and in March 2022, Clearview AI faced a 20 million Euro fine from the Italian data protection agency over the unlawful processing of biometrics and geolocation data.

2.3.3 Substantive guarantees

Substantive guarantees applicable to the use of AI may be neatly divided into several categories. First, labelling and certification requirements in reporting EU countries are nascent and the reporting consensus seems to default to the terms of the AI Act in expectation of its advancement into law, as well as the requirements set forth in data protection and privacy legislation. Outliers include Russia, where certification procedures are mandated by the state but beyond software that deals in state secrets, there is no obligation that regular monitoring or auditing continually occurs.

Second, concerning the need for regular assessment of the accuracy and/or effectiveness of AI-based systems, it seems that the legal system of no reporting country poses any requirement of that matter. There is an exception, in the USA, where several municipal ordinances require periodic audits. It appears that a number of agencies that conducted such audits in recent years concluded that various algorithm-based programs were less effective than expected. Moreover, in Spain, a non-profit organization publishes statistics on the use of the VioGen program as well as a user manual, so the public may 'know the tool'. However, this is not due to legal requirements.

Third, in countries where data protection authorities are the overseers of potential predictive policing, regular auditing and reporting requirements apply; however, most countries have no standards in place specific to predictive policing. The German report discusses data protection principles in detail. It includes that the public consultation of the Federal Commissioner for Data Protection and Freedom of Information led to calls for 1. the need for broad public debate and empiric review of AI applications; 2. the requirement for a specific legal basis; 3. use of AI must comply with general rules on data

[57] See the German report, in this volume, p. 144-148.

[58] Per the Russia report <https://www.vedomosti.ru/politics/articles/2019/10/06/812955-moskvichka-prosit-sud> accessed 6 November 2023.

protection; 4. explainability, quality of data, and quality of training data; 5. preservation of the "core area of private life"; 6. data protection authorities must supervise the use of AI; and 7. there must always be a privacy-impact assessment before use.

Fourth, as regards accountability of organizations developing AI, there is very little reported as applying directly to predictive policing software or programs. Finally, almost every country reported that AI use by police will require that the police officer or entity using the technology be held accountable according to the relevant national regulations on policing or public authorities rather than the developer. Similarly, as in Belgium, several countries indicate the division between AI decision-making and police action, necessarily severing the chain of accountability with the developer.

2.4 General principles of the rights framework

The majority of reporting countries describe similar concerns over general principles of law as regards the use of AI, which would inherently apply to predictive policing as well. Where there is an applicable rights framework to speak of, the reports cite an available connection to predictive policing, rather than any existing instruments drafted for this purpose. This is the case in the majority of reporting countries, such as the USA where constitutional principles are cited as most applicable; as well as Spain and Portugal which report evolving approaches to digital rights that are based on the development of existing rights, rather than the need for a dedicated protection framework specified to AI. A number of the EU reporting countries, such as Germany, indicate that existing data protection laws are already applicable to the use of predictive policing, yet point out the forthcomoing AI Act and the need to transpose it into the repective national legal frameworks.

2.4.1 Consensus about existing threats to the rights to equality and privacy

The rights of non-discrimination and equality are raised in all country reports within the context of policing regulations, constitutional law, or policy on AI writ-large. Especially highlighted is the risk that predictive policing leads to over-policing certain categories of people, and to perpetuating racial bias, is almost systematically highlighted.

Similarly, the privacy issue is frequently discussed in EU countries within the context of GDPR, the Law Enforcement Directive,[59] or national constitutions; however, the Dutch report cites directly to Article 8 ECHR and the Greek one refers to the caselaw of the EU Court of Justice. Though no laws directly defined in terms of predictive policing are reported, the most frequently cited, common approach to an applicable rights framework appears to be within privacy and data protection regimes. One outlier is China, the report

[59] Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data.

for which states that privacy is generally within the realm of civil law and that enforcement is available via tort liability law rather than a constitutional principle.

2.4.2 Awareness of risks concerning the rights to liberty and security of persons

Concerns about the risk that a person be arrested based on mere probabilities calculated through AI are mentioned in numerous country reports. American legal scholars seem to agree that predictive policing in connection with big data (machine learning) 'has the potential to change the reasonable suspicion calculus because more personal or predictive information about a suspect will make it easier for police to justify stopping a suspect.'[60] In Greece, however, in terms of surveillance tools, there must be a reasonable suspicion that certain crimes are likely to occur in a certain place, based upon factual evidence or statistical data before an arrest can take place. The Turkey report states accordingly that it is unlikely that profiling technologies, like predictive policing, would meet the requirements for arresting someone, as evidence must be produced regarding a crime, which in the case of prediction is absent. The requirement for suspicion is specific and objective facts. One interesting outlier is China, the report of which states that, however, a proposal has been put forth for a remedy for large groups of individuals who are subject to algorithmic decision-making (monitoring, etc.). The report proposes that a class action system may be used to request decision-making processes and rationale, which if not provided may open up the possibility to file a class action suit.

The Spanish report mentions concerns over the freedom of movement, stating that it has been discussed whether the use of facial recognition affects freedom of movement. It is considered that even if facial images come from open-source materials, they are collected without consent. In France, the right to privacy is frequently discussed in the context of liberty, though not for predictive policing specifically. Any data used for technology used for policing purposes will be governed according to the type of data, the duration of its retention, and details of processing. These will also apply to predictive policing.[61]

2.4.3 Discussions about risks against the procedural rights in criminal proceedings

There is some variation in the apparent procedural requirements applicable to predictive policing. The Russian and Canadian reports indicate that there are little or inadequate applicable principles of procedural legality governing predictive policing. Similarly, in the Netherlands, it is reported that predictive policing with AI-based systems would not require reasonable suspicion as this aligns with a preventative investigation generally.

---

[60] Andrew Guthrie Ferguson, 'Big Data and Predictive Reasonable Suspicion,' University of Pennsylvania Law Review, (2015): <https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=9464&context=penn_law_review> accessed 30 November 2023.

[61] <https://documentation.insp.gouv.fr/insp/doc/SCOOPIT/254DC1AB72053BC20F7FCC83485E1F71/synthese-du-livre-blanc-surveiller-les-foules-nbsp-observatoire-de-l-ethique-publique?_lg=fr-FR>, accessed 30 November 2023.

As to Finland, it is reported that procedural legality over predictive policing would be based on the suites of fair trial and good administration rights.[62]

When asked whether it is possible to use outputs of predictive policing tools in criminal proceedings, reports seem to indicate that existing legal frameworks would again provide the standards necessary for such a determination. In Turkey, it is reported that if AI programs are in operation without a legal base or procedural guarantees, a prohibition of unlawfully obtained evidence would be implicated. The Constitutional Court puts forth a strict exclusion of illegal evidence and therefore to use preventive AI technologies would require applicable rules for their findings to be lawful evidence. Similarly, the report states that when algorithmic outputs are based solely on profiling data, they alone should not per se initiate a criminal investigation. As to Argentina, the investigation stage of a criminal proceeding cannot be initiated by the comprehensive electronic AI facial recognition surveillance; therefore, fair trial principles are not implicated as criminal proceedings will not rely on this form of evidence.[63] In contrast, the Brazil report indicates that the use of AI-generated evidence, or that secured by police at the impetus of predictive assessments, may be potentially applicable. The authors of the Finnish indicate that, although there is no categorical ban on evidence produced by predictive policing, the admissibility of such data depends on whether they may adequately prove facts relevant to the case at hand. They highlight that a prediction 'likely bears no relevance in proving that the accused is guilty of a specific past offense'.[64] Other countries, such as Italy, are reported to be very case- and situation-specific, as the assessment may be relevant to the extent that it refers to the most recent crime committed by the accused only for the purpose of crime linking. This is raised in the report as problematic, due to the fact there is no regulatory authority that oversees predictive policing, as such.[65]

Furthermore, some country reports mention the presumption of innocence and suggest that AI technologies endanger the principle, should investigative measures be taken based on predictions without enough suspicion.[66] In the Spanish report, it is noted that the Veripol software specifically may be problematic to both the presumption of innocence as well as the status of the victim.[67]

2.4.4 Concerns about threats to the freedom of expression

The report on China mentions the threat to the right to freedom of expression: 'Some scholars have pointed out that the use of large-scale monitoring in the investigation has

[62] Report on Finland, <https://www.penal.org/de/2023-2>, A-15, p. 10.
[63] Report on Argentina, <https://www.penal.org/de/2023-2>, A-17, p. 3.
[64] See the Finnish report in this volume, p. 306.
[65] G. Padua, *Intelligenza artificiale e giudizio penale: scenari, limiti e prospettive*, in *Processo penale e giustizia*, 2021, 1492; C. Parodi – V. Sellaroli, *Sistema penale e intelligenza artificiale*, cit., 58 ff.; M. Pisati, *Indagini preliminari*, cit., 958.
[66] Belgium, Canada, Spain, and Turkey. For further details see Kelly Blount, Applying the Presumption of Innocence to Policing with AI, in Artificial intelligence, big data and automated decision-making in criminal justice, RIDP, Vol. 92, 2021, p. 33.
[67] Spanish Constitution Art. 24.2.

a direct and indirect impact on freedom of expression. '[68] It goes on to state that' the filtering and interception of specific information by investigation organs will directly infringe on people's right to freedom of expression. Indirectly it will inhibit citizens' motivation to express their opinions, demands, and suggestions through various channels.'[69]

The report on Belgium refers to predictive policing as allowing for mass surveillance that will lead to infringements on group rights, such as the right to the freedom of expression and assembly. It further notes that for this reason, one solution may be to allow the use of mass surveillance only as a last-resort technique.[70] Similarly, in Greece, it is reported that many scholars fear the use of predictive policing, as it leads to the constant surveillance of public places, a lack of available anonymity, and ultimately has a chilling effect on the freedom of expression, in contrast to Greek constitutional principles.[71]

3 Predictive Justice[72]

3.1 Definition

Just as for 'predictive policing,' it appears from the national reports that the term 'predictive justice' lacks a legal definition in all reporting countries. In most of them, this is a logical consequence of the fact that they do not use any predictive justice tools – yet. Doctrinal definitions exist; however, they depend much on the study's scope in which scholars shaped them. There are, indeed, two very different realities behind the expression 'predictive justice.'

According to a first understanding, 'predictive justice' is a synonym of 'actuarial justice' and it is possible to define it as 'the use of analytics techniques across data sets with the goal to inform decision-making processes at different stages of the criminal justice system, including sentencing, release, parole and probation.'[73] The analytics techniques referred to are person-based risk assessment tools. Their objective is to foretell human behavior, and principally to evaluate whether a person will commit or re-commit a crime, to enlighten the penitentiary officer or the judge who has to decide on this person. Actuarial justice is far from new and several reporting countries have used it in their judicial or penitentiary system for a longer or shorter period of time.[74] However, for now, in contrast to policing, there are very few examples of AI systems serving judicial objectives – namely sentencing. It seems that such AI-based risk assessment tools operate only in the UK and the USA. Other countries like China and Russia show interest in developing

[68] Report on China, in this volume, p. 283
[69] Zong Bo: Legal Regulation of Large-scale Monitoring in Investigation, published in Journal of Comparative Law, Issue No. 5, 2018.
[70] Report on Belgium, <https://www.penal.org/de/2023-2>, A-09, p. 9.
[71] Article 11, Greek Constitution. Report on Greece, <https://www.penal.org/de/2023-2>, A-06, p. 10.
[72] Sarah Cherqaoui and Juliette Lelieur have written this part of the general report.
[73] Definition suggested by the Dutch report, <https://www.penal.org/de/2023-2>, A-04, p. 35.
[74] The Netherlands, Spain (Catalonia), the UK, and the USA.

them, while European countries are mostly reluctant to forecast human behavior for justice purposes.

A second understanding of 'predictive justice' is more recent and aims to foretell the outcome of a judicial decision based on the probabilistic analysis of former decisions that were rendered in similar cases, rather than based on the probable behavior of a person – though both analyses may obviously be combined in the future. The European Commission for the Efficiency of Justice (CEPEJ), an entity attached to the Council of Europe, thus defines 'predictive justice' as 'the analysis of large amounts of judicial decisions by artificial intelligence technologies in order to make predictions for the outcome of certain types of specialised disputes.' Many European country reports quote this definition, while the report on the USA refers to a similar doctrinal definition.[75] Moreover, many other expressions are used to designate 'predictive justice' in this sense: 'quantitative legal predictions' or 'quantitative legal analysis', 'jurimetrics,' 'legal analytics,' 'automated judicial decision-making,' 'algorithmic justice' or 'statistical justice,' and finally 'legal technology' or 'legal tech.' The reference to a *quantitative* operation is meaningful because probabilities are the true indication that AI-based instruments deliver. Additionally, since the term 'predictive' is inappropriate, as we have highlighted in the introduction of this report, we decided to use the term 'quantitative legal analysis' throughout our analysis. This new method of 'producing' legal decisions is still at its beginning in many countries. Its emergence generates, however, a fierce debate in the legal literature and may have a tremendous effect on criminal justice's future.

Eventually, several country reports mention the use of AI systems in court management, since this is a growing reality in their country, although the relation with the term 'predictive justice' is very thin in this domain. Especially in China, AI serves the administration of justice in a very broad sense, including the formal examination of evidence.[76] This evolution takes place along with an important political will to modernize Chinese justice and aims to support the Government's objective to guarantee 'similar judgment for similar cases.' Therefore, it includes standardization of sentencing through the introduction of AI technology.

There seems to be a common trait between AI systems used for actuarial justice, quantitative legal analysis, and justice management: Either the public institutions that eventually use them or private companies may be their creators and develop them. Mostly, private and public actors seem to move forward hand in hand. Especially concerning the quantitative legal analysis aspect, it is clear that the open data process of judicial decisions enables the private actors to develop legal tech algorithms working based on these data.

---

[75] According to Raffaele Giarda 'predictive justice involves the use of machine learning algorithms that perform a probabilistic analysis of any particular dispute using caselaw precedent', (2022), see the report on USA, p. 213 of this volume.

[76] See point 4. of the general report.

3.2 National practices

3.2.1 Risk assessment tools

Of the countries participating in the study, the USA is the one that uses the most AI-based risk assessment tools. Various jurisdictions in the USA rely on a well-known tool, the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), which was developed in the late 1990s by a private company. Because this commercially available instrument is proprietary software, there is very little transparency with regard to its working. According to some authors, COMPAS, referred to as a fourth-generation tool, uses machine learning while other authors affirm that it is a non-learning algorithmic tool. Originally, COMPAS was not developed for sentencing but rather for making decisions concerning the treatment, supervision, and parole of prisoners; however, it was – and still is – used for sentencing, however. A second tool mentioned in the USA report is the federal Prisoner Assessment Tool Targeting Estimated Risk and Needs (PATTERN), which was developed and implemented by the Federal Bureau of Prisons in 2019. It takes an AI-like approach but does not utilize a fully autonomous machine-learning algorithm. PATTERN determines eligibility for early release. A third AI-based tool mentioned in the USA report, this one developed in 2013, was funded by the National Institute of Justice. It is an advisory tool used by the Pennsylvania Board of Probation and Parole to inform parole release decisions.[77]

Although the use of AI-based risk assessment tools in the USA tends to be advisory, rather than presumptive or mandatory, the staff of the Federal Bureau of Prisons rely solely on PATTERN to decide whether an inmate is eligible for benefits such as early release. Moreover, in some jurisdictions of the USA, the use of risk assessment tools may be required at the pre-trial stage by court order, by the state supreme court, the judicial council, or by legislation.

In the UK, since 2013 prosecutors have relied on the Offender Assessment System (OASys) managed by the Ministry of Justice. OASys is used to assess the risk of harm and reoffending and thus informs decision-making about both sentencing and parole. The Offender Group Reconviction Scale (OGRS) is a key algorithmic component of OASys and is used independently in some circumstances, for instance as a short delivery pre-sentence report. OGRS is currently based on logistic regression, but introducing more advanced machine-learning methods seems to be under consideration. The risk assessments provided through an OASys assessment, combining the professional judgment of a probation officer and OGRS score, can influence judicial decision-making about a suitable sentence, including whether an offender is imprisoned or not. However, judges and magistrates are not under any obligation to follow recommendations based on risk assessments provided via OASys and OGRS.

[77] Report on the USA, in this volume, p. 214-216.

Apart from these two countries, there are a few other jurisdictions awaking interest in such risk assessment tools. In Canada, where there seems to be no statistic-based risk assessment for criminal justice purposes for now, a movement towards developing such tools with the support of AI technology is perceivable. Several research projects aim to create systems able to evaluate the risk of re-offending, to inform decision-makers in the context of bail. Governmental agencies share this interest, like in Ontario. However, Canadian observers of the experience in the USA – especially with COMPAS – warn against the partiality of machine learning tools and fear that discrimination against Indigenous people happens. Besides, some European countries are using statistical tools based on quantitative methods, but not for sentencing like as in the UK and the USA. This is the case in Spain, more specifically in Catalan prisons, where judges of penitentiary institutions use a tool named RisCanvi (Risk change in Catalan), based on a logistic regression system, for the granting of prison permits, parole, classification of the prisoner, and the adoption of supervision measures. In the future, RisCanvi might incorporate modern methods of machine learning. In the Netherlands, OxRec (Oxford Risk of Recidivism Tool) is a traditional actuarial risk assessment tool that provides a probability score about reoffending to the Dutch Probation Services. It is part of a diagnostic tool of the Probation Services, which assesses the offender's likelihood of reconviction, provides the criminogenic needs of offenders, and allows probation officers to formulate supervision plans.

Most European countries reject statistical methods for assessing the risk of reoffending and exclude using AI for such purposes. In Germany, the legal literature is highly skeptical regarding those tools. Assessment of the risk of recidivism in relation to decisions on probation and parole is based on expert reports and, according to case law; experts must not rely on a statistical analysis. They must conduct an individual and all-encompassing evaluation of the person. At most, they may take statistical base rates as a starting point; further individualizing such findings is mandatory. The Italian Code of Criminal Procedure would allow recidivism risk assessment after sentencing, in the correctional phase, but it seems that there is no use of AI by the digital risk assessment tools to deliver reports based on a general scientific evaluation. Nor in Finland do AI-based risk assessment tools operate in the justice system and the other European country reports included in this survey do not even discuss the question.

Finally, in China, scholars seem to promote the development of innovative risk assessment tools: Instead of measuring the risk that people become offenders, their research goal is to evaluate the 'social risk of arrest.' They work from the perspective of arrest and try to provide a quantitative assessment of this social risk, referring to the 'factors affecting people's social learning progress.' The first of the core eight indicators they take into consideration is criminal history – or 'litigation evasion history.' The other criteria are related to the respective person (antisocial personality, criminal attitude, drug abuse, entertainment, and rest habits) as well as to their family environment, educational background, and occupation.

3.2.2 Quantitative legal analysis

In the majority of the reported countries, the development of AI systems for quantitative legal analysis has either begun recently, mainly in the fields of civil or administrative law,[78] or is likely to begin soon.[79] Most examples of AI systems aiming to produce legal solutions to cases based on probabilistic calculation emanate from private companies, principally publishers specialised in legal matters,[80] however not – yet – in the field of criminal law. As the French report illustrates, start-ups tend to build partnerships with legal publishers or educational entities to help them develop AI-based statistic tools, or with courts to test them.[81]

At this stage, it matters to notify that the country reports were drafted before Law ChatGPT was launched; this is the reason why the use of generative AI for drafting legal consultation is not discussed here. Furthermore, it is important to stress that each country report, non- depending on whether AI systems are used for quantitative legal analysis or not in the given country, indicates that the long-term goal is not to replace judges in their function. AI tools are not intended to produce judicial decisions on their own; their purpose is, according to the reports, exclusively to assist judicial actors and facilitate their work while reducing discrepancies in decision-making. This said it seems that there are concrete examples of AI-based quantitative legal analysis tools already operating in the world.

The Chinese report constitutes the most impressive illustration. In 2017, the State Council issued the *Development Plan for a New-Generation Artificial Intelligence*, which puts AI technology at a national strategic level and provides guidance to this aim. The Chinese rapporteur notes: 'Under the new technological revolution, AI is now empowering traditional policing, public prosecution, and court trial to move towards the intelligent justice stage.' This evolution triggers a 'huge potential and application possibility of AI in the construction of criminal justice,' and 'China has successively issued pertinent policies and plans regarding the examination and prosecution and court trial, as well as promoted and guided the integration of AI with intelligent construction step by step.' This is to say that a global strategy is set throughout the country. However, while in some provinces the use of quantitative law analysis is already a reality, in others the process of digitalization of justice seems to be still ongoing. This is why the general report presents some applications of Chinese policy in the current paragraph dedicated to quantitative law analysis but reserves other applications to the following passage relative to digital justice.

---

[78] Argentina, Canada, France, Poland, Russia, Spain, and the USA.

[79] Germany, Greece, Italy, Portugal, and Turkey.

[80] The Argentinian report mentions Sherlock Legal, deriving from IBM's Watson Legal, as well as Legal One (Editorial La Ley), while the Spanish report refers among others to the applications of the Wolters Kluwer group and the French report to LexisNexis 360 Intelligence. There are various tools on the market worldwild, however, we do not provide here for an exhaustive list.

[81] See the French start-up Predictice, which however does not – yet – operate in criminal law.

Already in 2016, the Supreme People's Procuratorate issued a five-year development plan that promotes 'intelligent procuratorial work,' while the Supreme People's Court proposed for the first time to build 'smart courts.' Then, several opinions of the Supreme People's Court tended to enhance 'knowledge-based AI-aided decision-making' for various users, following the overall aim to implement the 'similar judgments for similar cases' objective besides optimizing the allocation of judicial resources. Based on this, the prosecutorial organs of the Guizhou Province, for instance, developed an 'intelligent case research and judgment system,' which makes a pre-research and judgment on the nature of the case and the standard of evidence. It can use the Chinese crime constitution theory and the specific provisions of criminal law to produce a knowledge map of different crime constitution elements. It also compares various sentencing circumstances to produce a standardized map of conviction and sentencing. In the context of pleas for leniency, a Chinese scholar proposes to add data-based prediction to the theoretical prediction system to form a 'dual-core' collaboration called 'AI-assisted accurate prediction and sentencing.'[82] The use of AI is widespread in China at the judgment stage too, especially in Beijing where courts have innovatively constructed the 'Smart Judge' system, a digital platform that offers a comprehensive analysis of previous cases heard by the judge as well as pushes all similar cases by relying on a legal database and a semantic analysis model. Smart Judge also creates a whole process data service, automatically generates a trial outline and record template regarding the trial stage, and finally generates judgment documents regarding the case closing stage. Finally, the 'Enforcement AlphaGo' of Guizhou High People's Court is an "enforcement big data application analysis system" with independent learning ability. It can assist judges in avoiding discrepancies in sentencing, where many problems have been identified by the legal literature – in China, big data and AI assistance are widely seen as a remedy to sentencing problems, and therefore some scholars call for its mandatory use to favor accurate sentencing.

Another Chinese development deserves attention. The 'deviation early warning' system is based on scientific research and a sentencing algorithm that operates through in-depth learning of many criminal documents. It automatically provides early warning for cases with great deviation, thus providing technical support for unifying the judgment standard. It is used in many provinces to avoid different judgments for similar cases.

The Russian report mentions for its part the chatbot LegalApe that was publically presented at the VIII. St. Petersburg International Legal Forum in 2018. The bot can answer questions of a legal nature, while preserving the logic of statements, formulate questions on the circumstances of the case in the context of previous statements, and draft a legal opinion.[83] The Russian report also mentions an electronic system for determining the optimal punishment measure, called 'electronic scales of justice,' which was developed – and tested – to assist courts in choosing what can be considered a fair punishment. Its

[82] Daocui Sun, 'Artificial Intelligence Assisted Accurate Prediction of Sentencing in China -- Taking Plea for Leniency Cases as the Applicable Field' (2020), through the report on China, p. 267 of this volume.
[83] The reports on the UK and the USA also mention the emergence of legal bots.

creators sought to weaken the influence of the subjective human factor, ensure uniform judicial practice, and strengthen the authority of the courts.

Even in Germany, there is an interest of some academics to improve harmonizing the sentencing levels across the country through the creation of a sentencing database with the help of AI technology. Nevertheless, skepticism against AI-assisted assessments of guilt, sentencing, and enforcement of imprisonment remains dominant in Germany. The German report emphasizes the anchoring effects that AI tools have on human decision-making and recalls that, according to constitutional law, judges shall be independent and only subject to the law.

Most countries, however, are currently preoccupied with the digitalization of the judicial process. Consequently, efforts and novelties are concentrated on the digital transformation of justice more than quantitative legal analysis itself. The digitalization of justice is indeed a prerequisite for the implementation of any AI-based system providing for quantitative legal analysis. It mainly implies collecting and retaining a large amount of data, especially judicial decisions. This development involves the development of private players, the so-called Legal Tech, which is emerging in many countries and getting specialized in AI systems before the public authorities. The rapid advent of quantitative legal analysis is, it seems, mainly the consequence of a craze by the private sector, which sees a huge economic market behind these technologies.

3.2.3 Digitalization of justice and court management with AI assistance

In promoting 'intelligent prosecutorial work' in China, the prosecution organs of several provinces have developed 'case management robots.' They do not have all the same capacities and most of them have an impact on evidence questions. In Jiangsu Province, for instance, the robot can compare and analyse the case card filling and various legal documents of the prosecutorial organs, to check the obtained data, and further remain, warn, and evaluate the possible qualitative or evidential problems of the case. It can find out mistakes and defects in case handling documents. In the Tianjin municipality, the prosecutorial work robot has a facial recognition function thanks to which the new visitor's face will be registered and remembered. The robot then handles preliminary business such as case management, prosecution and appeal reception, and business consultation according to the needs of the public.[84]

In the context of limited resources of courts or public institutions involved in the course of criminal justice, which additionally have to handle an increasing number of cases, AI solutions were developed in other parts of the world. South America is a good example since both the Argentinian and the Chilean reports present such evolution. In Argentina, the Public Prosecutor's Office of the City of Buenos Aires has been exploring an AI system deemed to optimize the justice system since 2017. It is called 'Prometea' and operates

[84] See the report on China, in this volume, p. 264.

under human supervision. According to the Argentinian rapporteurs, it can 'read, predict, write, and decide a judicial case in 20 seconds with a 96% accuracy rate.' It can also translate judicial decisions and other legal documents into English, French, and Portuguese. Moreover, the trend toward the digitalization of justice is politically supported in Argentina since a 2018 decree of the Ministry of Modernization urges the 'digital, complete, remote, simple, automatic and instantaneous processing of all documents, communications, proceedings, files, notifications, administrative acts, and procedures.'[85]

In Chile, a very creative tool was developed in 2020 by the Public Criminal Defender's Office, the office in charge of providing defense free of charge to defendants who are accused of a crime. The 'virtual assistant' helps public defenders prepare cases of first hearings for detention controls. It generates indications about the likeliness of a pre-trial detention request, the investigation periods as regards time limits, and allegations of the illegality of the detention. According to the Chilean rapporteurs,[86] it also delivers valuable legal arguments for the discussion of precautionary measures regarding the prosecuted crime. The defender can optimize her or his attention time and provide for a more qualitative defense. The 'virtual assistant' has been operating throughout the Chilean territory during the year 2021. Yet, it is not operational anymore, mainly because of the budgetary impossibility of having external data management services (sufficient data stocking place).

Several European countries have concretely engaged in the process of digitalization of justice, presumably to introduce quantitative law analysis as a next step. In Belgium, the 'gulf' of digitalization has taken place in the criminal justice system and the rapporteur estimates that 'another wave will follow and that will be the use of AI and techniques of predictive justice.' In France, an important law was passed in 2016 with the aim of 'digitalizing the Republic' *(loi pour une République numérique)*. It widely supports the principle of 'open data' concerning judicial decisions. In Spain, the *Digital efficiency project* is presented as a 'nuclear action' promoted by the Government within the 2030 justice program. In this context, both France and Spain have used software for pseudonymizing judicial decisions. Other European countries consider themselves late in this process and took decisions to tend towards it.[87] The Polish rapporteur even notices an already existing influence of AI at several levels: advanced case-law search engines have an impact on the court's decisions and their justification, as well as electronically processed data concerning previous convictions and detentions play a role in the court's decisions. [88]

3.3 Incentives for using AI systems

In all countries using AI systems in the course of criminal justice, independently of which precise form (risk assessment tools, quantitative law analysis, or court management), it appears that the main incentives are reducing costs – linked to human resources – and

[85] Report on Argentina, <https://www.penal.org/de/2023-2>, A-17, p. 6.
[86] Communication at the International Colloquium of Buenos Aires, 30th March 2023.
[87] Greece, Italy, and Poland.
[88] Report on Poland, <https://www.penal.org/de/2023-2>, A-20, p. 9.

improving efficiency – in targeting these resources. Additionally, more justice-focused incentives are mentioned in several reports, like reducing bias and arbitrariness, as well as increasing consistency and transparency of decision-making in the UK. Very similar concerns are shared in the USA, where some claim that AI systems contribute to the harmonization of the application of the law and to the consistency of the sentencing levels across the territory. The stark political will to equalize sentencing in China is a decisive incentive and the aim to modernize the Chinese criminal justice system in a harmonized way all over the country seems to be important too.

More specifically, incentives for using risk assessment tools in the USA include reducing the country's very high incarceration rates, decreasing the disparities caused by cash bail systems, and providing fairer, less punitive outcomes. It is, however, impossible to assess whether these objectives have been met because performance evaluations of risk assessment tools are very rare.

3.4 Assessment of the reliability of AI systems

The reliability of AI-based systems is often presented as high, which is probably a criterion of use for public institutions and a selling point for private companies. In Argentina, for instance, the degree of accuracy of Prometea had been evaluated at 93% but this software had been evaluated by its designers and developers and not by an impartial specialized scientific committee.

In the USA a study published in 2013 of 19 criminal risk and need assessment tools found that validity had been examined in only 'one or two studies.' Another study, this one conducted by EPIC (the Electronic Privacy Information Center) between September 2019 and July 2020, indicates which of 'the numerous tools in use had been subject to a validation study'[89] but does not identify which (in any) of the tools were AI-based and does not say who carried out the study. COMPAS has been evaluated by numerous entities, both independent and internal. In a summary of multiple studies, the (internal) Northpointe Reseach and Development Department concluded that COMPAS was reliable. Nevertheless, an evaluation carried out by ProPublica[90] concluded that risk scores calculated by COMPAS were 'remarkably unreliable in forecasting violent crime.'[91] It is therefore hard to draw a solid conclusion regarding this tool. Another software, PATTERN, is subject to annual review and validation by the Attorney General. The software used to help the Pennsylvania Board of Probation and Parole was evaluated in a paper written by Richard Berk in 2017.[92]

---

[89] Report on the USA, p. 222 of this volume.

[90] According to the report on the USA, ProPublica describes itself as 'an independent, nonprofit newsroom that produces investigative journalism with moral force.'

[91] Julia Angwin et al., Machine Bias, ProPublica (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[92] Richard Berk, An Impact Assessment of Machine Learning Risk Forecasts on Parole Board Decisions and Recidivism, (2017), through the report on the USA, p. 216 of this volume.

All authors of country reports addressing the question of the reliability of AI systems used by judicial authorities highlight that a periodical evaluation of the accuracy of those systems by their users is essential. Moreover, external and independent experts, if possible appointed by a public authority, must regularly review the reliability of the systems,[93] especially if they work based on machine learning, as this is a technology that is permanently evolving through 'self-learning' algorithms. This is without prejudice to auditing by the private companies that developed the tool and aim to improve it and adjust its functioning to the need of an efficient and fair trial.

3.5 Normative Framework

3.5.1 Scare pieces of legislation

General and countrywide legislation on the use of AI for actuarial justice and quantitative law predictions is still inexistent. Instead, most Governments favored a general approach by the Executive and submitted plans, like in China,[94] or enacted decrees to approve national strategies for the development of AI without specific reference to criminal justice, like in Russia.[95] In the USA, however, legislative activity has taken place at the state and local levels. For example, legislation enacted in the state of Idaho in 2019 specifically addresses questions of the transparency, accountability, and explainability of pretrial risk assessment tools. Concerning quantitative legal analysis, a French Act of 2019[96] prohibited the use of data that enable the identification of judges and other justice agents to profile or rank them, or to evaluate, analyze, compare, or foretell their professional practices.[97]

Legislation regarding data protection partly compensates for the lack of general legislative frameworks, as underlined by most of the national reports. Machine learning-based systems need a voluminous amount of data to be functional and more or less trustworthy. In the context of quantitative law analysis, data may be personal and sensitive, which raises the question of their protection. International regulations progressively appeared in that matter, especially on the European continent: The Convention for the Protection of Individuals with regard to Automatic Processing of Data of the Council of Europe (better known as the Convention 108+), the European Union's General Data Protection Regulation, and the Law Enforcement Directive. This directive protecting individuals about the processing of their data by police and criminal justice authorities is interesting

[93] See for instance the report on Spain, <https://www.penal.org/de/2023-2>, A-05, p. 14.

[94] See above, 3.2.2.

[95] Decree of the President of the Russian Federation of October 10, 2019 N 490 on the development of AI in the Russian Federation; decree of the Government of the Russian Federation of August 19, 2020 No. 2129-r on the approval of the concept for the development of regulation of relations in the field of AI technologies and robotics until 2024, through the report on Russia, <https://www.penal.org/de/2023-2>, A-07, p. 7.

[96] Law nb. 2019-222 of 23 March 2019, Loi de programmation et de réforme pour la justice.

[97] Art. L. 10 of the Code de la justice administrative and art. L. 111-13 of the Code de l'organisation judiciaire.

because Article 11 prohibits decisions affecting people based solely on automated processing, including profiling apart from a few framed exceptions.

However, data protection is not an issue limited to Europe. In Canada, for instance, at the correction stage, the collection, sharing, and protection of personal data by federal penitentiaries are specifically framed by the *Corrections and Conditional Release Act*. On its side, China has two specific laws about personal data, the Data Security Law of China and the Personal Information Protection Law of China, which helped clarify the boundaries regarding AI-based systems and personal data protection.

Even in the absence of specific laws on using AI in the course of criminal justice, many governments show an interest in that matter, and working groups and committees have been set up. They help develop legislation projects.

3.5.2 Projects of legislation

In some countries, there are attempts to introduce general legislation. In the USA, a bill, the *Justice in Forensic Algorithms Act of 2021,* was introduced in the House of Representatives. Had it passed, it would have established a federal framework to govern the use of computational forensic software.[98] In Spain, the *Draft Law on Digital Efficiency Measures of the Public Service of Justice* came out in 2022 and is still under discussion.

However, to this date, it seems that there is no existing or planed legal framework specifically dealing with the reliability of AI technology and the effective control of human operators over it. Similarly, the national reports could not provide information on legislation or projects of legislation regarding labelling or certification of AI systems, not even to ensure that they are compatible with the general principle of criminal justice and human rights.

3.5.3 Soft law

The first international soft law instrument specifically dealing with the use of AI systems for justice purposes is the *European ethical Charter on the use of artificial intelligence in judicial systems and their environment.* The European Commission for the Efficiency of Justice (CEPEJ) of the Council of Europe adopted it in December 2018 to guide policymakers, legislators, and justice professionals. The Charter sets out five ethical principles: 1) Respect for fundamental rights; 2) Non-discrimination; 3) Quality and Security; 4) Transparency, impartiality, and fairness; 5) and finally the principle of 'under user control.' The content as well as the various impacts of these five principles are further detailed.[99] They do not specifically concern criminal justice but provide interesting guidance on how to receive AI technology in judicial environments. Although the provisions on the rights regarding AI of the 2021 Portuguese Charter and Spanish Charter on digital

[98] Report on the USA, p. 229 of this volume.

[99] <https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment> accessed 30 November 2023.

rights[100] do not focus on judicial matters, they echo the European ethical Charter on several substantial points. In the UK, the government guidelines *Understanding artificial intelligence ethics and safety*[101] concerns building and using AI in the public sector, but they are relevant for judicial matters.

By contrast, an example of a national soft-law source focused on justice matters is the Model Penal Code in the USA,[102] which in its 2017 revision prominently endorsed the consideration of risk in the sentencing process (MPC-S).[103] Risks and needs processes developed by the sentencing commission – including, presumably, those based on AI – may be incorporated into the sentencing guidelines if they are sufficiently reliable. Turning to quantitative law analysis issues specifically, the French National Bar Council (Conseil National des Barreaux) adopted the *Charter on Transparency and Ethics in the Use of judicial data*, in October 2020. This text develops eleven ethical principles and representatives of the legal tech industry signed it.

Soft laws also concern data protection in many reported countries. For instance, Argentina, Chile, Portugal, and Spain are part of the Ibero-American Network of Data Protection, which, on June 20, 2017, approved the Standards for Data Protection for the Ibero-American States and prepared two guiding documents for the proper use of personal data in the design and implementation.

3.5.4 Case law

Existing case law mainly concerns the questions of the reliability and impartiality of AI risk assessment tools, the lack of which endangers the right to due process. This case law first appeared in the USA in 2016, in the highly controversial case of *State v. Loomis*.[104] One claim made by Loomis, who was sentenced after a risk evaluation provided by COMPAS had assessed him as a high-risk person, was that he had suffered a violation of his right to be sentenced on the basis of accurate information. The Wisconsin Supreme Court rejected all of Loomis's due process challenges, pointing out that variables used by the COMPAS algorithms were publicly available, that the outcome of the risk assessment was based either on Loomis's answers to a questionnaire or on publicly available information, and that risk scores were not used as the sole determinative factor in sentencing so that Loomis in fact received an individualized sentence. However, recognizing that 'risk assessment tools may not perform as well for non-whites as for whites' and that

[100] See p. 22 of this volume.

[101] UK Government, 'Understanding artificial intelligence ethics and safety' (2019) <https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety> accessed 30 November 2023.

[102] The Model Penal Code, first promulgated in 1962, is a model code assembled by the American Legal Institute.

[103] <https://robinainstitute.umn.edu/publications/model-penal-code-sentencing-proposed-final-draft-approved-may-2017> accessed 30 November 2023.

[104] *State v. Loomis*, 881 N.W.2d 749, 765 (Wis. 2016).

'the accuracy of such tools, without constant re-norming, is short-lived,'[105] the Wisconsin court 'essentially implemented a mandatory disclaimer on the practice of using a COMPAS risk assessment at sentencing' and stressed that risk scores may not be used as the sole determinative factor in sentencing. Cases from other jurisdictions of the USA have been decided in the same vein.[106]

Addressing more specifically the question of the reliability and impartiality of the *data* used by AI systems, the Canadian Supreme Court held in *Ewert v. Canada*[107] that the duty of reasonable verification of data accuracy extends to the *results* produced by an actuarial risk assessment tool, which supposes that the tool delivers strongly reliable assessment. In doing this, the Supreme Court poses a jurisprudential standard that aims to prevent discrimination – in the given case, the assessment system had been used towards Indigenous offenders. It is interesting to note that case law in both the USA and Canadan rules on the problem of discrimination against people of colour, while the questions posed to the courts concerned the reliability and accuracy of AI technology.

In addition, the question of transparency of AI-supported decision-making appears in the case law of the USA as well as that of the Netherlands. Transparency of AI calculations towards the addresses of decisions is a condition for verifying their accuracy and challenging them. In the USA, the case *State v. Walls*,[108] decided in 2017, provided an important ruling regarding the effectiveness of the right of defense. The Kansas Court of Appeals held that the defendant must have access to the risk assessment report to be able to review and verify the questions, answers, and scoring decisions contained in this report. Depriving the defendant of the report 'necessarily denied him the opportunity to challenge the accuracy of the information [provided by the AI tool] upon which the court was *required* to rely in determining the conditions of his probation.' The Dutch Supreme Court issued several civil law judgments concerning the use of an automated decision-making system by a government body and stated as a rule that 'stakeholders need to be able to able to verify the correctness of the decision made in the automated process as well as the correctness of the data and the assumptions underlying the process.'[109] The concerns on transparency and explainability of AI systems are emerging too when the technology is used for building evidence, as will be discussed in the next part of this report.

---

[105] Chris Miller, 'The Prospects of Constitutional Challenges to COMPAS Risk Assessment (26 April 2021)': see the report on the USA, in this volume, p. 233.

[106] Report on the USA, in this volume, p. 235-237.

[107] *Ewert v. Canada*, 2018 SCC 30 (CanLII), [2018] 2 SCR 165, <https://canlii.ca/t/hshjz>, accessed on 21 November 2023.

[108] *State v. Walls*, 2017 Kan. App. Unpub. LEXIS 487; 396 P.3d 1261 (Kann. App. 2017). According tot he Kansas Rules of Appellate Procedure, Rule 7.04, an unpublished memorandum opinion such as this one is not binding precedent and is not favored for citation.

[109] Supreme Court of the Netherlands, judgment of 17 August 2018, ECLI:NL:HR:2018:1316 (case nr. 17/01448).

3.6 General principles of criminal justice

Considering the questions raised in the case law, it is not surprising that the principle of non-discrimination is the most frequently discussed general principle in the reported countries (3.6.1). It mainly concerns the use of risk assessment tools, but also, to a smaller extent, the use of quantitative legal analysis. Then come a series of guarantees that are linked to the right to *due process* or a fair trial. The right to an independent judge (3.6.2) as well as the right to an adversarial trial that guarantees the equality of arms (3.6.3) and the right to appeal (3.6.4) seem to be significantly challenged by the use of machine learning in the realm of criminal justice. Interestingly, a new requirement linked to the right to a fair trial is emerging: should the right of access to a 'human judge' be recognised? (3.6.5). Moreover, in several countries, there are concerns about the transformation of law that would result from the use of quantitative legal analysis, since it uses mathematical calculation instead of legal reasoning (3.6.6). Finally, some country reports discuss the phenomenon of privatization of justice, which is increasing with the emergence of AI systems, and might collide with the right to equality of citizens before criminal justice (3.6.7).

3.6.1 Principle of non-discrimination

All reports on countries using risk assessment tools based on machine learning highlight that discrimination is a serious concern.[110] The danger of discrimination seems to be acknowledged worldwide by the legal literature and in the media, and it appears that the *Loomis* case has played an important role in this matter.

Discrimination is also a matter of concern in the context of the emerging quantitative legal analysis. It is referred to in many country reports, even where the given country is not using machine learning for preparing the production of judicial decisions yet.[111] The Chinese report emphasizes that 'if algorithm designers deliberately write programs with subjective judgment, algorithm manipulation will occur.'[112] In the same vein, the Canadian report highlights that data implemented in algorithms or the algorithms themselves may be problematic: 'just like any other technological artifact, code is not neutral, but inherently political [...].'[113] Choices are made as to whether or not to include certain variables in the algorithms, and it is a reality that some software will include data that will be excluded by other software. These choices are subjective and undermine the apparent objectivity of the statistical tool.[114] Beyond those choices, unintended discrimination may happen. As underlined by the Greek report, by using past decisions combined with other

[110] Reports on China, the Netherlands, the UK, and the USA.
[111] Argentina, Belgium, Canada, China, France, Greece, and Italy.
[112] Report on China, p. 280 of this volume.
[113] Report on Canada, quoting 'Code is Law' by Lawrence Lessig <https://www.penal.org/de/2023-2>, A-03, p. 52.
[114] Report on Canada <https://www.penal.org/de/2023-2>, A-03, p. 51.

data, AI systems 'may reproduce and entrench bias, discrimination, and inequality, particularly as far as minorities and disadvantaged groups are concerned – giving rise to the so-called algorithmic bias.'[115]

3.6.2 Principle of independence of judges

Whether the use of AI systems is compatible with the right to an independent judge, an element of the right to a fair trial according to Article 6 of the European Convention of Human Rights, is another important issue, except in China and North America where discussion on this matter does not seem to be widespread.

On the European continent, the notion of 'robot judge' has widely emerged. It expresses the fear that a machine is making judgments instead of a judge, which in the field of criminal justice seems particularly unsuitable. A significant part of the legal literature worries that judges and other judicial actors, who are usually very busy and have to face the pressure of performance, might be tempted to delegate some of their work to AI. The Spanish report underlines that following the suggestion of an AI system without any kind of subsequent verification is practically delegating the decision to the system. In Turkey, the 2021 report of the AI Working Group of the Istanbul Bar Association states that AI-based risk assessment tools may pose problems regarding the independence and impartiality of judges. The Greek report adds that the lack of specific training for legal professionals in AI technology can worsen the phenomenon: they are not educated to filter critically the outcomes exposed to them to keep their part of discretion. In the United Kingdom, however, probation officers and judges are trained to use OASys and OGRS so they can manage to use these tools without losing their professional judgment. However, the report underlines the possibility of a risk-averse approach regarding high scores. In the Netherlands, research shows that judges generally do not blindly follow the result of OxRec but use it alongside their evaluation.

Many voices clarify that AI systems only *assist* decision-making, they are not deemed to *replace* judges.[116] They insist that the final decision remains by the judge. However, when calculations by AI do not only deliver a risk assessment but also suggest a legal decision based on this assessment, the technological output is steering the judge enough to worry about the real independence of the judge. As the UK report highlights, AI tools may appear more objective than they are in fact. This encourages judges to place great reliance on them.[117] The formula 'automation bias' commonly expresses that humans tend to trust statistical results because of their scientific aura. It is in general very difficult for human decision-makers to refute a 'recommendation' made by a high-tech tool, usually they

[115] Report on Greece < https://www.penal.org/de/2023-2>, A-06, p. 13.

[116] Report on the Netherlands <https://www.penal.org/de/2023-2>, A-04, p. 43; report on Belgium, https://www.penal.org/de/2023-2, A-09, p. 17; report on Russia, https://www.penal.org/de/2023-2, A-07, p. 15.

[117] Report on the UK, https://www.penal.org/de/2023-2, A-14, p. 10.

even fear departing from solutions given by AI.[118] In this context, the judge may lose her discretion when relying on the software and therefore freely renounce a part of her independence.[119]

The Dutch report suggests that 'the differentiation between assistance and steering of decision-making is a useful starting point.'[120] The problem is how to make this differentiation in practice. Moreover, the German report warns that the use of AI tools to assist merely (human) judges should not be underestimated. Their findings may have a strong 'anchoring effect' on human decision-making even though they solely assist the judge.[121]

Discussions on potential threats to the independence of judges also concern quantitative legal analysis. As the report on Italy points out, social pressure could push judges to follow the 'normative force of numbers.'[122] When a solution is presented to the judge as a 'scientific, impartial and technological output,'[123] the chances are high that she will rely on the result presented to her. This may create an overreliance of judges on AI systems, and make them ignore any contradictory information. The risk is that they finally make their decision only based on automated decision-making systems, thus confirming the so-called 'automation bias' that several national reports are denouncing.[124] The Turkish report mentions that the developers of algorithms are mostly unfamiliar with the legal system and its principles, whereas those who implement the law use these technologies automatically in the face of complexity and obscurity in algorithms: These factors may indirectly harm the independence of the judiciary.[125] A similar concern arises in France, where scholars and practitioners both highlight the performative power of quantitative legal analysis software.[126] A Belgian author argues that when the tool is making the decision, it overtakes the task of the judge, which is contrary to the Constitution.[127] The author of the Belgian report finds that more precise rules should determine the role that future

[118] Ales Završnik, 'Criminal Justice, Artificial Intelligence Systems, and Human Rights', ERA Forum 20 (2020) 567–83, 574.
[119] Report on the Netherlands, https://www.penal.org/de/2023-2, A-04, p. 40; report on Germany, https://www.penal.org/de/2023-2, A-02, p. 31-32.
[120] Report on the Netherlands <https://www.penal.org/de/2023-2>, A-04, p. 40.
[121] Report on Germany <https://www.penal.org/de/2023-2>, A-02, p. 31-32.
[122] Report on Italy, p. 204 of this volume.
[123] Ozan Can Özbalçık, 'Artificial Intelligence-Based Risk Assessment Tools in Criminal Procedure and Its Legal Effects', through the report on Turkey <https://www.penal.org/de/2023-2> accessed 30 November 2023, A-16, p. 18.
[124] Report on Belgium <https://www.penal.org/de/2023-2> A-09, p. 17; report on Canada <https://www.penal.org/de/2023-2>, A-03, p. 58-59; report on Turkey <https://www.penal.org/de/2023-2>, A-16, p. 18.
[125] Özgür Taşdemir, 'Ceza Adaletini Dijitalleştirmek, Büyük Veri Vicdani Kanaate Karşı', in Yaşar Bilge (ed.), *Sağlık Alanında Büyük Veri Analitiği ve Uygulamaları* (Türkiye Klinikleri 2021).
[126] Report on France, p. 179 of this volume.
[127] G. Vanderstichele (2020), quoted in the report on Belgium, A-09 <https://www.penal.org/de/2023-2>, at p. 15.

quantitative legal analysis instruments may have, to protect the independence of judges.[128]

3.6.3 Right to an adversarial trial and equality of arms

First to mention is the fear that the contradictory procedure and its subtleties would be evacuated in favor of statistical reality if the quantitative legal analysis was implemented in courts.[129] Worst, if the AI systems used by the prosecution services and the defendant counselor are the same, the right to an adversarial trial would be deprived of its substance.[130] Second, it appears that most country reports perceive the absence of transparency around the AI tools used in trials as a major hurdle against the exercise of the right to an adversarial trial. The Turkish doctrine concentrates on the so-called 'black box problem' arising out of the use of unexplainable AI systems and stresses that the suspect's right to defense is impaired if the suspect cannot learn what kind of data the AI system processes and how it is programmed.[131] The Turkish report points out the need for a legal regulation that specifically foresees the obligation to expose which data are processed via an AI system and how the algorithm functions (source code and training data). In Argentina, the conceivers of the software *Prometea,* developed in collaboration with the Public Prosecutor's Office of the City of Buenos Aires, seem to have considered transparency. According to the report on Argentina, the software was designed following the principle of algorithmic transparency and traceability. Argentinian doctrine refers to those standards as 'white boxes' in opposition to the 'black box' phenomenon deeply linked to machine learning and accentuated by the rise of deep learning.

An important question is whether the addresses of a judicial decision based on machine learning are in the position to challenge the outcome of the tool. In none of the reporting countries, there is an adequate procedure allowing judicial review on the accuracy of the statistical results provided by AI. The author of the Belgium report firmly affirms that 'nobody will go against the idea that it should be possible to challenge AI in courts.'[132] However, it seems 'doubtful that parties will be able to challenge the outcome of predictive tools only based on their right to an adversarial trial' because of the lack of transparency and the complexity of the technology.[133] This is why software developers should be heard as witnesses in court, and lawyers will have to work together with computer scientists to make sure that the reliability of AI outcomes is properly tested. [134] In Italy, the administrative supreme court has recognized the right of those who suffer the effects of an algorithmic public decision to get a review of how the algorithm works and what the datasets used are. Although this position only concerns administrative decisions, it

[128] Report on Belgium <https://www.penal.org/de/2023-2>, A-09, p. 16.
[129] Report on Canada <https://www.penal.org/de/2023-2>, A-03, p. 41.
[130] Report on Greece <https://www.penal.org/de/2023-2>, A-06, p. 16.
[131] Report on Turkey <https://www.penal.org/de/2023-2>, A-16, p. 19.
[132] Report on Belgium <https://www.penal.org/de/2023-2>, A-09, p. 17.
[133] Report on Belgium <https://www.penal.org/de/2023-2>, A-09, p. 17; report on Greece <https://www.penal.org/de/2023-2>, A-06, p. 12.
[134] Report on Belgium <https://www.penal.org/de/2023-2>, A-09, p. 17.

seems difficult to reverse it for judicial decisions. Following this path, the idea of a specific procedure providing an appropriate framework to challenge the technological aspects of AI and the material it uses for judicial purposes should make its way. In the USA, however, while 'the right to challenge decisions with significant effects is a core principle of the rule of law,' it seems that 'the recent trend has been to favor systemic governance over the companies or government entities that build and use Al over establishing individual rights such as a right to contest.'[135]

Concerning the requirement for equality of arms, several reports stress that asymmetries may arise in courts due to the use of AI, in different constellations though. The report on the USA takes the example of law enforcement authorities being in the position to access data possessed by the private companies that have developed the AI system they are using, while investigators for the defense cannot have access to these data.[136] The Belgian report points to another problematic situation, where private parties can afford AI tools while the prosecutors and judges cannot, because of the restricted budget provided by the State. Finally, the report on China acknowledges that court informatization may bring inequality of litigation rights, and highlights that Chinese scholars have proposed to ensure the equality of prosecution and defense through 'information isolation' and 'information disclosure.' The term 'information isolation' refers to shielding the judge from information that she obviously should not know, while the term 'information disclosure' expresses the requirement that unfavorable information to the defense is fully disclosed to it.[137]

3.6.4 Right to appeal

Another important question concerns the efficiency of the right to appeal in case the same AI tool assists the judge of the first instance and the judge of appeal. Several country reports, including France, Greece, Italy, and the Netherlands, point out the paralysis of the appeal system if the software used at first instance and on appeal are identical: the right to appeal would simply become illusory. The Dutch report suggests two solutions: either the appeal should be left to the human judge alone, or the scope to overturn decisions made by an AI system should be limited, thus limiting the scope of the right to appeal – but considered as more efficient by a part of the literature.[138] The first option would indirectly guarantee access to a 'human judge' in appeal, as an element of the fair trial.

[135] Kaminski and Urban, through the report on the USA, p. 245 of this volume.
[136] 'Privacy Asymmetries: Access to Data in Criminal Defense Investigations' (2021) through the report on USA, p. 248 of this volume.
[137] Zheng Xi, 'Conflict and Coordination Between Court Informatization and Citizens' Criminal Procedure Rights' (2020), through the report on China, p. 284 of this volume.
[138] Hildebrandt, through the report on the Netherlands <https://www.penal.org/de/2023-2> accessed 30 November 2023, A-04, p. 43.

3.6.5 Right to a fair trial. Is there a need for a right of access to a 'human judge?'

The right to a fair trial includes the right to access a court. It is questionable whether this right is properly guaranteed when the court is not composed of human judges. What if a quantitative legal analysis tool delivers the decision? Alternatively and more realistically, what if AI assists the court, given the fact that human judges highly rely on the statistical results AI provides? This raises the question of the need for a right to a human judge to satisfy the requirement for a fair trial.

The authors of the report on Italy suggest that the whole discussion about the risk of jeopardizing the right to a fair trial concentrates 'on the question of whether a quantitative law prediction process is a trial at all.' They argue that a 'reliable automated decision-making process based on quantitative law prediction can be conceived exclusively in relation to simple cases, in which the number of the variables at stake, both material and procedural, are extremely limited. Outside these boundaries, there cannot be the illusion of accomplishing the task of a trial, nor fair, neither unfair.'[139] In France, an author emphasizes that an algorithm can hardly be considered a court within the meaning of Article 6(1) ECHR, at least as far as its ability to provide all the guarantees associated with that concept is concerned. That observation, coupled with the fact that a syllogistic algorithm does not reflect the reality and complexity of a judicial decision, raises doubts about the compliance of predictive justice tools with fair trial rights.[140] The Greek report seems to follow the same line, quoting Article 8 of the Constitution reading that 'no one shall be deprived of the judge assigned to him by law against his will.' According to the authors of the Greek report, it must be understood that the term 'judge' refers to a natural person who is a member of a court.

Still, many voices leave room for the assistance of AI in the course of criminal justice without considering that this violates the right to a fair trial. As mentioned above about the right to appeal, access to an appeal trial in which a human judge is exclusively handling the case might be a solution, at least for less significant cases. In this vein, the French National Consulting Commission on Human Rights recommends that the persons who are the subject of a decision based on algorithms are systematically informed of it, and have the right to judicial review by a human being if the decision has significant consequences for them.[141] In contrast, the French Data Protection Authority finds that the role of human agents could be placed at a collective level rather than humans supervising every single decision – which would annihilate the optimizing effect of AI systems. It suggests that the use of algorithms be controlled by examining their design and all the direct and indirect effects they produce on the justice system.[142] The Spanish report argues that, as no country plans to replace judges with software shortly, the debate should

[139] Report on Italy, p. 207 of this volume.
[140] S.-M. Ferrié, 'Les algorithmes à l'épreuve du droit au procès équitable,' (date) through the report on France, p. 191 of this volume.
[141] Report on France, p. 191 of this volume.
[142] Report on France, p. 192 of this volume.

focus on the 'necessity to have a human judge behind decision-making.' It cites article 117.3 of the Spanish Constitution, which recognizes the principle of jurisdictional exclusivity. Spanish doctrine finds that this principle requires judges and courts to exercise jurisdictional power, however, it does 'not specify how or through which tools.' Spanish authors therefore consider it acceptable to use AI systems in a complementary way, acting as a support for the decision that the judge must make.[143] This position follows the line of the European Union. Directive (EU) 2016/680 settles a general prohibition of 'decision-based solely on automated processing, including profiling, which produces an adverse legal effect concerning the data subject or significantly affects him or her.' Exceptions might though be authorized by the Union or a Member State, with appropriate safeguards including the right to obtain human intervention on the part of the controller. It, therefore, excludes an AI system from making a judicial decision in criminal matters without human intervention but does not guarantee the right of access to a human judge.

Finally, the report on the USA quotes an author who favors 'a right to a well-calibrated machine decision' rather than a right to a decision taken by a human judge, in part because 'machines have the capacity to classify and predict with fewer errors than humans.'[144]

3.6.6 Legal reasoning v. mathematical calculation

A major point of concern about steering judicial decisions through risk assessment tools relates to the core logic of criminal law. Court condemnations and punishments are a repressive answer to criminal facts. As an answer, they necessarily intervene after these facts have happened. By using risk assessment tools, though, authorities tend to use a reversed approach, the reports on several countries note.[145] As the French and Italian reports point out, it is partly the same approach as Lombroso's Italian positivist doctrine developed in the 19th century: preventing crimes based on variables.[146] Whether physical or social, the variables considered to establish the supposed dangerousness of an individual no longer serve a punitive, but a preventive goal. The Canadian report stresses that the use of a risk assessment tool blurs the boundaries between preventive and repressive measures. Do we punish the individual for the acts he has committed or those we foresee he might commit in the future?[147]

The quantitative legal analysis seems to be overall controversial as to the method it relies on. Because it works based on statistical reasoning instead of using the traditional legal syllogism and mathematical calculation to avoid the subjective appreciation of facts, the literature worries that judicial decision-making loses flexibility, nuances, and the essence

[143] R. Borges Blazquez, through the report on Spain <https://www.penal.org/de/2023-2>, A-05, p. 13.
[144] Aziz Z. Huq, A Right to a Human Decision, through the report on the USA, p. 244 of this volume.
[145] France, Greece and Italy.
[146] Reports on France and on Italy, p. 166 and p. 204 of this volume.
[147] Report on Canada <https://www.penal.org/de/2023-2>, A-03, p. 62.

of real-life situations.[148] Discretion is inherent to the exercise of justice, as the Canadian and the Portuguese report recall, and discretion includes intuitive thinking and personal valuations of factual situations. The Turkish report adds that 'cognitive characteristics such as the psychology of justice, risk-taking, and reasoning cannot be calculated mathematically.'[149]

The Canadian report emphasizes that machine learning provides statistical results that do not rely on causality as legal reasoning does. Instead, these results derive from the establishment of mere correlations, which is a truly different approach.[150] Moreover, quantitative legal analysis works based on the study of past decisions, to identify the most probable outcome out of former decisions rendered in a similar legal issue. It therefore links future case law to past case law, thus operating similarly to the system of common law, in which the rule of stare decisis is well established. The reports on France and Italy underline that this differs much from the civil law tradition where previous cases do not bind the judge. Instead, in civil law countries judges are asked to apply only the law. The French report therefore questions whether AI system reasoning is compatible with the principle of criminal legality, enshrined in Article 7 of the Declaration of the Rights of Man and of the Citizen of 1789.[151] The Belgian report adds that evolutions of the case law are seriously inhibited if judges refrain from departing from existing case law. The AI system could assess their 'new' solution as deviant whereas it legitimately provides an up-to-date appreciation of the law.

Besides, the reports on Belgium and Canada recall the symbolic and ritual function of criminal trials. Is it possible to serve this function when the issue of the trial is foreseeable? The report on Russia highlights a possible decrease in the authority of judges as a possible consequence of their use of AI too. Finally, this report formulates loudly what many lawyers in the world fear: after future young judges start to rely on machines, will there be a 'generation of incompetent judges?'[152]

3.6.7 Privatization of justice and equality of citizens before criminal justice

While several reports do not discuss the risk of privatization of justice through using AI systems,[153] one report expressively refutes this risk,[154] and other reports, in contrast, highlight that there is concern, in their country, about the fact that it is for private companies to develop most AI systems. This reduces transparency and accountability in criminal

[148] Report on Belgium, Report on Belgium <https://www.penal.org/de/2023-2> , A-09, p. 19 and report on Turkey <https://www.penal.org/de/2023-2>, A-16, p. 19.
[149] Report on Turkey <https://www.penal.org/de/2023-2>, A-16, p. 18.
[150] Report on Canada <https://www.penal.org/de/2023-2>, A-03, p. 53.
[151] Report on France, p. 192 of this volume.
[152] Report on Russia <https://www.penal.org/de/2023-2>, A-07, p. 15.
[153] China, Germany, The Netherlands, Portugal, Russia.
[154] Report on Italy, p. 210 of this volume.

justice.[155] Especially in the USA, where private developers play a significant role in sentencing determinations, it is problematic that they are not subject to the traditional constitutional accountability mechanisms. In France, judges try to resist standardization through quantitative legal analysis tools developed by private firms, which they cannot control, and the report argues that AI systems designed by private firms endanger the role lawmakers play in criminal law.[156] The Spanish report mentions that many of the 'AI solutions' are implemented by private companies in public institutions through public partnerships with very limited competition since there are very few companies specialized in these new technologies. It underlines the problem of the private participation in the management of data and sensitive information usually collected in police databases,[157] and recalls, as the Belgian report does, that the companies producing AI systems could obey the interests of certain lobbies, or aim the maximization of benefits to the detriment of ethical principles' implementation. Finally, the Canadian report highlights a specific influence that AI systems may have when used by law firms to inform their clients whether they should better accept plea-bargaining or out-of-court settlements, rather than go to trial.[158] In this situation, which goes well beyond the Canadian example, the decision of the accused person to avoid a trial heavily depends on the advice given by the law firm. In case a company is concerned and internal investigations are needed to prepare an out-of-court settlement, it probably lies in the interests of the law firms to push for an out-of-court settlement.

Another consequence of the use of AI systems by (some) law firms to foresee the outcome of criminal judgments is that not all litigants can afford the high fees of these firms. Existing inequalities of litigants before criminal justice may increase that way.[159] In the long term, however, the reverse situation may appear. If, in the future, all law firms invest in AI systems, the defense of accused persons may become standardized – and cheap. In contrast, tailor-made defense would develop into a luxury service and be affordable only to fortunate people. An inequitable two-tiered system could rise out of this situation. [160]

4 Evidence[161]

The rapporteurs of fourteen countries have participated in this part of the survey: Argentina, Canada, China, Germany, Finland, France, Greece, Italy, the Netherlands, Poland, Portugal, Russia, Spain and Turkey. These countries represent various legal traditions in criminal procedural law, and beyond this, their rules on evidence may largely differ from one country to another, even inside one legal tradition as on the European continent. All

[155] Report on the UK <https://www.penal.org/de/2023-2>, A-14, p. 9 and the USA, p. 247 of this volume.
[156] Report on France, p. 180 of this volume.
[157] Report on Spain <https://www.penal.org/de/2023-2>, A-05, p. 14.
[158] Report on Canada <https://www.penal.org/de/2023-2>, A-03, p. 60.
[159] Report on Belgium <https://www.penal.org/de/2023-2>, A-09, p. 19.
[160] Reports on Greece <https://www.penal.org/de/2023-2>, A-06, p. 16; report on the UK <https://www.penal.org/de/2023-2>, A-14, p. 4-5 and report on the USA, p. 247 of this volume.
[161] Eftychia Bampasika and Juliette Lelieur wrote this part of the general report.

over these countries, AI systems are used for several purposes: mostly to help gather (4.1) but also to produce evidence (4.2) – sometimes both at the same time, as the Dutch report illustrates. Assessment of evidence through AI is also making its way, particularly in China (4.3).

4.1 Evidence gathering through AI

4.1.1 National practice

A first finding is that more or less all participating countries have already deployed, or are about to deploy some kind of AI to facilitate the evidence gathering. AI systems are used to detect and obtain evidence for a wide spectrum of criminal activities comprising fraud, economic crimes, cybercrime, forgery, web-based child sexual exploitation, and violent crimes. Another common element is the lack of transparency and publicly accessible information as to these AI systems, their function, and the agencies or authorities that have access to them and their outcomes. Several national rapporteurs were confronted with police discretion. Sometimes, even without official state information about the use of AI systems in evidence gathering in their country, they suppose that some law enforcement and judicial units already use such tools, like in Finland.[162]

Russia is among the countries[163] that employ many different AI systems. The Department of Criminology of the Ural State Law University is developing an artificial neural network for identifying signs of forgery of signatures made without the use of mechanical and computer devices.[164] Moreover, several AI systems provide forensic support to law enforcement authorities, like for instance the *Block system* in the investigation of economic crimes, and the *Octopus system* in establishing contact to contacts of criminals. Further, the *Mirror program* allows the synthesizing of video images of people for detecting deepfakes.[165]

To find evidence among huge amounts of data gathered in contemporary criminal investigations, the National Forensic Institute (NFI) of the Netherlands developed an AI system called Hansken.[166] Several investigative bodies used Hansken in the Netherlands, including the Dutch National Police for criminal investigation, the Dutch Fiscal Information and Investigation Service for fraud detection in tax investigations, the Netherlands Food and Consumer Product Safety Authority, and the Human Environment and Transport Inspectorate. Hansken allows the extraction and processing of data from all types of digital devices, such as laptops, smartphones, hard disks, and even whole servers. This concerns various kinds of structured and unstructured data, including names, keywords, phone numbers, chat messages, photos, videos, various types of metadata, and location data. The Dutch report notices that new AI tools first interpret the data and

[162] Report on Finland, in this volume, p. 290.
[163] The same is true for Canada, China, France, Germany, Italy, and the Netherlands.
[164] Report on Russia, <https://www.penal.org/de/2023-2>, A-07, p. 20.
[165] Report on Russia, <https://www.penal.org/de/2023-2>, A-07, p. 20.
[166] Hansken: The Open Digital Forensic Platform, www.hansken.nl.

then find correlations or links between them. This means that they do not simply *gather* data that may be relevant for evidence but also *produce* new data that may be pieces of evidence themselves.[167]

Similarly, in many countries, law enforcement authorities are equipped with software that extracts voluminous sets of data from digital devices and/or analyzes them. *UFED* (Universal Forensic Extraction Devise) designed by the Israeli company *Cellebrite* is capable of extracting data from encrypted or locked phones and allows for a very fast collection of evidence.[168] *UFED* seems to be widely used by national police authorities for mobile forensic analysis of smartphones and tablets, however, many other commercial software exist. For instance, *GrayKey* is an alternative used in Canada and *X-Ways Forensics* in Germany. In Italy, law enforcement authorities use a wide range of malware (malicious software) which may be based on AI applications, mainly to intercept images, conversations, screenshots, and other row data.[169] Furthermore, national police authorities also use different AI tools – like *MERCURE* in France – to manage and analyze telecom data. In addition, multiple systems of domains and vehicle number plate recognition are being operated through optical character recognition (OCR) in several countries.

At another stage of investigation, automated databases are used to support serial crime analysis. In France, for instance, the national police and gendarmerie deploys since 2003 a serial analysis software imported from Canada, called SALVAC (Système d'Analyse des Liens de la Violence Associée aux Crimes). This software allows the matching of violent criminal cases (homicides, rapes, sexual assaults, and attempts). It stores the identity of many thousands of persons, who have been convicted or cited in a procedure, on the orders of the public prosecutor. It also integrates data related to alleged criminal disappearances, discoveries of unknown bodies, as well as various other data (modus operandi, time of occurrence, victim's habits, and words spoken by the perpetrators). The law enforcement authorities of neighboring countries also use SALVAC (Belgium, Germany, Switzerland, and the UK), which facilitates cross-checking of information. Although SALVAC was not originally based on AI, new developments will introduce AI into the system to improve its efficiency. The same is true for several other serial analysis software used in France.[170] French law enforcement authorities also deployed a similar tool, ANACRIM, since 2011 to analyze data and compare information on the modus operandi of crime. Just like SALVAC, ANACRIM will be progressively augmented with AI.[171]

Similarly, in Russia, the *Maniac system* is used in the investigation of serial murders on sexual grounds. Additionally, the project *FORVER* of the Nizhny Novgorod University

[167] Report on the Netherlands, this volume, p. 318.
[168] See for instance the report on Argentina, <https://www.penal.org/de/2023-2>, A-17, p. 10.
[169] Report on Italy, <https://www.penal.org/de/2023-2>, A-07, p. 23.
[170] Report on France, <https://www.penal.org/de/2023-2>, A-10, p. 56.
[171] Report on France, <https://www.penal.org/de/2023-2>, A-10, p. 58.

provides investigators with a system that helps identify murders. It provides several versions of the crime ranked by probability. Based on the data obtained, the investigator instructs the operational staff to search for persons endowed with specific characteristics defined by the program: gender, age, occupation, remoteness of the criminal's place of residence from the crime scene, and the nature of the relationship with the victim.[172]

Further, Law enforcement authorities in Canada and Germany are reported to use AI technology to detect child sexual abuse material out of huge amounts of electronic information, either device- or web-based. In 2004, Public Safety Canada, which includes all national security departments, established the National Strategy for the Protection of Children from Sexual Exploitation on the Internet. Part of this strategy is the project *Arachnid*, managed by the Canadian Centre for Child Protection (CCPC), which is described as 'a web bot that detects and processes tens of thousands of images per second and sends takedown notices of sexual abuse material to web service providers worldwide'. To accomplish this, *Arachnid* uses Microsoft's PhotoDNA technology, and refers to a database of digital fingerprints. The fingerprints are associated with each prohibited photo and were obtained from the Royal Canadian Mounted Police and Interpol.[173] The Sûreté du Québec and other police forces in Canada use a substantially similar spyware tool, called the *Child Protection System*, which was developed by the Child Rescue Coalition (CRC), a US non-profit organization. This system does not employ AI technology yet, but researchers are studying this possibility. [174] The German report highlights that the ZAC-AIRA tool (rapid assessment through AI) merely aims at filtering evidence. It is designed to prefer false positives to false negatives, and further verification is needed before its results are considered as evidence to be submitted to courts. [175]

Finally, the public institutions of several countries use AI tools to detect different kinds of fraud and to further investigate them. In Poland, as an example, in response to the inquiry presented for this survey report, the National Revenue Administration has confirmed the use of machine learning and deep learning to develop an analytical system that identifies tax fraud.[176] Similar developments are observable in other countries, such as in the Netherlands where many municipalities use SyRI (Systeem Risico Indicatie), a policing tool to detect various forms of fraud, including social benefits, allowances, and tax fraud.[177]

4.1.2 Legislative framework

The first important question is whether the use of AI is allowed at all to gather criminal evidence. Since AI is a nascent topic in criminal procedural law, the adoption of fully-fledged regulations would not have been possible yet. Still, even in countries where AI

[172] Report on Russia, <https://www.penal.org/de/2023-2>, A-07, p. 20.
[173] Report on Canada, <https://www.penal.org/de/2023-2>, A-03, p. 77-78.
[174] Report on Canada, <https://www.penal.org/de/2023-2>, A-03, p. 78-79.
[175] Report on Germany, <https://www.penal.org/de/2023-2>, A-02, p. 41-42.
[176] Report on Poland, <https://www.penal.org/de/2023-2>, A-20, p. 12-14.
[177] Report on the Netherlands, , <https://www.penal.org/de/2023-2>, A-04, p. 13.

tools have been used for several years, it seems that the general trend is the absence of specific AI-related rules. Therefore, the national rapporteurs examine the possibility of filling the arising legislative gaps through the existing national regulatory framework. EU countries are further bound by the EU regulatory framework and thus rely to a great extent on European legal initiatives. Another trend is the preference for soft-law instruments such as self-regulatory schemes and private internal guidelines. In Canada, police agencies may establish an internal review board to conduct a risk audit of the use of new technology for evidence-gathering purposes – that was the case for GrayKey. There are also a number of principled guidelines from civil society actors that can be used to guide investigative and evidence-gathering practices by Canadian police forces.

Legislation is slowly making its way, though. In France, after serial crime analysis software has been deployed outside any legal framework, legal and decretal provisions have been adopted to authorize and regulate the use of serial crime analysis software, especially under data protection considerations.[178] Besides this, Art. 230-1 to 230-5 of the Code of Criminal Procedure provides the legal framework for extracting and decrypting data from a mobile device. Judicial authorization is needed to proceed to the extraction and decryption of the data. However, the use of AI tools to manage and analyze them is not referred to in the law.

Some regulation has appeared in Poland too.[179] Since 2018, the Tax Ordinance and the National Revenue Administration Act have been gradually updating the legal basis for data-gathering and its automatic processing for identification tax fraud by fiscal agencies. For example, the Tax Ordinance has introduced the *Clearing House's Information and Communications Technology System* as part of the effort to close loopholes in the tax system, enabling the National Revenue Administration (NRA) to daily gather and analyze data from all bank accounts. Two important provisions have been introduced to the NRA Act. First, the Head of the NRA has been authorized to perform analytical and reporting tasks, which include processing data gathered based on the NRA Act and Tax Ordinance. Second, the NRA bodies have been authorized to conduct automatic data processing and automatic decision-making including profiling, which produces legal effects on the individual, when undertaking analytical, forecasting, research activities, and risk analysis regarding fiscal areas. Nevertheless, the data-gathering processes covered by the law, are rather generic and blurred when one asks what data are processed, for which purposes, and in which circumstances. Regrettably, the legal framework is silent on the question of whether such 'analytical, forecasting, research activities, and risk analysis' may turn into evidence in a criminal trial or to what extent it may be used against the defendant.

In the current Dutch legal framework, there are no provisions regulating the use of Hansken. The Netherlands plans to introduce a specific provision concerning open-source intelligence, which commonly employs AI systems in the upcoming modernization of the

[178] Report on France, <https://www.penal.org/de/2023-2>, A-10, p. 61-63.
[179] Report on Poland, <https://www.penal.org/de/2023-2>, A-20, p. 12-14.

Code of Criminal Procedure (CCP).[180] In the latest development of the draft CCP, mention is made of a special 'technical tool' assisting the investigatory judge in their task to sift the data protected by the legal professional privilege (LPP) out of the data set relevant for the criminal investigation. In the Explanatory Memorandum to the draft CCP, it is stated that the tool will enable the sifting of LPP data whereas the person conducting the sifting would not gain any knowledge of these data. This would allow the investigating officer to conduct the sifting, instead of the investigatory judge, who is the only authority that may gain knowledge of LPP data.[181] Besides, in Spain, the Draft Bill on Digital Efficiency Measures includes some AI-related norms too.[182]

In the majority of countries participating in the survey, there are neither legal rules nor projects of legal rules to regulate the use of AI in evidence gathering. Still, general principles of law are applicable. In Canada, for instance, section 8 of the Charter of Rights and Liberties, which protects privacy, must be taken into consideration. When police officers want to obtain a seizure warrant issued to retrieve data primarily from automated or AI-based software, they must demonstrate that they have 'reasonable grounds to believe that the data they might discover will contain the things they are looking for'. Moreover, they must make the judge aware of the type of technology that will be used to extract the evidence so that the warrant clearly sets out the appropriate limits on this form of evidence collection.[183]

An example of a country where the principle of legality of evidence applies – and regulation for evidence gathering is needed, at least in principle – is Finland. As enshrined in section 2(3) of the Finnish Constitution (731/1999), the principle of legality of evidence requires that the exercise of public powers shall be based on an Act. In the absence of a parliamentary act granting law enforcement authorities the power to use AI systems in evidence gathering, the use of such systems should be illegal. However, the Finnish normative framework also acknowledges the principle of technological neutrality, which finally annihilates the principle of legality. The Finnish report describes that to conduct criminal investigations, law enforcement authorities must consider different acts, including the Coercive Measures Act. Since this act does not explicitly prohibit the use of AI systems – like any analytical methods or tools –, given the principle of technological neutrality, it should not be interpreted as precluding the use of AI-based software. Rather, institutionalized legal principles such as proportionality, minimum intervention, sensitivity, and provisions safeguarding legal privileges limit certain methods, tools, or means subject to a case-by-case analysis.[184]

Like in many other countries, in Portugal, under Art. 125 of the Code of Criminal Procedure, all forms of evidence that are not expressively forbidden by law, are admissible.

[180] See Chapter 7, Article 2.8.8 of the draft Code of Criminal Procedure.
[181] Report on the Netherlands, in this volume, p. 320.
[182] Report on Spain, <https://www.penal.org/de/2023-2>, A-05, p. 17
[183] Report on Canada, <https://www.penal.org/de/2023-2>, A-01, p. 83.
[184] Report on Finland, in this volume, p. 292-293.

Limitations to this principle exist in case of evidence obtained by torture, unlawful coercion, and infringement of personal physical or moral integrity. There is no specific normative framework concerning evidence gathering through AI in Portugal – AI tools do not seem to be used at the moment in Portugal anyway. As mentioned before, the Portuguese Charter on Human Rights in the Digital Age, adopted in May 2021, establishes a set of innovative standards regulating the digital environment and the provision of new rights and duties. For instance, Article 17 of the Charter protects against abusive geolocation. As Article 9, which specifically concerns AI technology, does not address the question of obtaining criminal evidence through AI, this process must be considered admissible.[185]

Besides the question of the lawfulness of gathering evidence with the help of AI, the second substantial problem discussed in the national reports concerns the feature of defendants' rights when law enforcement authorities deploy AI tools. Again, in the absence of specific sets of rules, it is necessary to interpret existing national laws and commonly acknowledged principles, such as the adversarial character of the criminal proceedings and the equality of arms, in a way to apply them at best in the AI criminal justice era. The common denominator of this (temporary?) solution is a certain lack of enforceability and materialization of the defendant's interest to be informed as to the use of an AI system and consequently challenge AI gathered evidence. Such is the case among others in Greece, Finland, Turkey, and the Netherlands, where in general the accused has the right to know the background of the investigation, and to access the case files that may include information about an AI system used in the evidence gathering stage and to challenge the evidence obtained in breach of fundamental guarantees.

In Finland, the *audiatur et altera pars* principle (right to be heard) guarantees the defendant the possibility to present evidence, as well as to challenge and comment on evidence submitted by other parties, which further necessitates access to such evidence. Yet, even law enforcement authorities may not have full access to information when using proprietary AI tools. In this case, such information remains practically unavailable to the defendant, as well. Interestingly, according to Finnish commentators, the principle of equality of arms requires that the defendant is granted access not only to prosecution evidence but also to material that has not been named as evidence by the prosecution. It may deal with information that has surfaced during the investigation and is supporting the defense's position. Therefore, the defendant should be able to gather evidence from the same 'pool of potential evidence' that the criminal justice authorities have access to, including sources that have been left out of the official police protocol.[186]

In Turkey, the general principles of adversarial jurisdiction and equality of arms require that in case AI systems are used during criminal proceedings, their results must be shared with the parties. These principles enable the parties to have the right to access the

[185] Report on Portugal, <https://www.penal.org/de/2023-2>, A-11, p. 23-24.
[186] Report on Finland, in this volume, p. 295.

file (CPC, Criminal Procedures Code, art 153) and examine the information and documents that form the basis of the accusation. Furthermore, the defendant and his counsel must have the right to challenge the presented evidence (CPC art 215). Particularly interesting is how Turkey could fill the regulatory gap about AI evidence. According to the national report, in the absence of a regulation that foresees exceptions for evidence gathered through AI systems, 'these systems and their conclusions must be shared with the victim and defendant, and the defense must be given the right to object.' Moreover, if the operation of these technologies requires technical knowledge or if there is doubt about their integrity, according to Article 63 of the Code of Criminal Procedure, they should be examined by an expert. Further, under Article 63/5, the parties have the right to object to the expert examinations and to request a new expert opinion. The rationale behind these rules is that the defendant must be allowed to understand the basis of the allegations against him and challenge the given results. [187]

4.1.3 Relevant Case Law

As the use of AI in the field of evidence gathering is still in its infancy, there are not many cases adjudicated already in national courts, with the striking exception of the Netherlands. Since 2016, a surge of Dutch court cases concerning cryptophones – phones that use encryption for anonymous communication –, in which the Hansken system has been used to gather evidence from huge digital data sets, has appeared. In 2016, a whole server was seized by the Dutch police to access the content of encrypted communications ('Ennetcom cases') and in 2020, the EncroChat cryptophones of more than 30.000 users were hacked by the French police, acting in cooperation with the Dutch police ('EncroChat cases'). Dutch courts are generally rather reluctant to request information on the functioning of Hansken from the NFI or to provide such information to the defense. They quickly rejected motions from the defense questioning the reliability of the functioning of Hansken and the evidence gathered through it. In general, Dutch judges seem to consider that the functioning of this AI system is completely unproblematic. For instance, the Amsterdam court stated in a 2018 judgment, that Hansken was merely used to view (not even to gather) the evidence already collected so that no specific legal basis is needed for its use.[188] Judges also seem to have a largely uncritical belief in the proper functioning of Hansken, perhaps related to the fact that the system has been developed 'in-house' (by the NFI itself), rather than by a private actor with commercial interests in mind. This 'presumed correctness' can be seen in a judgment by the Gelderland court, which ruled with very brief reasoning that the incompleteness of the results due to a software update, had no bearing on the integrity of the results and that the defense did not manage to prove otherwise.[189]

[187] Report on Turkey, <https://www.penal.org/de/2023-2>, A-16, p. 17.
[188] District Court of Amsterdam, judgment of 19 April 2018, ECLI:NL:RBAMS:2018:2504 (case nr. 13/997097-16), para. 7.3.
[189] District Court of Gelderland, judgment of 26 June 2019, ECLI:NL:RBGEL:2019:2833 (case nr. 05/780092-17), p. 9.

In the Netherlands again, the Hague District Court ruled in 2020 that the SyRI tool did not comply with Article 8 of the European Convention on Human Rights (ECHR).[190] According to the Dutch court, 'the application of SyRI is insufficiently transparent and verifiable. As such, the SyRI legislation is unlawful, because it violates higher law and, as a result, has been declared as having no binding effect.[191]

4.1.4 Academic debate and literature

The concerns around the use of AI systems to facilitate the gathering of evidence in criminal proceedings are relatively similar across the participating countries. There seems to be an overall consensus among scholars – irrespective of being in favor or against the use of AI evidence – that whatever use law enforcement and judicial authorities make of AI, this should happen only within a legislative framework. The report on Turkey underlines the need for legislative action to establish specific AI-related sub-rights stemming from the right to a fair trial and transparency obligations as to the data used and processed. In addition, doctrinal discussions propose that AI systems be used in criminal justice only for suspicions of crimes of a certain severity, and they must be exposed to regular auditing.[192] In Portugal, like in many other countries, scholars raise issues about privacy and the defendant's rights. Given the possible ramifications such systems could have on the right of the defendant to effectively challenge the evidence against them, it is proposed to use only explainable and transparent AI systems in this context.[193]

4.2 Evidence generated through AI

4.2.1 National Practices

AI technology does not only facilitate access to evidence but also contributes to producing new types of probabilistic information, such as the outcomes from facial and voice recognition systems. The law enforcement authorities of many different countries have started using facial recognition systems. [194] The report on Greece offers a comprehensive definition of how facial recognition works. The biometric data to be extracted, *inter alia*, by a digital photograph are compared with the data available in other databases employed by law enforcement authorities. This comparison leads to the so-called 'matching', which is followed by the calculation of the similarity score.[195] The limits against

[190] Right to respect for private and family life, home and correspondence. Triggering event was the 2020 Child Benefits System scandal, in which approximately 26,000 families were wrongly accused of social benefits fraud by Dutch tax authorities, which led to the then government's resignation. 'Dutch Childcare Benefit Scandal an Urgent Wake-up Call to Ban Racist Algorithms' (Amnesty International Netherlands, 25 October 2021) www.amnesty.org/en/latest/news/2021/10/xenophobic-machines-dutch-child-benefit-scandal/.

[191] The Hague District Court, judgment of 5 February 2020, ECLI:NL:RBDHA:2020:1878 (case nr. C-09-550982-HA ZA 18-388).

[192] Report on Turkey, <https://www.penal.org/de/2023-2>, A-16, p. 17.

[193] Report on Portugal, <https://www.penal.org/de/2023-2>, A-11, p. 23.

[194] Argentina, Canada, France, Italy, the Netherlands.

[195] Report on Greece, <https://www.penal.org/de/2023-2>, A-06, p. 8.

which the similarity score is compared are set by individuals and, thus, determined – at least to a certain degree – based on subjective criteria. In the case of a high limit, false positives will appear, while in the case of a low limit, there will be false negatives. At a practical level, the Hellenic Police is expected to acquire smart policing devices that will enable, *inter alia*, facial recognition for identifying individuals during on-site controls.[196]

In the Netherlands, facial recognition technology is increasingly used in public spaces, both by the police and municipalities, often in public-private partnerships constituted within smart city initiatives.[197] The Dutch police use the Central Automatic Technology for Recognition (CATCH), which compares images taken from a photo or a video with a large police database of current or past suspects and convicted persons. For the time being, the Dutch police do not deploy real-time facial recognition in public areas.[198] In France, police officers are allowed to take biometric pictures of suspected persons for investigative purposes, under the conditions of Art. 55-1 the Code of Criminal Procedure.[199] The facial recognition process works based on the *Traitement des antécédents judiciaires*, a large police database comprising not only convictions but also various information about the former implication of persons into a criminal case. In contrast to the Netherlands, facial recognition in public areas is not allowed in France. The law of 19 May 2023 has exceptionally authorized the use of AI applications to detect suspicious behaviors in the context of the Olympic and Paralympics games of Paris in 2024, however, biometric identification of persons remains excluded.[200]

In Finland, data processing acts permit law enforcement authorities to use facial recognition technology to prevent, detect, and investigate criminal offenses. Subsequently, a specific automated facial recognition system (KASTU) was developed and the police began to use it in May 2020.[201] KATSU should only provide indices to direct investigations, not produce evidence itself – as in Germany, where biometric identification could not be used as evidence in a criminal trial, but only as an indication to apprehend the suspect and identify them.[202] In September 2021, the controversial use of the 'Clearview AI' facial recognition application caused the issuing of a reprimand by the Finnish Data Protection Ombudsman. The National Bureau of Investigation had processed personal data in violation of the Act on the Processing of Personal Data in Criminal Matters.[203]

---

[196] Report on Greece, <https://www.penal.org/de/2023-2>, A-06, p. 17.

[197] See e.g., T van Arman, 'Smart Cameras for a Smart City' amsterdamsmartcity.com/updates/news/smart-cameras-for-a-smart-city.

[198] Report on the Netherlands, in this volume, p. 325.

[199] Report on France, <https://www.penal.org/de/2023-2>, A-10, p.71.

[200] Report on France, <https://www.penal.org/de/2023-2>, A-10, p. 16-17.

[201] Report on Finland, in this volume, p. 296.

[202] Report on Germany, <https://www.penal.org/de/2023-2>, A-02, p. 46.

[203] Data Protection Ombudsman, Decision, 20 September 2021, 3394/171/21. Further, the NBI were ordered to request the service provider to delete any personal data relayed to it by the NBI through the use of the Clearview AI software. See 'Police reprimand from Deputy Data Protection Ombudsman – police have initiated measures ordered' (28 September 2021) poliisi.fi/en/-/police-reprimand-from-deputy-data-protection-ombudsman-police-have-initiated-measures-ordered.

In 2017, the Italian Scientific Police Department acquired SARI, 'Automatic Image Recognition System', an automated face-based human recognition software. Among others, it is used for the investigation, detection, and prosecution of criminal offenses. The software has two operating functions. The 'Enterprise' function allows operators to search for the identity of a face using one or more facial recognition algorithms within a large database. The 'Real-Time' modality makes it possible to analyze the faces of subjects captured by cameras, comparing them with a wide-size watch list, nevertheless, it is not used by the Italian Police since the negative opinion of the Italian DPA.[204]

Regarding voice recognition, the Italian Scientific Police Department uses automated and semi-automated systems, allowing a faster analysis of the physical characteristics of the voiceprint. The technical analysis is divided into three operational phases. First, the operator must choose the phonic material. Secondly, special programs – IDEM and SMART, in Italy – isolate some parameters for the characterization of the voice. The final phase consists of a statistical interpretation of the data and a comparison made between the measurements obtained, to establish the compatibility between the anonymous voice and that of the known subject. Italian law enforcement authorities are collaborating in the Interpol 'Speaker Identification Integrated project' to create a large database of voice tracks.[205] The law enforcement authorities of other countries, including France,[206] also deploy voice recognition.

A third important illustration of evidence generated through AI consists of the technology of probabilistic genotyping. In Canada, the STRmix™ software application is the first commonly used AI tool for producing evidence, in the context of DNA mixture analysis. The Centre of Forensic Sciences of Ontario (since 2016), the *Laboratoire de Sciences Judiciaires et de Médecine Légale* du Québec (since 2018), and the British Columbia Institute of Technology (since 2018) are all said to be using this specialized software application. Experts generally operate STRmix™ when the DNA samples that are at their disposal are of poor quality and do not permit a 'match' after traditional genotyping. With probabilistic genotyping, they analyze a DNA mixture composed of several DNA fragments of the suspected person, with statistical methods and algorithms. To help identify this person, the tool evaluates different hypotheses that an expert has previously selected. For instance, one hypothesis might be that a certain person A is the suspect, and the other hypothesis is that A is not the suspect. The tool calculates which one of the hypotheses is more likely to be true – which does not exclude that both hypotheses are wrong. The expert might rather ask whether the suspected person is more likely to be A, B, or C. Even if the probabilistic genotyping tool designates B with the highest score, this does not mean that the offender is not D, who was not included in the expert's question. This is why probabilities provided by STRmix™ must be considered with much caution.[207]

[204] Report on Italy, <https://www.penal.org/de/2023-2>, A-01, p. 24.
[205] Report on Italy, <https://www.penal.org/de/2023-2>, A-01, p. 24.
[206] Report on France, <https://www.penal.org/de/2023-2>, A-10, p. 72-73.
[207] Report on Canada, <https://www.penal.org/de/2023-2>, A-03, p. 85-86.

Several national reports finally acknowledge that AI-generated evidence that is not proffered by law enforcement authorities – like for instance data from the drowsiness detection system of a vehicle –interferes more and more with the curse of criminal justice.

4.2.2 Nature and classification of AI-produced evidence

The question of whether AI-produced evidence constitutes a new means of evidence or belongs to one of the traditional legal categories was answered differently in the national reports. Some countries have not yet addressed this issue[208] while others could apply their general regime. [209] In China, some scholars have pointed out that big data-driven evidence, which is similar to electronic evidence can be examined according to electronic evidence examination rules and methods. Nevertheless, the uniqueness of AI-enabled evidence that concludes machine analysis requires a new examination system.[210]

The German report usefully reminds us that the electronic data provided by an AI system is not evidence by itself but must be transformed before the court can assess it as a piece of evidence. In Germany, it may be a transformation in documentary evidence, real – or material – evidence, or a testimony – or expert testimony. Its validity and significance may be questioned in court by all participants in the trial.[211]

In Canada, while digital information is usually considered documentary evidence, recent developments in common law seem to acknowledge 'electronic information recorded automatically without human intervention' as real evidence. However, because of the controversial nature of AI tools, their potential bias, their opacity, and the high level of expertise required to assess the reliability of their outcomes, Canadian authors urge to ensure that AI evidence be systematically introduced into court through expert testimony, to determine its reliability as a piece of evidence. Following this literature, it would be wise to acknowledge probabilistic genotyping as special expert testimony.[212]

The report on Turkey mentions that risk assessment profiling technologies show similarities with 'personality testimony' because they refer to the character of the suspect. According to the Turkish Criminal Procedure Code, such a testimony shall not be used to prove the material event. Moreover, a Turkish doctrinal approach considers AI systems as anonymous witnesses if the information about the software is not shared by claiming intellectual property rights, or if self-learning algorithms are used, and consequently, the results of this software are difficult to be explained by humans. Otherwise, AI systems can be considered as providing electronic (digital) evidence, which does not constitute a separate type of evidence. Since the evaluation and analysis of such evidence require

[208] France, Greece, Italy, Portugal, Spain.

[209] In Finland, for example, all five categories of evidence may be used to relay AI-produced information to the court. Report on Finland, this volume, p. 299. The same seems to be true in Germany and Turkey, and probably in many other countries.

[210] M Guoyang: On the Examination of AI-driven Evidence in Criminal Procedure, published in Criminal Science, Issue No. 5, 2021.

[211] Report on Germany, <https://www.penal.org/de/2023-2>, A-02, p. 45.

[212] Report on Canada, <https://www.penal.org/de/2023-2>, A-03, p. 88.

technical knowledge, an expert opinion must be sought, and the digital evidence must be explained using scientific methods.[213]

4.2.3 Legislative framework

While the production of evidence through AI may offer great opportunities to criminal justice systems, it also encompasses dangers. First, AI-generated evidence may not be as reliable as its scientific origin suggests, as the problem of false positives and false negatives shows. Second, AI technology does not secure the absence of bias and, third, it provides for results that always present themselves in the form of probabilities. It is not new that evidence proffered to criminal courts is neither perfections nor certitudes. However, in the case of AI, the risk of an overreliance of judges upon science is real. [214] This is why changes in the law of evidence could be expected. For instance, new rules could be adopted regarding the admissibility of AI-generated evidence and the verification of the reliability of its outcomes. It could also be appropriate to discuss the legal recognition of exclusionary rules specifically fitted for 'AI going wrong' risks. In Canada, the *Commission du droit de l'Ontario* has even called for a reform of the rules of evidence, in particular by *including a presumption of inadmissibility* of evidence generated by an AI tool.[215]

The general trend among the participating countries, nevertheless, is again the absence, at least for the time being, of specific regulations on AI-related evidence and the reliance upon the existing legal framework.[216] In most countries, there is not an exhaustive list of admissible means of evidence or admissibility rules. The prevailing system is the admissibility of any kind of evidence – including AI-produced evidence – and their free assessment by the judge.[217] Exclusionary rules designed for example to avoid the outcomes of unreliable AI systems being proffered as evidence in courts do not seem to exist.

However, traditional evidence rules come into consideration. In Italy, Art. 189 of the Code of Criminal Procedure deploys a 'useful test', measuring the demonstrative potential of a proffered evidence. In advocating such demonstrative potential about automated generated evidence, parties must elaborate upon the transparency and the explainability of the automated process that generated the information that they want to use as evidence. Thus, an adversarial debate can arise between defense and prosecution.[218] Similarly, in Canada, when expert testimony is based on new science or used for new purposes in Canada, the party wishing to present such testimony must demonstrate by a balance of probabilities its scientific and legal 'reliability'. In keeping with her role as

[213] Report on Turkey, <https://www.penal.org/de/2023-2>, A-16, p. 24.
[214] Eftychia Bampasika, Artificial intelligence as Evidence in Criminal Trial, WAIEL, September 3, 2020, <https://www-ceus.ws.org>, Vol. 28-44, Ethics7, consulted 30.11.2023.
[215] Report on Canada, <https://www.penal.org/de/2023-2>, A-03, p. 87.
[216] In Italy, however, according to the report of Italy, the Italian parliament adopted in in law no 205/2021 a controversial regulation of the use of automated facial recognition systems by law enforcement authorities, see report on Italy, <https://www.penal.org/de/2023-2>, A-01, p. 24.
[217] Argentina, Finland, France, Greece, Italy, the Netherlands, Portugal, Turkey.
[218] Report on Italy, <https://www.penal.org/de/2023-2>, A-01, p. 25.

gatekeeper, the judge must assess the admissibility of expert evidence involving new science against the criteria set out in R. v. Mohan: (1) whether the theory or technique can be and has been tested, (2) whether the theory or technique has been peer-reviewed and published, (3) whether there is a known or potential rate of error or whether standards exist and (4), finally, whether the theory or technique used is generally accepted.[219] Finally, in the Netherlands, as to the reliability of evidence, Article 359(2) of the Code of Criminal Procedure states that when the prosecution or the defense argues that evidence submitted by the other party is unreliable, they can enter a 'plea against the use of unreliable evidence.'[220]

4.2.4 Soft Law

In the absence of specific legislation as to AI-generated evidence, soft law occupies the room and plays a role in ensuring the robustness of AI-produced evidence. In the Netherlands, for instance, the reliability and neutrality of the CATCH facial recognition system are preserved through the guidelines for the use of the system. These guidelines require a 'double human verification' in the decision-making process. This procedure is designed to reduce the risk of false positives and to protect the rights of data subjects. After the comparison, the AI-generated list of candidates is presented to a trained expert. If the expert believes that there is indeed a match with one of the candidates, the match is shown to two other experts who assess the match independently (it is unknown what kinds of experts are meant here and in which way they are trained). If the experts do not come to the same conclusion, the most conservative conclusion is reported. In Canada, since the production of evidence using AI tools comes from laboratory practice, the use of this new technology is also indirectly regulated through the standardization standards that govern laboratory activities. In Ontario, laboratories must comply with the ISO 17025 standard to be accredited, and in Quebec, the *Laboratoire de sciences judiciaires et de médecine légale* with the ISO 900268, ISO 17025, and CAN-P-1578.[221]

4.2.5 Defendant's rights

Most countries adopt the position that their criminal justice regimes have already the safeguards needed for the effective participation of the defendant in the criminal trial. Many national rapporteurs mention the principle of equality of arms and the right of the defendant to have access to incriminating evidence and to challenge it, to consult an expert, and to bring exonerating evidence to the trial. The academic debate revolves mainly around issues that constitute ultimately different aspects of the procedural 'mother right' to a fair trial. This signifies the relative consensus in the literature about the dangers and challenges AI-produced evidence brings to national legal orders. In practice, concrete difficulties arise, like in Canada, where the operation of the STRmix™ does not appear

[219] Report on Canada, <https://www.penal.org/de/2023-2>, A-03, p. 88.
[220] Report on the Netherlands, in this volume, p. 319.
[221] Report on Canada, <https://www.penal.org/de/2023-2>, A-03, p. 86-87.

to meet any minimum guarantee of transparency. The Ontario Centre of Forensic Sciences, which uses the tool, admits that it does not have access to the source code and does not intend to share information related to its internal validation reviews regularly. Thus, a specific request before a judge must be made each time by the accused.[222]

In Greece many concerns have been expressed about the defendant's rights and the presumption of innocence, the aura of infallibility and objectivity that surrounds AI, and the rising asymmetries, that could de facto lead to a reversal of the burden of proof.[223] The report on Turkey further underlines the importance of a procedural right to challenge the originality and integrity of AI-generated evidence.[224] Besides, the report on Italy highlights that the principle of equality of arms includes the right to 'effectively influence the court's decision'. Thus, the impossibility of assessing the reliability of an automated generated piece of evidence proffered by the prosecution may deprive the defendant of the chance to 'effectively influence the court's decision.' This is why the rapporteur of Italy suggests that, under the principle of the equality of arms, the court discharges such automated calculations. [225] Lastly, the Finnish rapporteur argues for a general 'auxiliary questions' framework to assist triers of fact in assessing electronic evidence. The non-exhaustive list of auxiliary questions could additionally serve as a checklist that may help parties in supporting their evidence and challenging evidence presented by other parties. AI-produced evidence would not simply be presumed reliable and trustworthy, and the presumption of innocence would be guaranteed. By contrast, if AI-produced evidence is presented in support of the innocence of the defendant, however, the requirements of providing supporting information should not be interpreted to be as stringent.[226]

4.2.6 Use of non-investigative authorities' information as evidence

The possibility of using information produced by non-investigative authorities in the criminal justice context becomes a crucial issue. As the rapporteur of Italy highlights, under the current evidentiary law, the daily usage of commercial devices based on AI systems is a main source of information that may be considered as evidence in court.[227] The same is true in most countries, because of the principle according to which all forms of evidence that are not forbidden by law are admissible, even if they are atypical. The exception to this only exists when the atypical evidence implies a significant restriction of fundamental rights. Therefore, the drowsiness detection and distraction warning system embedded in an automated vehicle, for example, could be used as evidence in criminal proceedings, unless such evidence implies a sensitive restriction of fundamental rights. Given the fact that AI-generated evidence produced by an on-investigative structure was not meant to be used in a trial setting, it may lack the safeguards needed. As

[222] Report on Canada, <https://www.penal.org/de/2023-2>, A-03, p. 87.
[223] Report on Greece, <https://www.penal.org/de/2023-2>, A-06, p. 14 and 19.
[224] Report on Turkey, <https://www.penal.org/de/2023-2>, A-16, p. 24. Court of Cassation 1. CD, T. 16.01.2012, E. 2008/10249, K. 2012/48, see Arslan, p. 261.
[225] Report on Italy, <https://www.penal.org/de/2023-2>, A-01, p. 21.
[226] Report on Finland, in this volume, p. 309.
[227] Report on Italy, <https://www.penal.org/de/2023-2>, A-01, p. 25.

long as no law regulates AI-generated evidence, the discussion on potential fundamental rights restrictions will be left to the courts.

Another example of AI evidence produced by non-investigative authorities can be found in the Dutch draft CCP. It introduces a new provision, according to which the public prosecutor may order companies or institutions that can 'reasonably be suspected of having access to certain data' relevant to the investigation to process these data and then submit the result of this processing to law enforcement (Article 2.7.51(1) draft CCP). Google, Facebook, and Apple are examples of companies that may be asked to perform such processing. The main feature of this operation is that it produces 'new' data which are then supplied to the police. To put it differently, the police only receive the results of the data analysis performed by the company that collected the data. Ratio of this provision is the limitation of the amount of data that is provided to law enforcement.[228] Here again, the defendants need specific protection of their rights.

4.2.7 Case law

Even though the majority of countries have not developed important case law yet, there are some exceptions, especially regarding facial recognition. In the Netherlands, so far, the Zeeland-West-Brabant District Court concluded in its 2019 judgment[229] that the results of the CATCH system alone, even after they have been 'confirmed' by two human experts, do not suffice in establishing the link between the suspect and the crime for a criminal conviction. Additional corroborating evidence is necessary and seems to compensate for incertitude on the reliability and neutrality of such systems. Similarly, in Italy, the Supreme Court decided evidence drawn from a facial recognition system cannot be the 'sole or decisive piece of evidence' to apply pre-trial coercive measures.[230] In Turkey, the Constitutional Court decided that the assessment of the reliability of digital evidence requires technical knowledge, thus using solely the law enforcement's report on digital evidence and not sharing it with the defense infringes on the principle of equality of arms and the right to a fair trial.[231]

Finally, In Quebec, Canada, in a case where the STRmix™ had been used to produce one of the forensic biologist's reports, a judge denied a motion for disclosure made by an accused seeking information related to the internal validation process of the STRmix™

[228] Report on the Netherlands, in this volume, p. 320.
[229] District Court of Zeeland-West-Brabant, judgment of 17 May 2019, ECLI:NL:RBZWB:2019:2191 (case nr. 02-665274-18), para. 4.3.
[230] Cass. pen., sez. IV, 18 June 2019, n. 39731; Cass. pen., sez. I, 21 July 2020, n. 21823. Report on the Netherlands, this volume, p. 328.
[231] Report on Turkey, <https://www.penal.org/de/2023-2>, A-16, p. 23. Turkish Constitutional Court, Application No: 2014/253, Decision Date.: 9.1.2015, para. 55, 76, 77, https://www.resmigazete.gov.tr/eskiler/2015/05/20150512-12.pdf, A.D.17.07.2021.

software stating that 'currently, the burden is on the accused to show that there is a reasonable possibility that the information has probative value with respect to an issue or the competency of a witness to testify.'[232]

4.3 Evidence assessment through AI

4.3.1 National practices

In most countries participating in the survey, AI systems are not used for the assessment of evidentiary materials proffered in criminal trials. This finding is explicit in the reports on Argentina, Canada, Finland, France, Germany, Greece, the Netherlands, Poland, Portugal, and Russia. However, the Netherlands seems to find the use of AI systems realistic to detect fake images, videos, or audio files among the evidentiary material. According to the national report, the development of such systems to be used in law enforcement has begun in the Netherlands.[233]

Additionally, when an AI system performs evidence assessment, the latter could further be used as a basis for a claim concerning the defendant's guilt. In general, AI systems for guilt assessment neither are in place nor are likely to appear in the future.

China is an outstanding country regarding evidence assessment and other practices regarding evidence in criminal matters. In China, the judicial authorities are already using many AI tools in the context of evidence, for the guidance of evidence standards, evidence verification, and evidence chain examination. Since 2016, for example, Guizhou Province has formulated the 'evidence standard guidelines' for cases handled by public security organs, procurators, and courts. They use big data to embed element-oriented and structured evidence standards into the case handling system, to promote a more unified use of evidence and prevent wrongful conviction. The *Shanghai intelligent case handling aided system for criminal cases* developed by the Shanghai High People's Court in 2018 functions similarly. [234] The Guizhou Province also uses the *intelligent case research and judgment system*, which among other functions systematically analyzes and weights criminal evidence and the probative force of the evidence chain against the standards involved in the Criminal Procedure Law (刑事诉讼法).[235] Further, the procuration organs of Jiangsu Province have developed a *case management robot*. Through a comparative analysis of the case file and various legal documents of the procuration organs, the robot can check the obtained data and further remind, warn, and evaluate the possible qualitative or evidential problems of the case.[236] The prosecutions' office of Tianjin Municipality has conducted evidence presentation by multimedia through all links of court trials, forming

[232] Report on Canada, <https://www.penal.org/de/2023-2>, A-03, p. 87.
[233] Report on the Netherlands, in this volume, p. 331.
[234] Report on China, in this volume, p. 271.
[235] Report on China, in this volume, p. 262.
[236] Report on China, in this volume, p. 264.

a new mode of multimedia-driven cross-examination evidence in special case handling.[237] Furthermore, the Ziyang Prosecution's office of the Sichuan Province has developed an *integrated platform for court appearances* containing pre-trial preparation, charges during court trial, background support, and multimedia-driven evidence presentation based on electronic files, which endeavors to solve contradictions between different types of evidence and improve the public prosecution in court.[238]

4.3.2 Normative Framework

According to the national reports, no country has addressed this issue through legislation to this moment. This could be traced back to the fact that most of the participating countries have a system of free assessment of evidence. Greek judges, as an example among many others, are not obliged to follow concrete rules on evidence assessment. They shall decide on their conscience voice and be guided by the impartial judgment concerning the factual truth, the credibility of the witnesses, and the value of other evidence, and provide a specific and detailed justification as to the evidence used and the reasoning based on which their judgment has been formed (Art. 177 (1) GrCCP). Under these conditions, it would not be impossible to deploy AI systems to assess evidence. However, it remains questionable whether this would be compliant with different aspects of the right to a fair trial.[239] The report on Italy adds that the duty of judges to deliver a reasoned judgment gains significant importance when it comes to the evaluation of evidence. Judges need to give reasons as to why they followed a recommendation made by an AI system and how they tested their reliability, to justify their decision. They must also explain why inculpatory evidence should prevail over exculpatory ones or vice-versa. In the case of evidence assessment based on machine learning, the black box would hamper such justification. [240]

It is additionally worth noting that Portugal referred in its report to the European Commission's *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence* (AI Act), which does not explicitly prohibit any uses of AI in the judiciary. However, 'AI systems intended to be used by law enforcement authorities for evaluation of the reliability of evidence in the course of investigation or prosecution of criminal offenses' (Annex III, point 6) are classified as high-risk.[241]

4.3.3 Discussion in the Chinese literature

In China, substantial literature has grappled with various issues of AI being used at the trial stage. While some authors point out the downsides of such use and propose specific safeguards to completely dismiss it, other scholars support AI-enabled evidence. However, they call for the application of guarantees, either through the so-called auxiliary

[237] Report on China, in this volume, p. 264.
[238] Report on China, in this volume, p. 265.
[239] Report on Greece, <https://www.penal.org/de/2023-2>, A-06, p. 19.
[240] Report on Italy, <https://www.penal.org/de/2023-2>, A-01, p. 26.
[241] Report on Portugal, <https://www.penal.org/de/2023-2>, A-11, p. 14.

principle,[242] limitation principle,[243] and rebuttable principle[244] or through the establishment of unified and electronic evidence standards.[245] Others have put forward handling suggestions against the weakening of rational factors in evidence judgment due to AI and the hidden worries of case-handling personnel suffering from case-handling inertia and path dependence.[246] Further, to effectively avoid the legitimacy risk caused by AI technology in the criminal trial field, it is proposed to establish a concept of power regulation and regulate the intelligent case handling system from three aspects: the application mechanism (automatic judgment rendering), the participation mechanism (equalization of the defense) and the research and development mechanism (reliable decision-making), to protect the right of the accused to effectively participate in the intelligent system.[247]

With regards to evidence validity, some believe that AI cannot conduct substantive examination, but only formal examination, such as whether the interrogation meets the procedural requirements; in terms of probative force, AI cannot function independently, and may play an auxiliary and reference role in examining the authenticity of evidence; and in terms of standard of proof, the role of AI is not to judge the standard of proof regarding evidence specification and analysis, but is only an auxiliary means for judges to judge the standard of proof.[248] In terms of the data, the defense lawyer of the accused can request to view, modify, correct, and interpret the data related to their rights and interests in the intelligent system.[249] In general, it is underlined that the integration of AI evidence standards should be moderate rather than absolute and the legal problems must not be completely trusted to the algorithm, which would lead to the weakening or even elimination of factors such as human rationality and goodness in judicial case handling.

At the same time, part of the literature is against the use of AI in this stage, in the evaluation of evidence and judgment rendering, since the substantiation of court trial requires judges to form an inner conviction during the court trial following the principles of directness and verbalism in the court trial, so that 'the investigation of factual evidence is

[242] According to the auxiliary principle, AI can only play an auxiliary role in evidence judgment. It cannot replace the judge's examination and assessment of evidence.
[243] According to the limitation principle, when AI is used for evidence judgment, this must be limited to specific aspects, and not all evidence assessment can be made by AI.
[244] According to the rebuttable principle, when AI is used in one aspect of evidence assessment, it must be clear that the calculation results of AI are not 'absolutely accurate', but refutable and revocable. Not only can judicial personnel directly abandon the calculation results of AI with justified reasons, the party concerned may also raise an objection to the AI calculation results and ask the judicial organ not to consider unreasonable calculation results.
[245] Z Weimin, 'Some Thoughts on the Application Prospect of Legal Artificial Intelligence in China', published in *Tsinghua Law Journal*, Issue No. 2, 2018.
[246] Report on China, in this volume, p. 269.
[247] Report on China, in this volume, p. 276.
[248] Report on China, in this volume, p. 265.
[249] W Chenshu, 'Power Logic of Artificial Intelligence in Criminal Trial', published in *Journal of Xi'an Jiaotong University* (SOCIAL SCIENCE), Issue No. 1, 2021.

conducted in the court and the judgment results are formed in the court.' [250] Following, it is considered that the use of AI in this stage of the court trial process is bound to affect the judge's hearing and judgment of evidence, damaging the authority and seriousness of the court trial. Therefore, AI can be used as an aid to supplement knowledge and support calculation, but it should not become a 'vending machine' for judicial decision-making; it should rather turn from 'evidence guidance' to 'evidence assistance.' [251]

[250] Z Weimin, 'Some Thoughts on the Application Prospect of Legal Artificial Intelligence in China', published in *Tsinghua Law Journal*, Issue No. 2, 2018. Z Fuli & Z Haishan, 'Positioning, Prospect and Risk Prevention and Control of Artificial Intelligence Assisted Sentencing in the Era of Big Data', published in *Guangxi Social Sciences*, Issue No. 1, 2019. L Hongyang & L Xianglong, 'Ethical Issues in Intelligent Justice and Their Countermeasures', published in *On Politics and Law*, Issue No. 1, 2021.

[251] X Shu, 'How Can Artificial Intelligence "Unbiasedly" Help Criminal Justice -- From "Evidence Guidance" to "Proof Assistance", published in Science of Law (Journal of Northwest University of Political Science and Law), Issue No. 5, 2020.

NATIONAL REPORTS ON PREDICTIVE POLICING

PREDICTIVE POLICING IN CANADA*

Karim Benyekhlef and Gabriel Lefebvre[**]

Abstract

Canada's report on artificial intelligence and the administration of justice is divided into three parts. In this first part, we present the technological innovations, powered by algorithms, artificial intelligence and facial recognition, used by police forces in Canada. Although the use of these innovations does not appear to be widespread currently, police forces have a clear interest in making greater use of these predictive technologies in the future. Informed by the experience with these technologies in the United States, civil groups and legal researchers have expressed significant resistance to the biases, lack of transparency, and aura of scientificity that characterize these tools. Taking note of this resistance, we also present the most recent normative innovations in Canada as well as the classic principles of law that could frame the use of these technologies.

Background

To provide some background, the Canadian confederation includes ten provinces, each with legislative power in the areas of jurisdiction specified in section 92 of the Constitution Act, 1867. There is also a central government—the federal government —that draws its legislative powers from section 91 of the Constitution Act, 1867. Since 1982, many rights and freedoms have been protected by the Canadian Charter of Rights and Freedoms.

Organization of law enforcement. Canada's national police force is the Royal Canadian Mounted Police (RCMP). Section 92(14) of the Constitution Act, 1867, provides that each province has the exclusive power to make legislation on the 'Administration of Justice in the Province'. This allows each province to constitute a provincial police force, but only three provinces have done so: Ontario (the Ontario Provincial Police—OPP), Québec (the Sûreté du Québec—SQ) and Newfoundland and Labrador (the Royal Newfoundland Constabulary—RNC). These police forces have jurisdiction everywhere in the province concerned, except in municipalities that have constituted their own police forces. The other provinces have not constituted provincial police forces, but are protected by the RCMP, with which they have contracted to provide police services. The RCMP has the jurisdiction to investigate certain matters that are federal in nature, and it offers police services to the provinces without provincial police forces, and also to certain Indigenous communities and the federal territories. As we will see, most large Canadian cities have chosen to constitute their own police services.

Criminal law. In Canada, under section 91(27) of the Constitution Act, 1867, the power to legislate on criminal matters is reserved exclusively for the federal Parliament. This

[*] The second and third part of the Report (on predictive justice and on right of evidence) are available in French on the website of the International Association of Penal Law.

avoids regional compartmentalization. It aims for uniformity and consistency in the fundamental norms that guarantee that public order is maintained in Canada. Unlike in the United States, where there are both federal crimes and state crimes (since each state can pass its own criminal laws), the Criminal Code and the criminal procedure rules are the same throughout Canada.

1 National practices

Our presentation of the AI tools used by the police in Canada requires defining and recontextualizing the approach in terms of 'predictive policing'. Today, the expression 'predictive policing' does not seem to be widely used in official government and police department communications in Canada. In the United States, the National Institute of Justice (NIJ) sponsored a publication that proposed a definition in 2013: 'Predictive policing is the application of analytical techniques—particularly quantitative techniques—to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions.'[1] With regard to Canada, we managed to find an initial definition dating only from 2018 in a summary report by the Department of Justice's Research and Statistics Division (RSD): 'Predictive policing: when law enforcement identifies criminal activity using mathematical, predictive, and analytical techniques.'[2] However, this definition is not that of the government: it comes from an outside specialist mandated by the RSD.

To understand this approach to policing, we need to begin by establishing a definition that does not bring commercial potential into consideration. The expression 'predictive policing' supports the marketing idea that criminal activity could in fact be predicted using algorithmic statistical processing. The appropriate definition should instead convey the idea that 'predictive policing' is limited to statistical processing of quantifiable facts performed using algorithms, and that it provides suggestions about the location, time and persons at risk. The approach is all the more limited by the fact that it looks only at the 'when', 'where' and 'who' aspects of criminal activity. Understanding and anticipating criminal acts always require interpretation and human experience: the 'why' and 'how' are indispensable data for a true understanding of criminal activity. Uncom-

---

** Karim Benyekhlef is Full Professor, Director of the Cyberjustice Laboratory, holder of the LexUM Chair on Legal Information, Centre de recherche en droit public (CRDP), Faculty of Law, Université de Montréal and Gabriel Lefebvre is a researcher and doctoral candidate at Université McGill.

[1] Walter L. Perry et al., 'Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations' (*OJP website*, September 2013), 1 <https://www.ojp.gov/ncjrs/virtual-library/abstracts/predictive-policing-role-crime-forecasting-law-enforcement> accessed April 2022.

[2] Dennis D. Draeger, 'Justice Trends 2: Automated Justice Get the Gist of the future for technology in justice' (*Justice Canada website*, June 2018) <https://www.justice.gc.ca/eng/rp-pr/jr/jt2-tmj2/index.html> accessed February 2023. The author is a representative of *Shaping Tomorrow* – a company that offers research, analysis, strategy and planning services using an AI tool that it claims can 'anticipate trends' for clients in the public and private sectors. See Shaping Tomorrow's website: <https://www.shapingtomorrow.com/webtext/10>.

fortable with the received definition, the Citizen Lab and University of Toronto researchers behind the first Report of Canada on 'predictive policing' (2020) chose to define it outside of its commercial connotation and in a broader manner so as to include the other forms of police surveillance performed using algorithms. They defined 'algorithmic policing' as 'the use of algorithms by police services for the pre-emptive monitoring and forecasting of potential crime before any crime has occurred.'[3]

If we look at it alone and in isolation, it is difficult to understand the turn toward 'predictive policing'. In fact, it is part of a much broader series of police reforms that have been taking place in Canada and the United States since the 1990s: 'predictive policing' is intertwined with 'community policing', 'hot spot policing', 'problem-oriented policing' and 'intelligence-led policing'. According to Bilel Benbouzid, all these reforms share one feature: they seek to [translation] 'make policing more proactive and vigilant, and less reactive and emergency-oriented—policing more engaged in producing security than in repressing criminals'.[4] As we will see, predictive policing in Canada shares many characteristics with the policing approaches that we have mentioned: **(i)** collaboration with other actors in the community, **(ii)** interventions motivated by the analysis and processing of personal information and **(iii)** preventive action based on the risk that there will be a victim. In the United States, the National Institute of Justice (NIJ) also recognizes that police officers' day-to-day work has changed substantially: 'Today more than ever, law enforcement work is also proactive. In proactive policing, law enforcement uses data and analyzes patterns to understand the nature of a problem. Officers devise strategies and tactics to prevent or mitigate future harm.'[5] This broader change in approach toward preventing the risk of crime can also be seen in Canada, and seems to have accelerated after the September 11, 2001 attacks. At the turn of the new millennium, police departments found themselves under pressure from a populace that was demanding more 'results' in terms of improved security as well as tangible proof of those results (police accountability), but the police also found themselves having to operate with ever-shrinking resources in a context of fiscal restraint: 'As a result, Canadian police services are turning to information technologies and innovations as a means "to create smart, efficient processes and . . . to leverage technology to move away from reactive to proactive policing"'.[6] Collecting, aggregating and analysing information (intelligence-led policing) would then

[3] Kate Robertson, Cynthia Khoo and Yolanda Song, *To Surveil and Predict: A Human Rights Analysis of Algorithmic Policing in Canada* (Toronto: Citizen Lab and International Human Rights Program, University of Toronto, 2020), 29 (hereinafter 'Citizen Lab').

[4] Bilel Benbouzid, 'Quand prédire, c'est gérer, La police prédictive aux États-Unis', (2018) 211-5 *Réseaux* 221, 223. See also National Institute of Justice, 'Overview of Predictive Policing,' (*NIJ website*, 2014) <https://nij.ojp.gov/topics/articles/overview-predictive-policing> accessed April 2022; Rich LeCates, 'Intelligence-led Policing: Changing the Face of Crime Prevention,' (*Police Chief online*, October 17 2018) <https://www.policechiefmagazine.org/changing-the-face-crime-prevention/> accessed April 2022.

[5] National Institute of Justice, *supra*, note 4.

[6] Carrie B. Sanders, Crystal Weston and Nicole Schott, 'Police Innovations, "Secret Squirrels" and Accountability: Empirically Studying Intelligence-Led Policing in Canada', (2015) 55-4 *The British Journal of Criminology* 711, 711-712, referring to a quote of a Police Chief during the Ontario Association of Law Enforcement Planners Meetings in 2011.

make it possible for police to 'substantially' modify their approach by reorienting it toward pro-active, targeted surveillance, efficient management of crime risks and preventive strengthening of security.[7] Recourse to AI would fit into this quest for results by offering police intervention quantifiable measurement while at the same time making it possible to save resources.[8] We would like to draw attention here to the fact that there are inherent limits to what can be 'measured' and 'quantified' in terms of 'security production'. Once security is understood through the broader notions of harmony and social peace, it becomes difficult to quantify, and it even seems irreconcilable with the intensified and hyper-targeted pro-active police surveillance and vigilance suggested by the predictive tools in question. Once it is understood in terms of social peace, security cannot be reduced to a rate of return that would flow from law enforcement by the police.

The change in approach toward preventive policing can be seen everywhere in Canada, to begin with in the stated missions of the various police intelligence agencies integrated into police departments.[9] It can also be seen in the adoption of new public security policies focussing on prevention, such as the Departmental Crime Prevention Program (*Politique ministérielle en prévention de la criminalité*) adopted by the SQ in 2001 and its federal equivalent, the National Strategy on Community Safety and Crime Prevention, which have made the police's preventive mission official.[10] Even though these strategies specify that 'preventive' police intervention should be accompanied by non-repressive action, there is reason to fear that these approaches have in the end expanded the criminal realm, bringing police attention to people who are perfectly innocent, but socially vulnerable.

The desire to intervene even before a crime has occurred predates the advisability of using AI to predict that a crime might be committed. As early as 2004, researchers were already on a quest for predictive power spurred on by the feeling that innovation was required to fight the novel forms of criminal activity facilitated by new technologies. Around the same time, the RSD published a summary report in which the author recommended that the federal government fund an integrated multi-sector research group that 'examines and maps crime trends, forecasts future crime rates and patterns, and estimates the impact of crime (i.e., costs) for both the present and the future' to reduce and prevent crime. The research group would bring in players from the private sector, such as computer engineers, telecommunications services and Internet providers.[11] At the turn of the new millennium, the fear that the advent of new technologies would lead to an

---

[7] Ibid.

[8] B. Benbouzid, *supra*, note 4, 240.

[9] For example, the Ontario Provincial Police's Provincial Operations Intelligence Bureau (*OPP website*, 30 June 2016) <http://www.opp.ca/index.php?lng=en&id=115&entryid=576bf77e8f94ace216355e0f> accessed April 2022: 'Their goal is to anticipate, prevent and monitor criminal activity in Ontario. The members of this bureau collect, assess and share intelligence data within the OPP and with other law enforcement agencies.'

[10] Sûreté du Québec, Politique ministérielle en prévention de la criminalité, 2001, 10 and 15.

[11] Stephen Schneider, *Predicting Crime: A Review of the Research* (Summary Report prepared for the Department of Justice Canada, 2002), 30. See also 2–3 for fears expresses concerning the new possibilities that the technologies offer to criminals.

increase in crime led to the demand for greater innovation in the fight against crime. This is when the predictive policing turn was taken in Canada.

In this context, in which law enforcement seems to adopt an approach focussed on intelligence, innovation, prevention and efficient resource management, it is understandable that an AI tool would be an attractive piece of technology for police in charge of public security. There is still clear interest in Canada for the development of AI tools for predictive policing. For example, the recent 2020–2021 Departmental Report contains a new Digital Policing Strategy designed to 'connect' the RCMP. The Strategy provides for massive development of new technologies to prevent crime: 'the future of the RCMP is mobile and online'. One of its objectives is to 'make better use of data to predict, prevent and fight crime'.[12] In fact, police are already using AI tools in Canada, and it is reasonable to think that this will continue and increase.

In the next part, we present the survey of AI tools currently used in Canada that was done by Citizen Lab researchers in 2020, and we complement those findings with our own observations. We will see that the different reasons that justify recourse to these algorithmic prediction and surveillance tools are rooted in a quest for results in the fight against crime and in the view that measuring those results will demonstrate to the population that security has been strengthened efficiently. This quest entails that police need to show that they are innovating, which means developing their technology and intelligence wings, and one of the repercussions of this is that they are changing their approach and seeking to act preventively (*ex ante*), in other words, upstream, in comparison with the traditional (*ex post*) way of fighting crime. In the end, this leads them to intervene in a targeted manner, following a 'risk' analysis, with respect to persons who are 'vulnerable', in the sense that they have been assessed as at risk of being the victim or author of a crime. We argue that this predictive quest on the basis of AI reasoning is a threat to the 'justice' effect initially sought by criminal law (pacification, harmonization, feeling of security and perception of justice 'rendered'); the justice effect can result only from the 'just' application of criminal law, that is to say, state power used sparingly and with restraint, and humane application of the law, which results from interpersonal relations, deliberation and judgment based on human experience, and enforcement that respects the rule of law and our constitutional rights and freedoms. We agree with the observations of French authors Antoine Garapon and Jean Lassègue:

> [translation] When justice is required to manage offences in real time, through a flow and process, and particularly when it seeks to incriminate before an act is committed, as when terrorism is concerned, the principle of presumption of innocence is threatened. There is a risk that fact may conquer the story, and it is law that will be the poorer, and at the same time we will lose our guarantees. . . . The

[12] Royal Canadian Mounted Police, 'The Connected RCMP', (*RCMP website*, 14 october 2020) <https://www.rcmp-grc.gc.ca/en/connected-rcmp?wbdisable=true> accessed April 2022.

coming world is cognitive, not normative, which means that it is a world of facts, not of the idealities that are the foundations of law.[13]

1.1 Will to perform and innovate: a panorama of algorithmic tools

In September 2020, the University of Toronto and Citizen Lab published a landmark survey on Canadian law enforcement's use of algorithmic technologies to predict crime.[14] It was the first and also the most recent exhaustive survey of AI technologies used by the police in Canada. We will present their findings and then provide a panorama of the different surveillance tools that function through the use of an algorithm.

Vancouver – GeoDASH algorithmic policing system (Geodash APS).[15] In 2017, after a 6-month pilot test in 2016, the Vancouver Police Department (VPD) became the first police department in Canada to integrate the use of algorithmic technology into everyday practices in order to guide and coordinate police actions within its jurisdiction.[16] GeoDash APS is an application designed to predict the places and times when property crime is likely to occur.[17] In 2017, the Chief of Police suggested expanding the use of the application to predict car theft and thefts committed using a car.[18] The system aggregates historical data processed according to the type of crime, geographical coordinates, date and time. Processing is done every 24 hours, and it tells the police the 'high risk' areas (detailing the location as precisely as 100 m^2 or 500m^2) according to the time of day (per 2-hour block of time). Patrol officers are then assigned throughout the city of Vancouver according to the predictions in order to prevent crime from actually occurring, by their simple presence, while also carrying out 'proactive' surveillance.[19] GeoDash APS is the product of a public-private partnership between the VPD, the Latitude Geographics/Geocortex company and university researchers. The technology is driven by a will to innovate and get results in the fight against crime, and by the hope to be able to short-circuit criminality. For the Chief of Police, it means developing new, innovative strategies to prevent crime and intervene before it even occurs.[20]

Toronto.[21] Concerning the same family of algorithmic tools, the Toronto Police Service (TPS) has expressed interest in using an algorithmic tool able to identify areas where there is a 'high risk' that a property crime or crime involving a firearm might occur. The tool would also provide a suggestion regarding the number of patrol officers to be deployed in a high-risk area for the next 12 months. The tool results from a partnership that

---

[13] Antoine Garapon and Jean Lassègue, *Justice digitale* (Paris: Presses Universitaire de France 2018) 249.
[14] Citizen Lab.
[15] Ibid., 42–44.
[16] Vancouver Police Department, 'Vancouver Police Adopt New Technology to Predict Property Crime', (*VPD website*, 21 July 2017) <Vancouver Police Adopt New Technology to Predict Property Crime - Vancouver Police Department (vpd.ca)> accessed April 2022.
[17] Citizen Lab, 42.
[18] Vancouver Police Department, *supra*, note 16, video at 10:28.
[19] Citizen Lab, 43 ff.
[20] Vancouver Police Department, *supra*, note 16.
[21] Citizen Lab, 44–45.

began around 2016 between the police and a private firm, Environics Analytics, which offers companies data analysis services. The predictions would take into account various factors, including the crime rate for the preceding year, and the age, income and type of housing of the offenders in each neighbourhood.[22] The need for better public security performance, better management of police resources and better services offered to citizens are the main reasons provided for using the technology.[23]

Edmonton – Community solutions accelerator (CSA). Inspired by the entrepreneurial 'business accelerator' model, the CSA, implemented by the Edmonton Police Service (EPS) in 2020, is a law enforcement innovation laboratory bringing together private and public actors who will develop technological solutions for problems affecting the community in Edmonton. Those behind this initiative include the Edmonton Police Foundation, and private partners such as the University of Alberta, ATB Financial, TELUS and Motorola Solutions Canada.[24] The corporate partners will provide services such as workspace, computer infrastructure and expertise.[25] The laboratory will develop applications

[22] Ibid. Richard Boire, 'Data-Driven Decisions for Law Enforcement in Toronto' (*Machine Learning Times*, 17 august 2018) <Data-Driven Decisions for Law Enforcement in Toronto Machine Learning Times (predictiveanalyticsworld.com)> accessed April 2022.

[23] Citizen Lab, p. 44–45. Environics Analytics, 'Environics Analytics Names Toronto Police Service as Client of the Year' (*EA website*, 19 january 2017): Toronto Police Service is 2016 Client of the Year | News | Environics Analytics <https://environicsanalytics.com/en-ca/resources/media-room/press-releases/2017/01/19/environics-analytics-names-toronto-police-service-as-client-of-the-year> accessed April 2022.

[24] Caley Ramsay and Vinesh Pratap, 'Edmonton police use data, artificial intelligence to combat crime' (*Global News*, 12 February 2020) <https://globalnews.ca/news/6535688/edmonton-police-data-ai-community-solutions-accelerator/> accessed April 2022. In September 2021, the Edmonton Police Foundation and its associates partnered with the Silicon Valley business accelerator Alchemist to launch a new 'social problem management' accelerator: the TELUS *Community Safety & Wellness Accelerator*. See Edmonton Police Foundation, 'Community Solutions Accelerator: A better Alberta With and For Everyone' (*EPF website*) <https://edmontonpolicefoundation.com/csa> accessed April 2022. In the Edmonton Police Foundation's September 23, 2021 press release, examples are given of the kinds of tools that could result: 'Predicting domestic violence earlier, for early intervention; empowering homeless people with tools that predict needs and match solutions; technology-based addiction management/reduction solutions; solving cold cases on missing people; gamified platform to provide racial bias awareness and corrective solutions; proactive mental health and wellness platforms for individuals and businesses/entities; predictive tool to enable law enforcement to help offenders of certain crimes go through rehab instead of putting them through the criminal justice system.' The new accelerator targets 'ventures that apply technology solutions, especially artificial intelligence, machine learning and advanced analytics, to community safety & wellness.' 'Safety challenges' are defined as 'Solutions that increase safety in the community (e.g., theft reduction, improved road safety, food safety, etc.)'. Some of the content of the Press Release appears now on this website, Alberta Innovates, 'Unique business accelerator to grow tech-based ventures that improve community safety and wellness' (*Alberta innovates website*, 23 September 2021) <https://albertainnovates.ca/impact/newsroom/unique-business-accelerator-to-grow-tech-based-ventures-that-improve-community-safety-and-wellness/> accessed February 2023.

[25] Caley Ramsay and Vinesh Pratap, 'Edmonton police use data, artificial intelligence to combat crime', *supra*, note 24.

that can combine data from various sources and might use a machine-learning AI system.[26] The technologies might one day be commercialized. For example, one of the first projects announced—funded by Alcanna Inc.—is to develop technology to prevent theft in liquor stores.

At a press conference in February 2020, the EPS Chief of Police said he also hoped to use the technological innovations to analyse the relationship between criminality and methamphetamine use in order to better target vulnerable people likely to use such drugs. The goal would be to intervene in a preventive manner to refer such people to the healthcare system.[27] The CSA laboratory takes inspiration from the HUB approach, which has been adopted by some provinces and aims to solve social problems such as drug addiction by bringing different organizations together to identify people who are most 'at risk'.[28] The cross-sectoral approach taken by the EPS seems to make it possible to pool information from a variety of sources, such as the healthcare system, social services, child protection services and the police.[29] Although the Chief of Police said that most of the data that will be used by the CSA are already available to these bodies, that the future technologies will be subject to a privacy and impact assessment and that they plan to work with the Privacy Commissioner to regulate their practices,[30] it still remains that this collaboration raises fears relating to exchanges of personal information between bodies.[31]

In this case also, the initiative is motivated by a desire for innovation in the fight against crime. The vulnerability of the persons concerned would be the justification for the project concerning drug abuse.[32] The need for better performance with regard to security is also given as a justification for the approach: reference is made to the need to use limited

[26] Kelly Cryderman, 'Edmonton police create Community Solutions Accelerator with aim to reduce crime' (*Globe and Mail*, 28 February 2020) <https://edmontonjournal.com/news/local-news/edmonton-police-launch-community-solutions-accelerator-using-data-to-reduce-crime> accessed on April 2022.

[27] Edmonton Journal, 'Community Solutions Accelerator to fight crime', (*YouTube*, 11 February 2020) <https://www.youtube.com/watch?v=GqeBnDXR9bI&t=10s> accessed April 2022.

[28] Citizen Lab, 55. For more on the HUB model in general: Public Safety Canada, 'The Hub Model/Situation Table' (*PSC website*, 29 November 2021) <Crime Prevention Inventory (publicsafety.gc.ca> accessed April 2022.

[29] EPS, 'Partnering with technology to fight crime and improve public safety' (*Motorola Solutions website*, 11 February 2020) <https://newsroom.motorolasolutions.com/news/partnering-with-technology-to-fight-crime-and-improve-public-safety.html> accessed April 2022; K. Cryderman, *supra*, note 26.

[30] Anna Junker, 'Edmonton police launch Community Solutions Accelerator, using data to reduce crime' (*Edmonton journal*, February 11, 2020) <https://edmontonjournal.com/news/local-news/edmonton-police-launch-community-solutions-accelerator-using-data-to-reduce-crime> accessed on April 2022. Personal information will be managed by the EPS and the transfer of data to the Edmonton Police Foundation, which is responsible for transmitting the data to the participants ('Challenge contestants'), would be limited to data that do not make it possible to identify individuals. See the CSA Charter, available on the Edmonton Police Foundation website, *supra*, note 24.

[31] K. Cryderman, *supra*, note 26.

[32] Motorola Solutions website, *supra*, note 29.

resources efficiently, and to reduce pressure on the healthcare system, the police and the justice system, which are described as overloaded.[33]

Saskatchewan – Saskatchewan police predictive analytics lab (SPPAL).[34] A technological innovation laboratory, SPPAL, was also set up in Saskatchewan in 2015. It involves the active collaboration of the Saskatoon Police Service (SPS), the University of Saskatchewan, the Government of Saskatchewan and Saskatchewan's social services. SPPAL has developed algorithmic technology that predicts and targets individuals who are 'at risk' of being victims of crime. The technology is used to guide police interventions. The algorithmic model developed by SPPAL would make it possible to identify children and youth likely to abducted. SPPAL intends the technology to be used to intervene preventively with respect to repeat offenders and persons living with drug abuse or mental problems, and to prevent domestic violence. At this time, the technology uses only SPS data, but there are plans to incorporate data from all the municipal police forces in Saskatchewan and from the RCMP's Division 'F', which is the RCMP division associated with the province of Saskatchewan.[35] SPPAL also intends to integrate data from social media into the development of its algorithmic models in the future.[36] Although at this time this model does not seem to operate using massive sharing of data between social services and the police, it is nonetheless the case that this approach—described as an extension of the HUB model already implemented in Saskatchewan[37]—could integrate such data into its algorithm as Citizen Lab has suggested: 'the potential use of Hub model data in algorithmic policing methods was recognized by Public Safety Canada in 2015 when it reported that "[i]ntegrated health, social services, education and criminal justice data analysis will help to identify and plan predictive risk patterns at local, regional and provincial levels".'[38] This innovative approach is also justified by a will to intervene more efficiently with respect to populations judged vulnerable or 'at risk' in order to ensure their security, which should automatically ensure the security of the whole community.[39]

---

[33] K. Cryderman, *supra*, note 26. The Edmonton Police Foundation says that the CSA's main goals are 'Diminishing harm to individuals. - Disrupting, mitigating, and decreasing crime and disorder. - Creating new opportunities for social and economic prosperity including better healthcare outcomes for our most vulnerable.' The following principles are also mentioned: 'Principle 1: Above all, our focus will be on Community Safety and how best to maximize this for all Albertans - Principle 2: Work to create a better experience for Albertans most in need through human-centred design and innovation. … Principle 4: Create new opportunities for social and economic prosperity for Albertans most in need.' See the Edmonton Police Foundation *website, supra, note 24*.

[34] Citizen Lab, 51-52.

[35] Ibid., 51.

[36] Ibid., 51-52

[37] For more on the HUB approach in Saskatchewan, see Public Security Canada website, <https://www.publicsafety.gc.ca/cnt/cntrng-crm/crm-prvntn/nvntr/index-en.aspx?txt=saskatchewan&SORT=Title&BY=ASC> accessed February 2023.

[38] Citizen Lab, 55. See also Public Security Canada, 'Economics of policing and community safety. Policy Makers' Dialogue on Privacy and Information Sharing', (Workshop Report, 2015), 13.

[39] Citizen Lab, 52.

What is the situation in Québec? The Citizen Lab report does not discuss the possible use of AI tools by Québec police services. According to the Ligue des droits et Libertés ('League of Rights and Freedoms'—the League), there would nonetheless be [translation] 'good reasons to think that, as the second largest municipal police force in Canada, the Service de Police de la Ville de Montréal (Montréal City Police—SPVM) may have integrated statistical crime prediction tools into its crime fighting strategies'.[40] In November 2019, in the context of a public hearing of the Commission de la sécurité publiques de Montréal ('Montréal Public Security Committee'), the SPVM refused to confirm this possibility, which had been brought up by the League, on the pretext that it concerned mere 'technicalities of police investigations'.[41]

Algorithmic surveillance tools. According to the Citizen Lab report, many algorithm-based surveillance tools are used by police in Canada. We will present them here briefly. To begin with, automated licence plate reading technologies are used by police services in Ontario, British Columbia, Saskatchewan, Alberta, Nova Scotia, Québec and Prince Edward Island.[42] The SPVM has acknowledged using automated licence plate recognition systems and readers.[43] The SPVM's system is meant to ensure that vehicle drivers comply with the Highway Safety Code, and it checks in particular whether drivers have paid for their licence and licence plate. It can also be used to search for and find stolen vehicles and to search for vehicles in the case of an AMBER (missing person) alert.[44] The checks are done using a database that is updated by the *Société de l'assurance automobile du Québec* (SAAQ), by the Canadian Police Information Centre (CPIC) and, in the context of specific investigations, by the SPVM.

The Calgary Police Service (CPS), the TPS and the RCMP use or have used algorithmic systems to monitor social media. According to the Citizen Lab report, the TPS used AI-based social media analysis technology by the Sysomos/Meltwater company.[45] The RCMP also seems to have been monitoring social media, in its case using a system called Social Studio made by the Carahsoft and Salesforce companies. The RCMP recently entered into a contract with an American company to do social media monitoring of various online communities using the company's AI software, which 'analyzes relationships between the content and its senders, translates content into hundreds of languages, and

[40] Ligue des droits et libertés, 'Mémoire : Étude des technologies de reconnaissance faciale et des lecteurs automatiques de plaques d'immatriculation' (*LDL website*, 30 octobre 2020) <https://liguedesdroits.ca/memoire-reconnaissance-faciale-lapi-csp-montreal-2020/#_ftnref10> accessed April 2022.
[41] *Ibid.*
[42] Citizen Lab, 57.
[43] Commission de la sécurité publique de Montréal, 'Rapport sur l'Utilisation par le SPVM de technologies de reconnaissance faciale et de systèmes de reconnaissance de plaques d'immatriculation' (Ville de Montréal, June 2021).
[44] SPVM, 'Processus d'utilisation du Système de reconnaissance de plaque d'immatriculation (SRPI)' (*SPVM website*) <https://spvm.qc.ca/fr/Fiches/Details/Processus-dutilisation-du-Systeme-de-reconnaissance-de-plaque-dimmatriculation-SRPI> accessed April 2022.
[45] Citizen Lab, 58–59.

filters it based on geographic areas and expressed sentiments'.[46] The Ontario Provincial Police (OPP) also seems to have developed and used a chat room scraping tool called the ICAC Child On-line Protection System.[47] According to the Citizen Lab researchers, this technology would be able to scan online chat rooms, automatically save the content at the end of the chat, upload it and then store it in a database with a search engine available to police officers.

Canadian police are also enthusiastic about facial recognition (FR) tools. The police services in Edmonton, Calgary, Vancouver, Toronto and Halifax have apparently confirmed using or having used this technology.[48] The CPS allegedly uses NEC Corporation FR software called NeoFace Reveal. It would make it possible to associate photographs and composite drawings of unidentified suspects with existing or new mug shots. Again according to Citizen Lab, the TPS also uses FR software, and the York Regional Police and Peel Regional Police Service have taken steps to acquire such a system.[49] While the SPVM says it has not used this technology, in 2020 it said it was ready to use the services of third parties who have it in the context of major investigations.[50] In August 2020, the SQ signed a contract with the French company Idemia[51] to acquire FR and fingerprint recognition technology able to automatically, and in real time, match licence plates or persons and their tattoos on the basis of the provincial fingerprint and mug shot database in the context of specific criminal investigations.[52]

Recent AI tool developments in Canada. In Pennsylvania, the Mila Institute, McGill University and the Carnegie Mellon University School of Computer Science have developed a new AI tool to help police identify potential victims and people involved in human trafficking on the Internet and social media.[53] The InfoShield algorithm is in line with the National Strategy to Combat Human Trafficking 2019–2024 launched by the

[46] Anastasia Konina, 'The Privatization of Law Enforcement: Promoting Human Rights through Procurement Contracts', (2021) 1-1 *McGill GLSA Research Series* 1, 14; See Public Works and Government Services Canada, 'Request For a Standing Offer M7594-184225/B' (14 April 2020).
[47] Citizen Lab, 60–61.
[48] Céline Castets-Renard, Émilie Guiraud and Jacinthe Avril-Gagnon, 'Cadre juridique applicable à l'utilisation de la reconnaissance faciale par les forces de police dans l'espace public au Québec et au Canada - Éléments de comparaison avec les États-Unis et l'Europe' (International Observatory on the Societal Impacts of AI and Digital Technology, Accountable AI in a Global Context Research Chair, 2020), 12.
[49] Citizen Lab, 62.
[50] Commission de sécurité publique de Montréal, *supra*, note 43, Appendix 3, 20.
[51] Ibid., 9.
[52] Céline Castets-Renard, Émilie Guiraud and Jacinthe Avril-Gagnon, *supra*, note 48, 12.
[53] Pascal Robidas, 'Un algorithme de conception québécoise contre l'exploitation sexuelle en ligne' (*Radio-Canada*, 28 juin 2021) <https://ici.radio-canada.ca/nouvelle/1800791/mila-intelligence-artificielle-algorithme-police-exploitation-sexuelle> accessed April 2022.

Government of Canada and the RCMP. The strategy calls for the development of new technologies to fight against new forms of sexual exploitation.[54]

Researchers at the University of British Columbia have created AI software able to predict which new synthetic drugs are most likely to be released into circulation on the market.[55] In order to evade drug regulations, clandestine laboratories work on modifying the molecules of certain well-known drugs so that they will not be identified by police services. Public broadcaster Radio-Canada has explained that [translation] 'to help government agencies identify these new, potentially dangerous psychoactive substances, [these] researchers have trained an artificial intelligence algorithm using a database of 1800 synthetic drugs. Based on the molecular structure of those 1800 substances, the neural network algorithm generated nearly 8.9 million potential synthetic drugs'.[56] The model developed by the University of British Columbia is already being used by the US Drug Enforcement Agency, the United Nations Office on Drugs and Crime, the European Monitoring Centre for Drugs and Drug Addiction and the Federal Criminal Police Office of Germany.[57]

Algorithmic tools rejected by police services. Certain AI technologies used to be used by our police services but have been dropped. After having tested the NeoFace Reveal FR application for three months, the Ottawa police said that they did not want to implement it without consulting the community to protect people's privacy and human rights.[58]

Following the release of the report on the Joint Investigation of Clearview AI, Inc. in 2021, many police services stopped using the FR technology offered by Clearview AI, Inc. The various privacy protection offices across Canada had recommended that the company stop offering its tool in Canada. At the time, a number of law enforcement bodies, including the RCMP, were using the technology. According to the investigation findings, Clearview's AI technology collected images from social media to create a bank of biometric data. The offices found that Clearview AI inc. was required to obtain express consent from the people whose images were collected, which it had not done.[59] In December

---

[54] Mila Institute website (20 May 2021) : <https://mila.quebec/des-chercheurs-de-mila-participent-au-developpement-dun-outil-pour-lutter-contre-le-trafic-de-personnes-et-lexploitation-sexuelle-en-ligne/> accessed February 2023.

[55] M.A., Skinnider, F., Wang, D. Pasin et al., 'A deep generative model enables automated structure elucidation of novel psychoactive substances', (2021) 3 *Nat Mach Intell* 973.

[56] Radio-Canada, Les années Lumières radio show, 'Des algorithmes pour prévoir les nouvelles drogues de synthèse' (*RC website*, 21 November 2021): <Des algorithmes pour prévoir les nouvelles drogues de synthèse (radio-canada.ca)> accessed February 2023.

[57] University of British Columbia, 'UBC researchers train computers to predict the next designer drugs', (*UBC website*, 15 November 2021) <https://www.med.ubc.ca/news/ubc-researchers-train-computers-to-predict-the-next-designer-drugs/> accessed April 2022.

[58] Céline Castets-Renard, Émilie Guiraud and Jacinthe Avril-Gagnon, *supra* note 48, 12.

[59] Office of the Privacy Commissioner of Canada, *Joint investigation of Clearview AI, Inc.* (PIPEDA Findings #2021-001, 2 February 2021) available online: https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2021/pipeda-2021-001/ (hereinafter 'Clearview AI Investigation').

2021, Québec's *Commission d'accès à l'information* (CAI) issued an order giving Clearview AI, Inc. 90 days to delete all the photos of Quebecers that it had collected.[60]

Lastly, there are other technologies that have been cancelled for legal or technical reasons. In 2019, the TPS had to cancel the ShotSpotter automated gunfire detection system following fears expressed by civil society that the system could violate the right to privacy provided for in section 8 of the Canadian Charter of Rights and Freedoms (hereinafter the 'Charter').[61] Another example concerns the police services of Toronto and Calgary, which stopped using the Media Sonor social media surveillance software because it became useless when police services were banned from Facebook and Twitter for having violated privacy policies.[62]

1.2 Reception of AI tools in Canada: precaution in view of the American experience

As there is little public information circulating on the AI tools currently used by Canadian police forces, it is difficult to find large-scale studies on the performance of or safeguards against bias incorporated into specific tools now in use. In consequence, weighing the risks associated with these technologies takes the form of a general precautionary attitude informed by the American experience. The report 'The Rise and Fall of AI and Algorithms in American Criminal Justice Lessons for Canada', published in October 2020 by the Law Commission of Ontario (LCO), is a good illustration of the Canadian approach. The report expresses concerns about the introduction of these new technologies in Canada owing to their potential impact on human rights and calls for protective measures to be taken, with 10 lessons that Canada should learn from the American experience.[63]

---

[60] Tristan Péloquin, 'Clearview AI sommée de détruire ses photos de québécois' (*La Presse*, 14 December 2021) <https://www.lapresse.ca/actualites/2021-12-14/commission-d-acces-a-l-information/clearview-ai-sommee-de-detruire-ses-photos-de-quebecois.php> accessed April 2022. Recently, Clearview AI challenged this order before the courts, saying that their technology did not make it possible to identify which photographs are of Quebecers in order to delete them, Isabelle Ducas, 'Clearview AI dit qu'elle ne peut détruire les photos de Québécois' (*La Presse*, 7 February 2022) <https://www.lapresse.ca/actualites/national/2022-02-07/logiciel-de-reconnaissance-faciale/clearview-ai-dit-qu-elle-ne-peut-detruire-les-photos-de-quebecois.php> accessed April 2022.

[61] Jeff Gray, 'Toronto police end ShotSpotter project over legal concerns' (*Globe and Mail*, 13 february 2019) <https://www.theglobeandmail.com/canada/toronto/article-toronto-police-end-shotspotter-project-over-legal-concerns/> accessed April 2022; Canadian Civil Liberties Association, 'Shotspotter is Not Coming to Toronto, and that's a Win' (*CCLA website*, 14 february 2019) <https://ccla.org/en/privacy/surveillance-technology/shotspotter-is-not-coming-to-toronto-and-thats-a-win/> accessed April 2022 ; Andrea Janus, 'Toronto police scrap plans to acquire controversial gunshot-detection system' (*CBC News*, 14 february 2019) <https://www.cbc.ca/news/canada/toronto/toronto-police-scrap-plans-to-acquire-controversial-gunshot-detection-system-1.5019110> accessed April 2022.

[62] Citizen Lab, 58.

[63] Law Commission of Ontario, 'The Rise and Fall of AI and Algorithms in American Criminal Justice: Lessons for Canada', 2020 (hereinafter 'CDO1').

Along the same lines, the Citizen Lab report devotes considerable attention to the fears expressed by Canadian academics and NGO activists, who, in light of American experiences, have reasonable concerns about the potential impact of these tools on segments of our population that are already marginalized and over-represented in police interventions in Canada, such as racialized persons, persons living with mental health issues, persons with diverse gender identity and sexual orientation, and Indigenous people.[64] These fears are shared by the authors of the Citizen Lab report. They base their analysis on numerous studies and journalistic investigations conducted in other countries, in particular in the United States and Britain. Among other things, the authors express fears about the Calgary and Toronto police forces' use of NEC Corporation FR technology in light of a study done in Britain on other NEC products that showed inaccuracies and bias in the way data was processed.[65] In a brief submitted in 2020 to the *Commission de la sécurité publique de Montréal*, the League also expressed its misgivings, in light of the [translation] 'worrisome trend [toward police use of AI technology] that has been growing in North America since approximately 2011', about the impact of the use of FR technologies on segments of the population that are already subject to racial profiling.[66] The League based its concerns on the Armony-Hassaoui-Mulone Report submitted to the SPVM in 2019, which established the over-representation of racialized and Indigenous persons among those subjected to street checks, arrests and detentions in Montréal. The Armony-Hassaoui-Mulone Report also anticipated the negative effects of the use of predictive AI technologies on these segments of the population:

> [translation] Criminal profiling has been fine-tuned in recent years through the development of information technologies and the advent of big data, which have made it possible to create increasingly advanced predictive tools, whether for use in geospatial analysis (to identify the probable locations of future crimes) or to access the risk of recidivism (to identify individuals who are more likely to (re)offend). Since even the smallest police services have already integrated these predictive technologies into their practices, there is every reason to believe that these strategies for analysing criminality will play an even bigger role in the future. Yet, beyond their possible efficiency or inefficiency in lowering the crime rate, the emphasis placed on these tools can reinforce existing profiling. As soon as criminal profiling (prediction) is based on elements linked directly or indirectly to "racial"

---

[64] Citizen Lab, 26–28.

[65] Citizen Lab, 92: 'Further, a 2018 report by Big Brother Watch indicated that NeoFace Watch, a facial recognition product by NEC Corporation—the company from which the CPS and the TPS procured their facial recognition technologies—was found to produce inaccurate matches 91 to 98 percent of the time, in usage by the Metropolitan Police and South Wales Police in the United Kingdom. . . . Such findings raise questions about the reliability of a technique that can lead to arrests or criminal charges on the basis of misidentification.'; For the study, see Big Brother Watch, 'Face Off: The lawless growth of facial recognition in UK policing', 2018.

[66] Ligue des Droits et Libertés, *supra*, note 40. The League bases its concerns on the following study by Will Douglas Heaven, 'Predictive policing algorithms are racist. They need to be dismantled' (*MIT Technology Review*, July 17 2020) <https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/> accessed April 2022.

> belonging (the colour of one's skin, of course, but also the way of dressing and walking, the kind of body language or simply where one lives), existing racial disparities will necessarily be accentuated. And, at the same time, there will probably be more street checks of citizens who are not criminals but who belong to the targeted group.[67]

In the Clearview AI Investigation, the federal and provincial privacy commissioners did not assess the technical accuracy of the facial recognition technology, but they nonetheless said that they 'recognize a number of concerns related to facial recognition technology, generally'.[68] Their misgivings concerned FR technology's efficiency and accuracy, and the possibility that it could make identification errors, and were based on a study by the United States National Institute of Standards and Technology.[69] The privacy commissioners also expressed special concern about the high rate of false positives 'when assessing the faces of people of colour, and especially women of colour, which could result in discriminatory treatment for those individuals'.[70]

Canada would be advised to create a public list of the different AI tools used by law enforcement. Such a list would allow and encourage independent research on the performance, reliability and impartiality of these tools. At this point in time, there is no list aside from the survey done by the Citizen Lab researchers. It could be useful to conduct studies in Canada similar to those that have been done elsewhere in the world on the tools that are currently being used by our police forces.

2 Normative framework

To date, there is no law, directive or major policy by the Canadian government or the provincial governments specifically regulating AI tools used for predictive policing or algorithmic surveillance.[71] The legal framework remains lacking and provides no basic guarantees with regard to transparency, accountability, performance, safeguards against bias, reliability, certification or labelling, which are aspects generally identified in technology law as necessary given the potential of these new technologies. We propose **(2.1.)** to examine the principles of law that provide the foundation for the first attempts to regulate these new technologies and **(2.2.)** to see how existing legislation on the accuracy of personal information provide a minimal guarantee that the AI tools used for predictive policing will have a degree of reliability. Lastly, **(2.3.)** we will review the primary obstacles to ensuring transparency in the way AI tools operate.

---

[67] Victor Armony, et al., 'Les interpellations policières à la lumière des identités racisées des personnes interpellées', (Final report to the SPVM, August 2019), 19–20.
[68] Clearview AI Investigation, [91]–[97].
[69] Patrick Grother, Mei Ngan and Kayee Hanaoka, 'Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects' (*NIST website*, 2019), Face Recognition Vendor Test, Part 3: Demographic Effects: <https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf> accessed April 2022.
[70] Clearview AI Investigation, [95].
[71] Citizen Lab, 9.

2.1 Normative framework and principles of law

Even though the norms that we will present here do not have the status of laws or Normative densification. Even though the norms that we will present here do not have the status of laws or regulations, they aspire to regulate certain forms of behaviour related to the use of AI technologies by police and decision-makers. Norms, as imperatives meant to regulate, normalize and prescribe forms of behaviour, can take many forms (oral or written, published in different formats), govern a smaller or larger number of individuals or parties (internal or public directives), follow different development processes (diplomatic agreements, public inquiries, collaboration with private actors), be stated with different degrees of coercive force (recommendations, guides, requirements, principles) and come from different sources and types of authority (diplomatic agreements between ministers from different countries, standards organizations, the Treasury Board, the Office of the Privacy Commissioner of Canada, senior officers in police forces, civil society). Today in Canada, sources of norms (standards bodies, private companies, police departments, multi-sectoral councils, etc) are increasing in number and becoming more diverse, and there is also a trend, though still weak, toward intensifying norms (in other words, norms that were initially in civil society in the form of principles and internal directives have in some cases been reproduced entirely, in part or in amended form in directives issued by administrative and governmental authorities), and toward enriching normative content (the federal government's Directive – which we will describe – can be amended and may change; other internal directives have also been amended following inquiries). Based only on the evolution of normative activities over the last four years, we also predict that there will be an increase in the volume of norms, and possibly an extension of their scope (from administrative to criminal law), and we can already see an increase in the number of players concerned by these norms (police officers, AI tool designers, laboratory technicians, decision-makers).[72]

Our presentation will therefore pay attention to 'normative density', in other words, norms' aspiration and capacity to regulate behaviour, their degree of detail, their coercive force and even their authority. Normative densification is both quantitative and qualitative—and both of these dimensions should be considered. Normative densification can also be described as a 'polarizing process': on one hand, the number of fields regulated by the norm expands and the number of sources and the volume increase, and on the other hand, the norm becomes more concentrated and expressed with greater precision and strength.[73] Based on the first normative attempts to regulate AI technologies, it is also possible to identify a common core of fundamental principles of law that could probably be redeployed in future legislative efforts to regulate in a more precise manner the use of AI technologies by police and decision makers.

[72] Catherine Thibierge, ed., *La densification normative. Découverte d'un processus*, (Paris: Éditions mare & martin 2013), pp. 1123–1124.
[73] Ibid., p. 1108.

Principles upheld by Canada internationally. The Canadian federal government's commitments abroad reveal the importance that it places on protecting human rights with regard to the use of AI technologies. For example, there is the 2018 Canada-France Statement on Artificial Intelligence, in which Canada and France committed to establishing an international study group on these technologies and to promoting 'a vision of human-centric artificial intelligence grounded in human rights, inclusion, diversity, innovation and economic growth'.[74] The Statement is intended as a reminder of the G7 Innovation Ministers' Statement on Artificial Intelligence adopted in Montréal on March 28, 2018, in which the G7 representatives made a commitment to fostering the development of 'human-centric' AI technology while sustaining economic growth and innovation.[75]

There is also the OECD's 2019 Recommendation of the Council on Artificial Intelligence, to which Canada adheres. It is the 'first intergovernmental standard on AI' and is organized around 'five complementary values-based principles for the responsible stewardship of trustworthy AI'. These guiding principles are 'inclusive growth, sustainable development and well-being', 'human-centred values and fairness', 'transparency and explainability', 'robustness, security and safety', and 'accountability'. The second principle, 'human-centred values and fairness', is especially pertinent for regulating the implementation of AI technology in criminal law. The principle is expressed in these terms: 'AI actors should respect the rule of law, human rights and democratic values, throughout the AI system lifecycle. These include freedom, dignity and autonomy, privacy and data protection, non-discrimination and equality, diversity, fairness, social justice, and internationally recognised labour rights.'[76] This principle ensures that new technologies will have to comply with the law applicable where they are deployed, and not vice versa.

In accordance with the Canada-France Statement on Artificial Intelligence, which announced the intention to create an international group of experts on AI, France and Canada planned to set up the International Panel on Artificial Intelligence, with a mission 'to support and guide the responsible adoption of AI that is human-centric and grounded in human rights, inclusion, diversity, innovation and economic growth'.[77] In the end, the panel was renamed the Global Partnership on Artificial Intelligence (GPAI). Canada is a

---

[74] *Canada-France Statement on Artificial Intelligence* (*Government Canada*, 7 June 2018) <https://www.international.gc.ca/world-monde/international_relations-relations_internationales/europe/2018-06-07-france_ai-ia_france.aspx?lang=eng> accessed April 2022.

[75] G7 Innovation Ministers, 'Annex B: G7 Innovation Ministers' Statement on Artificial Intelligence, Montreal, Canada' (*UofT website*, March 28 2018) <http://www.g8.utoronto.ca/employment/2018-labour-annex-b-en.html> accessed April 2022.

[76] OECD, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449 (2019) (*OECD website*) <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449> accessed April 2022.

[77] Prime minister, 'Mandate for the International Panel on Artificial Intelligence' *(Prime Minister website*, 6 December 2019) <https://pm.gc.ca/en/news/backgrounders/2018/12/06/mandate-international-panel-artificial-intelligence> accessed April 2022; *Declaration of the International Panel on Artificial Intelligence* (*Government Canada*, 16 may 2019) <https://www.canada.ca/en/innovation-science-economic-development/news/2019/05/declaration-of-the-international-panel-on-artificial-intelligence.html> accessed April 2022.

member of the GPAI, and one of the GPAI's centres of expertise is in Montréal (the International Centre of Expertise in Montréal for the Advancement of Artificial Intelligence). Canada's Minister of Innovation, Science and Industry, François-Philippe Champagne, is the past Council Chair (2020–2021) of the GPAI. The Partnership is recognized by the OECD and structured around the OECD Principles in the OECD Council Recommendation on Artificial Intelligence, and the OECD is a permanent observer of the GPAI and hosts the GPAI Secretariat. The GPAI's preliminary terms of reference confirm that it aims to create working and research groups to develop AI in a way that respects human rights, inclusion, diversity, innovation and economic growth.[78] In Montréal, the centre of expertise is home to two working groups: one on responsible AI and the other on data governance.

The CAN/CIOSC 101:2019 National Standard. The Chief Information Officer Strategy Council (CIOSC) has established 'the world's first standard that establishes minimum ethical protections in the design and use of automated decision systems' to provide 'a framework and process to help organisations address AI ethics principles, such as those described by the OECD'.[79] The CIOSC is a national non-profit corporation created in July 2017.[80] It brings together chief information officers from various sectors, including companies, provincial and federal governments, municipalities and non-profit organizations. Its mission is to build and influence Canada's technological ecosystem and to support Canada in the new data-based economy by developing national standards for new technologies.[81] Until legislation and regulations are passed by our Parliament, these national standards are intended as front-line norms to regulate AI tools. Alex Benay, CIOSC co-founder and past Co-Chair, says: 'We need to be laser-focused on developing next generation technology standards to fill gaps created by legacy regulation and legislation that just haven't kept up with the pace of change.'[82] The CIOSC has been accredited by the Standards Council of Canada, which is the primary standards accreditation body in Canada and a Crown corporation created under the Standards Council of Canada Act 'to enhance Canada's competitiveness and well-being'.[83] The CAN/CIOSC 101:2019 National Standard was developed by the CIOSC in 2019. It applies to all private, public, non-profit and government bodies seeking to use machine-learning AI in automated decision-making systems. In line with the OECD's mission, the National Standard stipu-

[78] See the GPAI's terms of reference document, (*GPAI website*) <https://gpai.ai/about/gpai-terms-of-reference.pdf> accessed April 2022.

[79] Senator Colin Deacon, 'Focusing on ethical AU will unlock social and economic opportunity' (*Senate website*, April 13, 2021) <Focusing on ethical AI will unlock social and economic opportunity: Senator Colin Deacon (sencanada.ca)> accessed April 2022. This norm is also said to be inspired by the Government of Canada's *Directive on Automated Decision-Making*.

[80] CIO Strategy Council website: <https://ciostrategycouncil.com/standards/> accessed April 2022. Since 2023, it is called the DIGITAL GOVERNANCE STANDARDS INSTITUTE.

[81] CIO Strategy Council website: <https://ciostrategycouncil.com/about/> accessed April 2022.

[82] CIO Strategy Council website: <https://ciostrategycouncil.com/history/> accessed April 2022.

[83] Standards Council of Canada, 'mandate, Mission and Vision' (*SCC website*): <https://www.scc.ca/en/about-scc/mandate-mission-vision> accessed February 2023.

lates that 'AI should benefit people and the planet by driving inclusive growth, sustainable development and well-being'.[84] It establishes minimum requirements with regard to transparency, accountability, safeguards against bias and performance for 'protecting human values and incorporating ethics in the design and use of automated decision systems'.[85]

Preliminary conclusion. It is clear from these first normative initiatives that the principles deriving from them, which are likely to be reflected in future legislative efforts, have their sources in Canada's diplomatic commitments with respect to the OECD and Canada's participation in the G7, and it is notable that the OECD and the G7 are both essentially economic bodies. The 'principles' upheld by Canada in its international commitments have an ethical dimension, but also a significant economic dimension. As can be seen from the Canada-France Statement, the OECD's Recommendation of the Council on Artificial Intelligence, GPAI's terms of reference and the CAN/CIOSC 101:2019 National Standard, the 'principles' promoted in these initiatives all refer to AI tools' potential with regard to 'economic growth'. Growth-centred initiatives seem unfit for application in a field as special and sensitive as criminal law and policing. It is obviously problematic to consider economic growth as a value in the same way as human rights principles. If it were applied in criminal law, this ordering of values could have negative consequences on the rights of suspects and the accused. It would not be tolerable, for example, for intellectual property rights or patent rights to limit or compete with a right as fundamental as that to a full and complete defence.

Directive on Automated Decision-Making. Committed to these same principles, the Government of Canada, through the President of the Treasury Board, issued its very first Directive on Automated Decision-Making (in force since April 1, 2019). The Directive applies generally, with no other specifications, to federal public sector services looking to use AI technologies in their decision-making. It has been criticized for the fuzziness of its field of application.[86] It seems difficult to apply to predictive policing practices and not to have been written with that context in mind. In fact, it was designed to apply to administrative decisions and makes no reference to police or criminal procedures. Moreover, it concerns federal institutions only, so it is not binding on provincial or municipal police services, where most AI technologies are used.[87]

Principles from the document entitled Privacy guidance on facial recognition for police agencies. In 2021, as follow-up to the Clearview AI Investigation, the Office of the Privacy Commissioner of Canada published a document entitled Privacy guidance on facial recognition for police agencies. It is meant to provide a framework for the use of

[84] CIO Strategy Council website: <https://ciostrategycouncil.com/normes/conception-ethique/?lang=fr> accessed April 2022.
[85] Ibid.
[86] CDO1, p. 39.
[87] Citizen Lab, p. 142.

FR technologies by federal, provincial and municipal police forces.[88] The goal of the guidance is to strengthen existing protection for privacy, which can be threatened by the technological potential of FR tools. The guidance states obligations that should be met prior to using FR technologies and that are related to the following guiding principles: 'necessity and proportionality' with regard to the use of the technology, 'accuracy' of the software, 'data minimization', 'decision-makers', 'accountability', 'openness', 'transparency' and 'individual access' to one's own personal information.

Principles from civil society. Some policy statements from non-governmental organizations may also guide police practices or inspire future legislative efforts. For example, there is the Montreal Declaration for a Responsible Development of Artificial Intelligence[89] (2018), which recalls the importance of ensuring that technology is developed in ways that respect the right to fairness, for privacy and the need to guarantee democratic participation. The Toronto Declaration[90] (2018) by Access Now argues for the need to ensure decision makers' accountability when they use AI tools. It calls for proactive measures to minimize these technologies' impact on the right to equality and to clearly limit their use to what is necessary. These declarations, which result from collaboration among various civil society stakeholders, aim to promote the ethical use of these technologies. The approaches they recommend are based on the principles of fairness, accountability, transparency and ethics ('FATE principles').[91] The Sedona Principles for E-discovery are also intended to provide a principle-based normative framework for evidence gathering by Canadian police.

Internal police initiatives and guidelines. Even today, the most abundant normative activity in this area is found within police departments. In the absence of a binding normative framework, police forces self-regulate their use of new AI technologies, and it is distressing that this has been left to their discretion. Contrary to what the SPVM claimed at a press conference, whether or not to use these new technologies and the way they are used are not simple 'technicalities'[92] pertaining to police investigations: they raise broader collective issues. The use of these technologies requires public debate about how we want to ensure social harmony and peace in the public space and how we want to weigh and balance the need for security against our fundamental rights. In the past, privacy commissioners have found some of the police's internal guidelines wanting.

---

[88] Office of the Privacy Commissioner of Canada, *Privacy guidance on facial recognition for police agencies*, (*OPCC website*, 2021): <https://www.priv.gc.ca/en/privacy-topics/surveillance/police-and-public-safety/gd_fr_202205/> accessed on April 2022. The guidance must be in line with and relate to other related directives, in particular the Treasury Board Secretariat's *Directive on Privacy Impact Assessment* (*Government Canada*, 18 June 2020) <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=18308> accessed on April 2022.

[89] *Montréal Declaration:* https://www.declarationmontreal-iaresponsable.com/.

[90] *Toronto Declaration*: https://www.torontodeclaration.org/declaration-text/english/.

[91] Citizen Lab, p. 9.

[92] Ligue des droits et libertés, *supra*, note 40.

For example, the Clearview AI Investigation into the use of FR found that the RCMP's security guidelines had previously directed that links made by the technology be treated as leads and not as confirmed results. The various federal and provincial privacy commissioners responsible for the Investigation determined that such a directive was not sufficient to govern the use of FR technology and that additional measures were required to ensure accuracy and to counteract the possibility of 'false positives' and discriminatory bias.[93] It was only as a result of this investigation, in March 2021, that the RCMP introduced the National Technology Onboarding Program to systematically review the compliance of new technological tools used in the course of investigations with the Privacy Act and the Canadian Charter of Rights and Freedoms.[94]

The CPS has developed its own written policy on the collection of information on social networks. Police officers are allowed to collect such information, but they must limit it to the specific purposes of the investigation. The policy also allows the police to collect any information that could be 'threat-related'. The authors of the Citizen Lab criticized this instruction for its vagueness.[95]

In the absence of binding state standards, the VPD has also, on its own initiative, tried to limit the impact of its AI technology on segments of the population that are already overrepresented in police responses. To limit the impact of police bias that is present in historical data, the VPD ensures that the data entered into its system come exclusively from break-and-enter cases reported by ordinary citizens. It also excludes some of the more "sensitive" areas, such as the Downtown Eastside, where the population is poorer, from its surveillance mapping and has ultimately instructed patrol officers not to use the application to justify routine voluntary identification checks (street checks).[96] The Citizen Lab researchers acknowledged that the VPD's approach and rules were useful for limiting some of the negative effects of the technology, but also expressed some reservations about whether they would be able to counter other forms of discriminatory bias and the potential for people to be biased in favor of predictions made by technology (automation bias).[97]

---

[93] Office of the Privacy Commissioner of Canada, "Police Use of Facial Recognition Technology in Canada and the Way Forward", 2021, [81]. (hereinafter "Special Report on FR Technology").

[94] Royal Canadian Mounted Police, 'Response to the Report by the Office of the Privacy Commissioner into the RCMP's use of Clearview AI'(*RCMP website*, June 10 2021) <https://www.rcmp-grc.gc.ca/en/node/91915> accessed April 2022. See also the Special Report on FR Technology.

[95] Citizen Lab, p. 58: 'Under the policy, officers may collect publicly available data, including data processed by third-party social network aggregators and software. Officers are restricted, however, to collecting only information that is linked to a specific investigative purpose, including "threat-related information'. The policy does not define what 'threat-related information" means nor does it restrict the CPS from using products like Media Sonar in the future, should they become useful once more for investigations.'

[96] Citizen Lab, p. 44.

[97] Citizen Lab, pp. 109–110 and 125–126.

2.2 Accuracy of personal information as a minimal guarantee of the reliability of AI tools

The adoption of specific rules to regulate the use of AI technologies in criminal law and policing, in particular with regard to FR, is necessary because of how sensitive the information in question is (e.g., unalterable biometric data) and because of the need to expand our existing protection as fast as the technological tools become more powerful. AI tools' processing capacity and the nature of the data processed (historical, multi-source, decontextualized) demand we remodel the existing framework. While there may not yet be any rules strictly regulating the standard of reliability that must be met by AI tools, existing guarantees regarding accuracy of personal information indirectly provide a minimal level of reliability.

However, these new tools require police databases to meet higher standards of reliability and accuracy and to be updated constantly because AI processing is automated, virtually instantaneous and continuous. The criminal records databases currently used by law enforcement do not seem to meet a high enough standard to supply AI tools. For example, the CPIC, which is the primary and most widely used criminal records database employed by Canadian police forces and government departments, is recognized as containing information that is outdated and inaccurate.[98] The way data is kept in the RCMP's exempt record database has also been severely criticized. In 2008, the Privacy Commissioner of Canada conducted a study to determine whether the data contained in the RCMP's exempt records had been assessed for reliability and the results recorded.[99] Under section 18 of the Privacy Act, the RCMP can declare certain of its records containing personal information collected during criminal investigations 'exempt from public access'; such records could possibly concern innocent people if they were 'in the wrong place, at the wrong time, talking to the wrong person'.[100] For these reasons, the bodies concerned must ensure that the content of these files is limited to what is 'legitimate', and they are responsible for classifying and organizing the information into files that are 'locatable'. Files must also be kept under unique numbers, be checked and be subject to

[98] Citizen Lab, 85 quoting Alyshah Hasham, 'Criminal-record database spotty and out of date, lawyers lament' (*The Toronto Star*, December 9 2016): <https://www.thestar.com/news/crime/criminal-record-database-spotty-and-out-of-date-lawyers-lament/article_39ef3ef1-e377-54bb-9430-63214a8931d3.html> accessed April 2022; Brigitte Bureau, 'RCMP database remains out of date, police and prosecutors say', (*CBC News*, March 10, 2015): <https://www.cbc.ca/news/politics/rcmp-database-remains-out-of-date-police-and-prosecutors-say-1.2989397> accessed April 2022. See also more recently: Nicole Brockbank, 'How a criminal charge laid in Calgary was linked to a Toronto woman who's never been there', (*CBC News*, January 21 2021): <https://www.cbc.ca/news/canada/toronto/false-identity-rcmp-database-1.5881006> accessed April 2022. See also Renata D'Aliesio and Kathryn Blaze Carlson, 'Substantial gap discovered in RCMP database of anonymous dead' (*Globe and mail*, March 16 2015) <https://www.theglobeandmail.com/news/national/substantial-gap-discovered-in-rcmp-database-of-anonymous-dead/article23467796/> accessed April 2022.

[99] Office of the Privacy Commissioner of Canada, *Examination of RCMP Exempt Data Banks*, (*OPCC website*, February 2008), [1.1]–[1.8]: <https://www.priv.gc.ca/fr/mesures-et-decisions-prises-par-le-commissariat/verifications/rcmp_080213/> accessed April 2022.

[100] Ibid., [1.7].

a periodic review process. The Commissioner's report revealed that almost none of the files had been subject to regular monitoring to ensure that it was appropriate for them to remain classified as exempt files, in compliance with RCMP policy.[101] It would be problematic for such a database or information obtained from such a database to be fed into an AI tool.[102]. In the end, there is also a problem with inputting incident reports by police officers into AI tools because of the potential police bias and lack of standardization that has been shown to result from doing so: 'The analyst above highlights the lack of standardization and training, as well as selective reporting, which raises concerns regarding data quality and integrity.'[103] The accuracy of the data input into an AI tool is a minimal guarantee that reasonable use will be made of the technology. Inaccurate or incomplete data can generate false results and lead to disproportionate or unwarranted intervention by the police. For example, biased or inaccurate data can lead to arbitrary detention, for the grounds for suspicion would be unreasonable; such data can also result in discriminatory police tactics, such as racial profiling, thereby violating the rights to equality and non-discrimination.[104]

A duty to fight pro-actively against risks of error? In current law, there seems to be emerging a kind of duty to take pro-active measures to ensure the accuracy of data when it is to be used by AI technology. To begin with, section 6(2) of the Privacy Act requires all federal bodies, such as the RCMP, to 'take all reasonable steps to ensure that personal information that is used for an administrative purpose by the institution is as accurate, up-to-date and complete as possible'. The Privacy Commissioner of Canada ruled that, when police use FR technology, a general internal instruction to investigators to consider FR-generated results 'as leads, not confirmed identity matches' could not be sufficient to discharge the section 6(2) obligation. This means that, when AI technology is used, additional measures must be taken to ensure data accuracy and to neutralize the possibility of false positives and discriminatory bias.[105] Along the same lines, a similar law in Québec provides that Québec police 'must see to it that the personal information kept by it is up to date, accurate and complete so as to serve the purposes for which it is collected or used'[106]: [translation] 'it may also be asked whether the accuracy obligation could apply to the process and to biometric data, which would provide a foundation for an obligation to fight against the risk of error'.[107] The public bodies of the other provinces are also required to take reasonable measures to ensure that the personal information that they

[101] Ibid., [1.9].

[102] For an example of obsolete information found in the data bank, see Ibid., Exhibit E: 'A resident alleged that an individual entered a rooming house in the neighborhood. Believing that drugs may have been involved, the resident contacted the police. The investigation revealed that the individual had dropped off his daughter at school (down the street from the rooming house), and he had stepped out of his car to have a cigarette. The file was concluded approximately seven years ago.'

[103] Carrie B. Sanders, Crystal Weston and Nicole Schott, *supra*, note 6, 720.

[104] Citizen Lab, 18–25 and 85. See also Citizen Lab, Section 2.2.

[105] Special Report on FR Technology, [80]–[85].

[106] Act respecting Access to documents held by public bodies and the Protection of personal Information, CQLR, c. A-2.1, s. 7 **(Québec)**.

[107] Céline Castets-Renard, Émilie Guiraud and Jacinthe Avril-Gagnon, *supra*, note 48, 41.

collect and use is accurate and up to date,[108] although certain police forces, such as those of Ontario, are exempt from that specific obligation.[109].

2.3 Obstacles to guarantees of algorithmic tool transparency

The inherent opacity and complexity of how AI tools operate threatens the fairness of criminal trials guaranteed by section 7 of the Charter. Guaranteeing the transparency of AI tools is essential to ensure the accused's right to a full and complete defence. To exercise this right, the accused must have access to all the information necessary to make their case and respond to the offence with which they are charged.[110] The opaque operation of AI tools deprives them of this right. Moreover, the exercise of other constitutional rights is also threatened by this opacity: the presumption of innocence (s. 11(d) of the Charter), the right to non-discrimination (s. 15 of the Charter), protection against arbitrary detention or arrest (s. 9 of the Charter) and the right to redress (s. 24(1) of the Charter) are among the constitutional rights affected by the technical opacity of AI tools.

Full transparency is also essential to ensure the accountability and responsibility of decision makers (police officers and judges) who must be aware of how the AI tool they are using works and what its recommendations are based on.[111] The structure of AI tools is in itself normative.[112] It is the result of human decisions, of reflection, whether conscious or not, of political and ethical choices in the ordering of parameters, and that ordering, which, through the tool's recommendations, influences and constrains the decision-maker. In this sense, its architecture must also be able to be the subject of adversarial debate, otherwise the judge's decision will be usurped by the choices of the algorithm's designer. As the authors Antoine Garapon and Jean Lassègue explain, [translation] 'If predictive justice does not want to be seen as magical divination, just as mysterious and intimidating as the ancient oracles, then it must make its algorithms public and not hide behind a trade secret (which means that copyright law will have to be changed). . . . [for] if there can be no objection, there can be no law'.[113] For these reasons, new legislation should recognize that persons must have some guarantee of transparency when they are

---

[108] Freedom of Information and Protection of Privacy Act, RSA 2000, c F-25, s 35 **(Alberta)**; Freedom of Information and Protection of Privacy Act, RSBC 1996, c 165, s 28 **(B.C.)**; Local Authority Freedom of Information and Protection of Privacy Act, SS 1990-91, c L-27.1, s 26 **(Saskatchewan)**.

[109] As an example, Citizen Lab, p. 86, cites the *Freedom of Information and Protection of Privacy Act*, RSO 1990, c F.31, s 40(3) **(Ontario)** and the *Municipal Freedom of Information and Protection of Privacy Act*, RSO 1990, c M.56, s 30(3) **(Ontario)**.

[110] *R.* v. *Seaboyer*, [1991] 2 SCR 577; Anastasia Konina, *supra*, note 46, 16–18: 'The opacity of algorithms, also referred to as the black box problem, strongly suggests that the law enforcements' use of technology is incompatible with the *Charter* right to make full answer and defence.' We will come back to this in greater detail in Part III on the law of evidence.

[111] Citizen Lab, 129-132.

[112] Lawrence Lessig, Code and Other Laws of Cyberspace, (New York: Basic Books 1999); Joel R. Reidenberg, Lex Informatica: The Formulation of Information Policy Rules through Technology, (1997) 76 Tex. L. Rev. 553; Langdon Winner, The Whale and the Reactor: A Search for Limits in an Age of High Technology, (Chicago: University of Chicago Press 1986).

[113] A. Garapon and J. Lassègue, *supra*, note 13, 242.

subject to policing and criminal charges based on the use of AI technology: they require the right to have access to the source code and to an intelligible explanation of how that code works.

At this point in time in Canada, the main obstacles to this guarantee of transparency are (i) the principle of trade secrecy, which prevents the source code from being made accessible to everyone, (ii) the plaintiffs and defendants' lack of training and knowledge of how the technology works ('technical illiteracy'), and (iii) the very operation of machine-learning AI tools, which, over time, causes the structure of the code to evolve to the point where the person using it (and sometimes even the coder) no longer knows how it works.[114]

There is currently no legislation in Canada requiring vendors of AI tools to disclose the source code of the tools they provide to police. Trade secrecy, intellectual property and patent rules appear to be the main limitations to ensuring transparency. For example, Article 19.16 of Chapter 19 of the Canada-U.S.-Mexico Agreement precludes requiring technology providers to disclose source code prior to import. It is therefore only after the fact—that is, after the accused's rights have been violated and in the context of a specific investigation—that the source code may be consulted, and then only under certain conditions.[115]

3 General principles of law

3.1 Right to privacy

Regarding privacy rights, both the federal and the provincial governments are competent with respect to organizations under their jurisdiction. There are therefore different privacy laws for public and private organizations, but the principles organizing privacy rights are pretty much the same at the federal and provincial levels. With respect to the private sector, the federal Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5 applies to all provinces that have not enacted "substantially similar" legislation. To date, only Québec, British Columbia and Alberta have enacted legislation to govern the private sector that has been found to be substantially similar to the federal legislation.

Despite language that appears to be limited to prohibiting unreasonable searches by the state, section 8 of the Charter provides individuals with broader moral protection 'from unjustified state intrusions upon their privacy'.[116] In this sense, the protection provided by section 8 encompasses three spheres of privacy claims, three specific expressions of the right to privacy: **(i)** privacy in relation to the body (the personal sphere), **(ii)** privacy

[114] Angèle Christin, 'Predictive algorithms and criminal sentencing', 283 in N. Guilhot and D. Bessner (eds), *The Decisionist Imagination* (Berghahn Books 2018).
[115] Citizen Lab, 131–132.
[116] *Hunter* v. *Southam*, [1984] 2 S.C.R. 145, 160.

in relation to places (the spatial sphere) and **(iii)** privacy in relation to personal information (the informational sphere).[117] This way of categorizing these non-watertight spheres of claim helps to illustrate the unsuspected scope of privacy protection: it is a rich, complex, fragmented right. Complex, because it enjoys both specific constitutional and specific legislative protection. Fragmented, because the constitutional division of powers between the provinces and the federal government[118] means that the right to privacy is protected by both provincial and federal laws in their respective jurisdictions and the requirements differ depending on whether one is dealing with public or private institutions. In order to function, AI tools require the collection, storage and use of personal information on a massive scale, so we will focus on the informational sphere of the right to privacy. We will begin by presenting legislative privacy protection **(3.1.1)**: the protection of informational privacy that legislation imposes on both federal and provincial public institutions (such as police forces) and the data protection obligations that legislation imposes on private companies when they interact with the police. After that, **(3.1.2.)** we will turn to constitutional privacy protection.

3.1.1 Legislative protection of information privacy

In principle, privacy rights encompass a variety of protections for 'personal information' when it is collected, used and shared by different organizations. While the wording of statutes varies from province to province, the underlying structural principles are similar. 'Personal information' includes any 'identifying information', such as any data or information that relates to or identifies a specific individual.[119] This means that data that, on its own, would not identify a person but that, when processed by the AI tool and coupled with other information collected by the algorithm, would identify that person could be considered 'personal information' and would be protected by law.[120]

---

[117] Karim Benyekhlef and Pierre-Luc Déziel, *Le droit à la vie privée en droit québécois et canadien*, (Montréal: Éditions Yvon Blais 2018) (hereinafter: 'Benyekhlef and Déziel') in reference to the classification proposed in *R.* v. *Dyment*, [1988] 2 SCR 417.

[118] *Constitution Act, 1867*, 30 & 31 Vict., c. 3 (U.K.), ss 91 and 92.

[119] Benyekhlef and Déziel, p. 262; **Federal government institutions**: the *Privacy Act*, S.C. (1985), c. P-21, s 3, which applies to the RCMP; **Provincial public bodies**: the *Act respecting Access to documents held by public bodies and the Protection of personal information,* CQLR, c. A-2.1, s 54 and the *Act to establish a legal framework for information technology*, CQLR, c. C-1.1 **(Québec)**; the *Freedom of Information and Protection of Privacy Act* (FOIP), R.S.A. c. F-25, s 1 **(Alberta)**; the *Freedom of Information and Protection of Privacy Act* (FIPPA), R.S.B.C. 1993, c. 165, s 1 **(British Columbia)**; the *Freedom of Information and Protection of Privacy Act* (FIPPA), R. S.P.E.I. 1988, c. F-15.01, ss 1.1. and 2 **(Prince Edward Island)**; the *Freedom of Information and Protection of Privacy Act*, C.C.S.M., c. F175, s. 1 **(Manitoba)**; the *Right to Information and Protection of Privacy Act*, S.N.B. 2009, c. R-10.6, s 1 **(New Brunswick)**; the *Freedom of Information and Protection of Privacy Act*, S.N.S. 1993, c. 5, s 4(i) **(Nova Scotia)**; the *Freedom of Information and Protection of Privacy Act*, R.S.O. 1990, c. F.31 **(Ontario)**; the *Access to Information and Protection of Privacy Act*, S.N.L. 2015, c. A-1.2, s 2 **(Newfoundland and Labrador)**; the *Access to Information and Protection of Privacy Act*, S.S. 1990-91, c. F-22.01, s 24(1)a) **(Saskatchewan)**

[120] Benyekhlef and Déziel, 266, and 268, where they explain that the threshold over which indirect information can be considered as identifying is not yet clearly defined in Canadian law, see *Gordon* v. *Canada (Health)*, 2008 FC 258.

General panorama of protection. Public bodies – Individuals can generally expect **(i)** that the collection of their information will relate directly to the activities of the public body and will be limited to what is necessary for those activities, i.e., it will be 'essential'[121] and not merely convenient;[122] **(ii)** that it will be used for the purpose for which it was collected or for a purpose consistent with that purpose;[123] **(iii)** that the collection is directly from the individual and that the individual is notified of the collection and the purposes of the collection,[124] with certain exceptions;[125] **(iv)** that disclosure to other organizations is permitted only under certain conditions specific to each province, for example, where it is 'necessary' for the enforcement of a law.[126] **Private bodies –** Currently, a corpus of federal and provincial legislation applying to private businesses ensures that **(i)** the consent of the individual must be obtained prior to the collection, use or disclosure

[121] **Federal government institutions**: the *Privacy Act*, S.C. (1985), c. P-21, s. 4 (interpreted in the Clearview case); **Provincial public bodies**: the *Act respecting Access to documents held by public bodies and the Protection of personal information*, s. 64. **(Québec)**. Bill 64 passed by the National Assembly of Québec now provides that such collection must be preceded by an assessment of privacy-related factors and in accordance with an agreement with the *Commission d'accès à l'information* (CAI - Québec's Privacy Commissioner).

[122] *M.L.* c. *Gatineau (Ville de)*, 2010 QCCA168.

[123] **Federal government institutions**: the Privacy Act, S.C. (1985), c. P-21, s. 7. **Provincial public bodies**: the Act respecting Access to documents held by public bodies and the Protection of personal information, s. 65.1. **(Québec)**

[124] **Federal government institutions**: the *Privacy Act*, S.C. (1985), c. P-21, s. 5(1)(2); Benyekhlef and Déziel, 318

[125] **Federal government institutions**: the *Privacy Act*, S.C. (1985), c. P-21, s. 8(2); **Provincial public bodies**: the *Freedom of Information and Protection of Privacy Act* (FOIP), R.S.A. ch. F-25, s. 34 **(Alberta)**; the *Freedom of Information and Protection of Privacy Act* (FIPPA), R.S.B.C. 1993, ch. 165, s 26 **(British Columbia)**; in Québec, this rule 'does not apply to judicial inquiries or to any investigation or report made by a body responsible by law for the prevention, detection or repression of crime or statutory offences', *Act respecting Access to documents held by public bodies and the Protection of personal information*, s 65 5th para. **(Québec)**

[126] For example, the *Freedom of Information and Protection of Privacy Act*, C.C.S.M., c. F175, s 44(r) provides that a public body may disclose personal information without consent for 'law enforcement purposes or crime prevention' **(Manitoba)**; the *Freedom of Information and Protection of Privacy Act*, R.S.B.C. 1993, c. 165, s 33.2(i): 'A public body *may* disclose personal information referred to in section 33 inside Canada as follows: (i) to a public body or a law enforcement agency in Canada to assist in a *specific* investigation' and s 33.1 (1)(t) (interpreted in the *Denham* case): 'to comply with a subpoena, a warrant or an order issued or made by a court, person or body in Canada with jurisdiction to compel the production of information.' **(British Columbia)**; the *Act respecting Access to documents held by public bodies and the Protection of personal information*, s. 67: 'A public body may, without the consent of the person concerned, release personal information to any person or body if the information is necessary for the application of an Act in Québec, whether or not the law explicitly provides for the release of the information' and recently passed Bill 64 provides in particular that such release must be under an Act that (1) '. . . explicitly provides for the release of the personal information' and (2) if 'the Act does not explicitly provide for the release of the information, the information is released on an ad hoc basis and, if personal information concerning any other person is also released, such information concerns only a limited number of persons'. **(Québec)**

of personal information,[127] with some exceptions, such as where the information is publicly available.[128] Otherwise, personal information held by a private body may be disclosed to the police if the police have obtained the legal authority to obtain it, such as a warrant issued by a judge.[129] The federal Act also provides that **(ii)** an 'organization may collect, use or disclose personal information only for purposes that a reasonable person would consider are appropriate in the circumstances'[130] and that such information must be protected.[131] **Specific protection for biometric data –** Biometric data, such as those in an image of a face, are generally characterized as sensitive and constitute personal information.[132] In Canada, the only legislation directly applying to biometric data is the Act to establish a legal framework for information technology (hereinafter 'ALFIT'), which applies in the Province of Québec. A private or public body that creates a database linked to a biometric system must not only **(i)** 'obtain the express consent of the person concerned' for their identity to be verified or confirmed with the help of a FR system (in line with ALFIT s. 44) and also **(ii)** 'disclose the creation or existence of the biometrics system

---

[127] **Federal legislation on private organizations**: Canada's *Personal Information Protection and Electronic Documents Act*, S.C. 2000, c. 5, s 5(1) and principle 3 of Schedule 1, which applies to the private sector is the province has not passed legislation governing the private sector that is 'essentially similar'. **Specific provincial legislation governing private bodies and replacing the federal law in those provinces**: the *Act respecting the protection of personal information in the private sector*, CQLR, c. P-39.1, s 14 and the *Act to establish a legal framework for information technology*, CQLR, c. C-1.1 (**Québec**); the *Personal Information Protection Act*, S.B.C. 2004, c. 63 **(British Columbia)**; the *Personal Information Protection Act*, S. A. 2003, c. P-6.5 **(Alberta)**. **Not considered as essentially similar to the federal legislation**: the *Privacy Act*, S.N.L. 1900, c. P-22 **(Newfoundland and Labrador)**; the *Privacy Act*, C.P.L.M., c. P125 **(Manitoba)**; the *Privacy Act*, R. S.S. 1978, c. P-24 **(Saskatchewan)**

[128] **Federal legislation on private organizations**: Canada's *Personal Information Protection and Electronic Documents Act*, S.C. 2000, c. 5, s. 7(1)d) (interpreted in the *Clearview* case). **Specific provincial legislation governing private bodies**: *Personal Information Protection Act*, S.B.C. 2004, c. 63, art. 12(1)*e*), 15(1)*e*) et 18(1)*e*) **(British Columbia)**; the *Personal Information Protection Act*, S. A. 2003, c. P-6.5, s. 14*e*), 17*e*) et 20*j*) **(Alberta)**

[129] **Federal legislation on private organizations**: *Personal Information Protection and Electronic Documents Act*, S.C. 2000, c. 5, s. 7(3)(c.1) (interpreted in the *Spencer* case). **Specific provincial legislation governing private bodies**: *Act respecting the protection of personal information in the private sector*, CQLR, c. P-39.1, s. 18(3): 'A person carrying on an enterprise may, without the consent of the person concerned, communicate personal information contained in a file he holds on that person . . . (3) to a body responsible, by law, for the prevention, detection or repression of crime or statutory offences who requires it in the performance of his duties, if the information is needed for the prosecution of an offence under an Act applicable in Québec;' **(Québec)** In this case, this section must be read in light of the Supreme Court's interpretation in *R.* v. *Spencer,* (2014) 2 S.C.R. 212 of s 7(3)(c.1) of the federal Act. Enterprises have a fiduciary duty to their clients and must protect their personal information. Consequently, they cannot voluntarily disclose such information if the police has not obtained a warrant from a court. In sum, section 18(3) must be read so as to allow an enterprise to disclose information without the consent of the person concerned when the Québec police requires it to do so in the—*lawful*—performance of its functions, that is, in accordance with the rule of law and the right to privacy under the *Charter* and with a proper warrant issued by a judge ahead of time.

[130] **Federal legislation on private organizations**: *Personal Information Protection and Electronic Documents Act*, S.C. 2000, c. 5, s. 5(3).

[131] **Federal legislation on private organizations**: Ibid., s 5(1) and the 10 principles in Schedule 1.

[132] Clearview AI Investigation; Special Report on FR Technology.

to the CAI' (in line with ALFIT s. 45). The CAI can prohibit the use of the database, require its destruction or order changes. It has also been noted that 'any secondary information revealed by biometric characteristics about an individual cannot be used as a basis for a decision concerning that person'.[133] In July 2020, the CAI published principles and obligations that are binding on public bodies and enterprises when they use biometric databases. The principles are structured around three themes: preliminary analysis and proportional collection, declaration to the CAI and express consent.[134] Special protection with regard to AI technology – In Québec, Bill 64, which was passed in September 2021, created two new forms of protection with regard to AI technologies used by public bodies. First, the individual concerned must be notified if the public body uses profiling technology and must be informed of the means available to deactivate it.[135] Second, a 'public body that uses personal information to render a decision based exclusively on an automated processing of such information must inform the person concerned accordingly not later than at the time it informs the person of the decision'. If the person so requests, the public body or enterprise must disclose other information on the functioning of the AI tool.[136] At this point in time, we do not know how these forms of protection will be applied to guide Québec police officers' use of AI prediction, surveillance or FR tools.

Collection of information on the Internet. By the public body directly – In an earlier special investigation from 2013—the Blackstock Case—the Office of the Privacy Commissioner had already ruled that information accessible on a personal Facebook page was personal information protected by the Privacy Act: 'Under the Act, restrictions on the collection of personal information apply, whether the personal information is available publicly or not.'[137] This meant that the public bodies targeted by the investigation, namely, the Department of Justice Canada and Aboriginal Affairs and Northern Development Canada, could not collect the information available on the personal Facebook page of an Indigenous activist because that information was not directly related specifically to those bodies' programs or activities: 'not obviously relevant to policy development by AANDC, as the department contended, or to the human rights lawsuit with which the Department of Justice was particularly concerned'.[138] The Office of the Privacy

[133] The Special Report on FR Technology gives as an example contextual information that an FR tool finds by following hyperlinks associated with images to the Internet addresses from which the images were scraped.

[134] Commission d'accès à l'information du Québec, 'Biométrie : principes à respecter et obligations légales des organisations. Guide d'accompagnement pour les organismes publics et les entreprises', (*CAI website*, 2020) <https://www.cai.gouv.qc.ca/biometrie/pour-davantage-dinformation/> accessed April 2022.

[135] Act respecting Access to documents held by public bodies and the Protection of personal information, art. 65.0.1.

[136] Ibid., s. 65.2.

[137] Office of the Privacy Commissioner of Canada, 'Aboriginal Affairs and Northern Development Canada wrongly collects information from First Nations activist's personal Facebook page', (*OPCP website*, 29 October 2013): <https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-federal-institutions/2012-13/pa_201213_01> accessed April 2022.

[138] Ibid.

Commissioner also saw the bodies' actions as seeming 'to violate the spirit, if not the letter, of the Privacy Act' because they violated the principle of transparency underlying that Act. From this, we infer that the collection was done without notifying the activist and without directly contacting her.[139] **By a public body from a private organization –** Consistent with this decision, the Privacy Commissioner reiterated in the Joint Investigation of Clearview AI, Inc., that personal information available on the Internet is not public information; it must be remembered that only 'publicly available' information can be collected without consent.[140] It was therefore decided that a private organization, such as Clearview AI. Inc, could not legally collect images from the Internet to feed their RF tool without the consent of the individuals concerned. A fortiori, a public body, such as the RCMP, could also not seek to obtain information from a private third party, given that the latter had collected its data illegally. Such collection would be in violation of section 4 of the federal Privacy Act, R.S.C. 1985, c. P-21, which requires federal public bodies to collect only information that is directly related to their programs or activities. The Office of the Privacy Commissioner interpreted section 4 to mean that a public body's collection must be related to programs or activities that are 'lawful' in order to ensure that the public body actively upholds the unwritten constitutional principle of the rule of law: 'To find otherwise would be to permit government institutions to advance their mandates while rewarding organizations whose personal information collection practices are unlawful, including non-compliance with Canadian privacy laws.'[141] This constitutional principle is expressly enshrined in the Constitution Act, 1982, and has been recognized by the Supreme Court as an implicit part of the preamble to the Constitution Act, 1867.[142] A federal public organization, such as the RCMP, is therefore required to ensure that the personal information collection practices of a private third party with which it wants to do business are legal. This obligation limits the RCMP's general investigative powers.[143] Given that the public organization must proactively uphold the rule of law, it is also required to assess the risks of FR technology and ensure that it complies with the principles of common law and the rights and freedoms provided under the Canadian Charter.[144]

Data sharing between organizations: the problem of trans-functionality. Today, institutional functions are converging. For example, police are called upon to play 'community' and 'social' roles, and social services and the health system are intertwined. Consequently, partnerships are increasingly frequent among public bodies and between public bodies and private third parties. This raises the problem of data sharing between organ-

[139] Ibid.

[140] Joint investigation of Clearview AI, Inc., [44]–[47].

[141] Special Report on FR Technology, [22] and [26]–[27].

[142] British Columbia v. Imperial Tobacco Canada Ltd., 2005 SCC 49, [57]; Re Manitoba Language Rights, [1985] 1 S.C.R. 721.

[143] Special Report on FR Technology, [26], citing s. 18 of the *Royal Canadian Mounted Police Act* and s. 14(1)(*a*) of the *Royal Canadian Mounted Police Regulations.*

[144] Special Report on FR Technology, [42].

izations with widely different functions. When we look at current and reasonably foreseeable police practices, we see problems related to the use of an AI tool processing a private database or information originally held by other public organizations (such as the health system, social services, the child protection system, etc.).

Public-Public Sharing - In general, a public body may share personal information with another public body, without the consent of the individual, if the sharing is for the purposes of the original collection, if it is for a use that is consistent with the purposes of the original collection or if the sharing is authorized by other legislation.[145] In other words, the use of personal information must be limited to the purposes for which it was collected.[146] However, each provincial statute provides for very specific situations where information sharing is permitted.[147] According to the Citizen Lab researchers, the design of AI tools for predictive policing that focus on a cross-sectoral approach (Hub model), such as those possibly being developed by the Saskatchewan Police's SPPAL, represents an increased risk of sharing personal information that is sometimes very sensitive and was not originally collected for that purpose. There is also a risk that the sharing could be counterproductive as it could undermine public confidence in public social services and deter those who need them from using them.[148] Other researchers in Canada are also concerned that personal information held by public bodies (e.g., 'passports, visas, work, study or driver's licences') will eventually be recycled for use in police investigations.[149] In this regard, as part of a special investigation by the Office of the Information and Privacy Commissioner for British Columbia in 2012, Commissioner Denham expressed concern about the informational recycling of the province's driver's licence database by police for police investigations. Without the police having had to obtain a warrant and without even any specific leads on the identities of the persons sought, the province's automobile insurer apparently voluntarily allowed the police to use the province's entire database of driver's licence information and the public insurer's FR technology in order to identify individuals suspected of vandalism. Naturally, the Commissioner found that this method violated the Freedom of Information and Protection of Privacy Act.[150] It was a use of driver's licence holders' personal information that was different from and inconsistent with the purpose for which the public insurer originally collected it.[151] Since the FR technology in question gave police access to the entire database of the province's public insurer, the data's purpose was hijacked by the use made of it by the police. The police

---

[145] Benyekhlef and Déziel, p. 320-322. **Federal public institutions**: *Privacy Act*, R.S.C. (1985), c. P-21, s 8(2)

[146] **Federal public institutions**: Privacy Act, R.S.C. (1985), c. P-21, s. 7. **Provincial public bodies**: Act respecting Access to documents held by public bodies and the Protection of personal information, ss. 65.1 and 65.3. **(Québec)**

[147] Benyekhlef and Déziel, Table 4.6, 324–325.

[148] Citizen Lab, 81–83.

[149] Céline Castets-Renard, Émilie Guiraud and Jacinthe Avril-Gagnon, *supra* note 48, 41. For example, in Québec, this could violate s. 65.1 of the *Act respecting Access to documents held by public bodies and the Protection of personal information.*

[150] Information and Privacy Commissioner, Investigation Report F12-01. Investigation into the use of facial recognition technology by the insurance corporation of British-Columbia, [2012] B.C.I.P.C.D. No. 5.

[151] Ibid., [111].

could not lawfully use it without the 'necessary legal authority', such as a warrant obtained from a judge.[152] Generally, the laws dealing with the disclosure of information between public bodies only allow for the voluntary collaboration of the public body with the police when the latter make a single request for information on a user previously identified in the context of a specific investigation.[153] In this case, the police used the entire database without even having a lead on the identity of the vandals and relied on the database to provide them with one; this was tantamount to investigating the entire population of licensed drivers in the province solely on the basis of the fact that they had drivers' licences.

Private-Public Sharing – In the event that a public body, such as a police force, were to partner with a private organization to develop an AI tool for predictive policing, information sharing would be limited under specific laws governing private organizations: 'Commercial privacy legislation in Canada . . . also does not authorize disclosure without consent to law enforcement unless law enforcement has "lawful authority" to access the information'.[154] In keeping with this principle, it was decided that, in order to provide Canadians with a degree of anonymity when they were using the Internet, it would have to be accepted that Internet users had a reasonable expectation of privacy with regard to their IP address when combined with their name, phone number or street address. Thus, a private company could not voluntarily disclose this personal information to the police without a judicial warrant, at the risk of violating the protection of informational privacy protected by section 8 of the Charter. This was the conclusion of the Supreme Court in Spencer.[155] Section 7(3)(c.1) of the federal Personal Information Protection and Electronic Documents Act, S.C. 2000, c. 5, must be interpreted as requiring private organizations, such as Internet service providers, to shoulder a fiduciary duty to protect the privacy of their customers.[156]

Conclusion. Despite the vagueness of the current legal framework applying to AI tools, police officers have a substantive obligation to ensure respect for the rule of law and for privacy. This duty is based on the constitutional requirement that all privacy violations must be based on authorization and judicial review: 'Allowing law enforcement agencies to access data they could not constitutionally obtain, through a private company that obtained the data lawfully, could represent an unconstitutional expansion of the state's ability to monitor and track individuals without justification or judicial oversight.'[157] Despite the desire to develop more effective prevention strategies based on intersectorality, trans-functionality and partnerships between institutions, it must be understood that, in the eyes of those concerned, these institutions and companies have different functions

[152] Citizenlab, 81; Céline Castets-Renard, Émilie Guiraud and Jacinthe Avril-Gagnon, *supra*, note 48, 84. As provided in s. 33.1 (1)(t) *Freedom of Information and Protection of Privacy Act*, R.S.B.C. 1993, c. 165 **(B.C.)**
[153] 33.2(i) Freedom of Information and Protection of Privacy Act, R.S.B.C. 1993, c. 165 **(B.C.)**
[154] Citizen Lab, 84.
[155] *R.* v. *Spencer*, (2014) 2 S.C.R. 212
[156] Benyekhlef and Déziel, 188.
[157] Citizen Lab, 84.

and the barriers between them are, in the stakeholders' imaginations, hermetically sealed. Thus, the consent given to one of these organizations may not correspond to the new purposes for which one wishes to redeploy the information.

3.1.2 Constitutional protection of privacy

Constitutional protection of privacy flows from section 8 of the Charter: 'Everyone has the right to be secure against unreasonable search or seizure.' Where there is a reasonable expectation of privacy on the part of the state, police officers are required to obtain a warrant issued by a judge. The warrant is issued on the basis of reasonable grounds to believe that an offence has been or will be committed and that information relating to that crime may be obtained as a result of the violation of privacy (ss. 487; 487.01; 184.2(3)).[158] In Canadian law, a search conducted without a warrant is presumed to be unreasonable and it is then up to the Crown to rebut the unreasonableness on a balance of probabilities.[159]

Compliance with the principle of proportionality. Obtaining a warrant makes it possible to limit the violation to what is necessary and reasonable for the purposes of the collection. This requirement directly concerns the harmonization of certain rivalling rights: 'The task of any section 8 analysis is to balance competing values: individual interests and rights against collective preferences and desire for security.'[160] It makes it possible to meet the standard of proportionality that must govern any state infringement of individual rights and freedoms (s. 1 of the Charter): 'Seeking warrants and court authorizations can assist with ensuring that a proposed FRT use meets the proportionality standard.'[161] Similarly, in Commissioner Denham's inquiry referred to above, the requirement to obtain a judicial warrant for the use of FR technology and the database held by the province's public insurer ensures that: 'any change in [the initially planned] use of this magnitude is proportional to the public good served by the infringement on privacy rights of citizens'.[162]

Expectation of privacy: Beyond the traditional public/private frontier. Over the past few decades, the traditional boundary that separated 'private' and 'public' space has been fading away, in part due to better performance of surveillance tools. For this reason, policy makers have needed a new way to enforce privacy protections. The Supreme

---

[158] Special Report on FR Technology, paras. 42-48; *Hunter* v. *Southam Inc.*, [1984] 2 S.C.R. 145, [27]–[29] concerning a warrant as a precondition for the validity of search and seizure. See also, *R.* v. *Collins*, [1987] 1 S.C.R. 265, [34].

[159] *Hunter* v. *Southam Inc.*, [1984] 2 S.C.R. 145; *R. v. Collins*, [1987] 1 S.C.R. 265.

[160] Lee-Ann Conrod, 'Smart Devices in Criminal Investigations: How Section 8 of the Canadian Charter of Rights and Freedoms Can Better Protect Privacy in the Search of Technology and Seizure of Information' (2019) 24 *Appeal: Rev Current L & L Reform* 115, 122; Citizen Lab, 78.

[161] Special Report on FR Technology, [42]–[48]: 'Before using such a service, a police body must, at a minimum, examine whether such a service is reasonably necessary to the investigation and consider the proportionality of the intrusion against the specific public interest being pursued.'

[162] Information and Privacy Commissioner, *supra*, note 150, [113].

Court has recognized that contemporary everyday life requires a certain degree of anonymity, even when acting 'in public', 'in plain view', as when on the Internet.[163] Although the reasonable expectation of privacy is a form of protection that is variable, it should presumably apply to information shared on social media that is collected by AI technology. Such heightened protection would be justified by the highly intrusive nature of these technologies, given their efficiency, the intimate detailed nature of the biographical and historical information found on the Internet and the importance we collectively place on protecting it—as evidenced by our privacy laws.[164] In this regard, the Office of the Privacy Commissioner, in its Clearview AI Investigation found that an FR tool "based on the systematic extraction and processing of billions of images of individuals innocent of any crime, is a major and substantial intrusion by the state into the private lives of Canadians'.[165] We believe these lessons apply equally to other AI-powered surveillance technologies, as the Citizen Lab researchers explain:

> The aggregation and analysis of metadata and other open-source electronic records without judicial oversight could provide questionable access to information that the Supreme Court has said cannot be obtained through direct means. Privacy safeguards, including prior judicial authorization, are therefore necessary when law enforcement agencies collect and analyze content and metadata that is captured from online platforms or other environments where individuals operate freely with relative anonymity.[166]

Considering social media data as 'public' resources available to all law enforcement bodies for surveillance or investigation is circular reasoning: crime is presumed and social media surveillance is meant to justify the violation of privacy ex post facto in the name of effective preventive criminal law enforcement. This circular reasoning also violates the spirit of the presumption of innocence protected by the Charter (s. 11(d)). What must be preserved by section 8 of the Charter is the need for any invasion of privacy to be based on 'reasonable grounds to believe' that an offence has been or will be committed and that information related to the crime may well be found as a result of the violation: 'Pre-emptive fishing expeditions [on social media] could hardly satisfy that standard.'[167]

Interception of private communication in real time. A fortiori, we add our voices to those of Citizen Lab to express concerns about the practices made possible by AI systems like the ICAC Child On-line Protection System. According to Citizen Lab, this system makes it possible to scan and store in real time the information in certain private chat rooms, which would be a violation of Criminal Code sections 184 (interception of communications) and 193 (disclosure of information). In order to comply with section 8 of

[163] Ibid., [42]; *R.* v. *Spencer*, 2014 SCC 43, [43]–[47]; *R.* c. *Jarvis*, 2019 SCC 10, [41].
[164] Special Report on FR Technology, [42]. See also *R.* c. *Jarvis*, 2019 SCC 10, [41].
[165] Special Report on FR Technology, [46].
[166] Citizen Lab, 77.
[167] Citizen Lab, 78.

the Charter, it seems necessary to obtain a warrant under Criminal Code sections 185 and 186 (authorization).[168]

3.2 Equality rights and protection from discrimination: encoding, prolongation and spreading of systemic discrimination by AI tools

Section 15 of the Charter guarantees everyone the right to equal protection and equal benefit of the law. Criminal law must apply equally to everyone, and individuals cannot be discriminated against on the basis of any of the grounds enumerated in section 15 or on grounds similar to those enumerated in section 15. According to the Citizen Lab researchers, constitutional protection against discrimination extends to federal, provincial and municipal police officers' actions and conduct, and even the manner in which they perform their investigations: 'For instance, section 15 and human rights legislation would likely apply where a policing policy relies on a biased algorithm, or where an algorithmic "prediction" contributes to an officer discriminating against a member of a marginalized community.'[169] A person who has been discriminated against under the Charter because of police conduct could then claim damages under s. 24(1) of the Charter.[170] In *Elmardy* v. *Toronto Police Services Board*, the fact that there was no other explanation for the wrongful detention and search of a Black person other than the conscious or unconscious racial prejudice of the two police officers allowed the plaintiff to claim compensatory and punitive damages from the state for the infringement of her right to the equal protection and benefit of the law under s. 15 of the Charter.[171] In Doe v. Metropolitan Toronto (Municipality) Commissioners of Police, a woman was awarded financial compensation for a violation of her s. 24(1) Charter rights because the police used their discretion in the manner in which they conducted their investigation, which was negligent and discriminatory because of gender bias.[172] In this case, her right to liberty and security of the person (Charter s. 7) and her right to non-discrimination (Charter s. 15) were found to have been violated by the conduct of the police officers.

Use of AI tools by the police raises several issues regarding the right to equality. Owing to the high processing capacity of the tools, AI algorithms can prolong, normalize and even spread the past history of discrimination experienced by certain communities. This is especially true when the algorithm is fed with or designed on the basis of historical

---

[168] Citizen Lab, 60 and the footnote 248.

[169] Citizen Lab, 104 referring to *Elmardy* v. *Toronto Police Services Board*, 2017 ONSC 2074 and *Doe* v. *Metropolitan Toronto (Municipality) Commissioners of Police*, 39 OR (3d) 487, 160 DLR (4th) 697, 126 CCC (3d) 12.

[170] *Vancouver (City)* v. *Ward*, 2010 SCC 27. One author, while recognizing this possibility, points out the complexity and difficulty of making such a request: Gabriella Jamieson, 'Using Section 24(1) Charter Damages to Remedy Racial Discrimination in the Criminal Justice System', (2017) 22 *Appeal: Review of Current Law and Law Reform* 71, 87. See also Ranjan Agarwal and Joseph Marcus, 'Where There is no Remedy, There is No Right: Using Charter Damages to Compensate Victims of Racial Profiling' (2015) 34-1 *NJCL* 75, 89.

[171] *Elmardy* v. *Toronto Police Services Board*, 2017 ONSC 2074, [20], [23] and [40].

[172] Doe v. Metropolitan Toronto (Municipality) Commissioners of Police, 39 OR (3d) 487, 160 DLR (4th) 697, 126 CCC (3d) 12.

data that may be biased or that was collected in the context of discriminatory practices.[173] Even when race, gender and other prohibited distinguishing characteristics are not expressly written into it, an algorithm may produce discriminatory results if it takes into account 'proxies', i.e. a factor other than the prohibited ground of discrimination but which, being strongly linked to the prohibited distinguishing characteristic, acts as a substitute for it, as if the prohibited ground were taken into account.[174] Certain communities, already over-represented in police and court data in Canada, are likely to become the primary targets of police intervention as a result of these predictive policing tools, and would be subject to increased, unreasonable, unwarranted surveillance.[175] More surveillance will lead to more arrests, which, once the data from those arrests is fed into the algorithm, will lead to even more surveillance directed at those populations or neighbourhoods (the ratchet effect).[176]

3.3 Right to protection from unreasonable detention and arrest: the establishment of generalized suspicion

We fear the insidious effects that crime prediction tools, because of their scientific aura, can have on police officers' judgments and interventions with suspects. How can we distinguish between reasons for the detention or arrest of a suspect that are based on the independent judgment of the police officer and those that are based on a prediction by an AI tool? Section 9 of the Charter guarantees the protection of all against arbitrary arrest or detention. An arbitrary intervention is one carried out in the absence of reasonable grounds. Can an AI tool's prediction be considered 'reasonable grounds' for intervention and can it be used as one of the legal reasons for a police officer's intervention with regard to a suspect?

Detention for the purposes of an investigation. To begin with, detention for the purposes of an investigation must not be 'arbitrary'. It must be based on 'reasonable suspicion'. To be reasonable, suspicion must be based on verifiable, objective facts. In contrast, a police intervention would be arbitrary and unreasonable if it were based on inaccurate, biased data or data collected in the context of discriminatory practices: 'a detention based on racial profiling is one that is, by definition, not based on reasonable suspicion'.[177]

---

[173] Citizen Lab, 104-106. Use of AI tools raises other concerns related to equality rights: **(i)** some social groups are under-represented in the data gathered by certain public social services, which raises fears when AI tools are used across institutions (for example, SPPAL's HUB model and the Edmonton police's CSA), Ibid., 122; **(ii)** the very choice of the crimes targeted by AI tools can be discriminatory (for example, it has been noted that the tools used by Canadian police focus more on street and property crime than on environmental and financial crime), Ibid., p. 115; and **(iii)** we also fear the encoding, prolongation and spreading of *designers*' unconscious bias and prejudice, Ibid., 120–121.

[174] A. Christin, *supra*, note 114, 280–281.

[175] Citizen Lab, 107 and 109.

[176] A. Christin, *supra*, note 114, 280; Bernard E Harcourt, *Against Prediction: Profling, Policing, and Punishing in an Actuarial Age* (Chicago, IL: University of Chicago Press 2006), 3.

[177] *R.* v. *Le*, 2019 SCC 34, [77]–[78].

Knowing that the predictions of AI tools may be generated from inaccurate data,[178] that historical data may have been collected in the context of discriminatory practices, that because of their processing methods AI tools can multiply these biases tenfold and that the algorithm itself could reflect the discriminatory biases of the designer, it seems difficult to conceive how police officers who use such tools in their daily practice, even if they say they do not rely exclusively on such tools, could base their interventions on anything other than biases stemming from the functioning of these tools and biases stemming from the data they process.[179]

While the grounds for reasonable suspicion need only objectively raise the possibility of criminality, the grounds supporting reasonable suspicion cannot be 'innocuous factors'. An innocuous factor is one that goes 'both ways' and does not necessarily indicate that the person is engaged in a specific criminal activity. It has been argued that the mere combination of several innocuous factors will not result, through 'a kind of alchemy', in grounds capable of supporting a reasonable suspicion of criminality if those factors do not reinforce each other to the point of indicating a possibility that the individual in question is engaged in criminal activity.[180] Because they combine a variety of general factors, external to the suspect or external to the particular context of the police investigation, with sometimes overtly innocuous factors (as per the marketing promise that AI tools have the ability to discover hidden patterns from facts that have no apparent logical connection to the human observer), it seems the algorithms very often work as if they were performing a kind of 'alchemy' on innocuous facts. It turns out that the possibility of criminality indicated by AI tools only makes sense within its complex method of calculation and says nothing to the experienced police officer. Very often, experienced police officers' consideration of the multiple decontextualized factors taken into account by the algorithm would not arouse any reasonable suspicion. The need to refer to the police officer's common sense and practical experience in order to establish the reasonableness of the suspicion is also emphasized by the Supreme Court:

Assessing whether a particular constellation of facts gives rise to a reasonable suspicion should not — indeed must not — devolve into a scientific or metaphysical exercise. Common sense, flexibility, and practical everyday experience are the bywords, and they are to be applied through the eyes of a reasonable person armed with the knowledge, training and experience of the investigating officer. (Our emphasis.)[181]

Even more substantively, the assessment of the reasonableness of the suspicion must relate to 'the extent to which the interference with individual liberty is necessary to perform the officer's duty',[182] which means that it has to be taken into consideration that

[178] *R.* v. *Bernshaw*, [1995] 1 SCR 254, which specifies that detention or arrest based on inaccurate data is necessarily unreasonable.
[179] Citizen Lab, 125 and 127.
[180] *R.* v. *Urban*, 2017 ABCA 436.
[181] *R.* v. *Mackenzie*, 2013 SCC 50, [73].
[182] *R.* v. *Mann*, 2004 SCC 52, [34].

'[i]ndividual liberty interests are fundamental to the Canadian constitutional order'.[183] A designer setting up an algorithm according to this teaching cannot avoid imposing on the police officer who will use the algorithm a specific judgment on how to balance these values. Basically, it comes down to the question of whether, collectively and in the name of fighting crime effectively, we accept that innocent people be detained for the purpose of investigation, and that this be done on the basis the outcomes of complex algorithmic methods which, notably because of machine learning, probably end up escaping the understanding of the police officers who use them.[184] Have we, in our collective imagination, come to place so much trust in machines? Do we accept delegation to machines to the point where a machine's reasoning can produce reasonable grounds for impeding individual freedom? As citizens, do we not need to be able to share a certain common form of reasoning with the decision-maker to whom we delegate a part of the power of detention and arrest in order to accept such an obstacle to freedom? In our view, one of the primary conditions for delegating to police officers the power to deprive others of liberty is specifically the possibility of sharing and understanding the reasoning of the persons to whom this power is delegated. This is what is behind the idea that the factors underlying the suspicion must be verifiable and 'objectively discernible', that is, they must be able to be subject to independent judicial review, adversarial proceedings, dialogue.185 It difficult to see how an AI tool's reasoning could possess these qualities, given the technology's transparency issues and the machine learning involved in its operation, which even the designer, not to mention the police officer, may not understand. The AI tool, by virtue of its opacity, shuts down the dialogue about reasonable grounds for deprivation of liberty and how to balance the values involved.

To be reasonable, the suspicion must also be based on the specific characteristics of a suspect and not on the individual's general characteristics, such as things the individual has no control over, the characteristics of the place where the individual is located, or the officer's prejudices about a cultural group. In sum, the police officer's action must be motivated by a clear link between the specific person being detained and a recent or ongoing criminal offence.[186] By their nature, algorithms can only produce inferences, that is, statistical correlations based on compilation of general characteristics: 'Algorithmic policing methods tend to rely on generalized inferences by definition.'[187] For this reason, a mere prediction by an AI tool regarding the likelihood of a crime in a given location

[183] Ibid., [35].

[184] Citizen Lab, 130: 'For example, in discussing the Vancouver Police Department's (VPD) GeoDASH algorithmic forecasting system, S/Constable Ryan Prox shared that VPD officers run their 'algorithm in its machine-learning retraining mode at 3-week intervals; every 3-week interval it rewrites its algorithmic code. That's why we do the independent audits. Because I can't tell you what factors it's weighting according to making the determinations for the boxes. I can tell you if it's doing it accurately, based on where the incidents are taking place, but I can't tell you the "why", and what weighting it's putting on what factors.'

[185] *R.* v. *Chehil*, 2013 SCC 49, [26].

[186] *R.* v. *Mann*, 2004 SCC 52, [34]–[35].

[187] Citizen Lab, 125. See also S. Du Perron and K. Benyekhlef, 'Les algorithmes et l'État de droit', Document de travail No 27, Laboratoire Cyberjustice, June 2021, 19.

cannot establish or be used to establish reasonable suspicion.[188] As the Supreme Court recognized in *R.* v. *Mann*, the 'high crime nature of a neighbourhood is not by itself a basis for detaining individuals'.[189] Consequently, the deployment of police forces on the basis of a prediction of criminality by algorithmic calculation is likely to taint, at the source, the reasonableness of the suspicion and to cultivate, because of the algorithm's suggestion and its scientific aura, suspicion even before its natural formation in the police officer's mind. The aura of 'objectivity' that accompanies the AI tool's suggestion is likely to impinge on the degree of discretion necessary for an experienced police officer to make the right decision in the context. The subjectivity and discretion in police decision-making, which AI tools seek to combat, ultimately prove necessary for the proper functioning of our law enforcement system:

> The Supreme Court recognizes that police discretion is an essential feature of the criminal justice system. As Justice La Forest wrote in R v Beare, eliminating police "discretion would be unworkably complex and rigid." . . . Police discretion requires both rational justification that is proportionate to the seriousness of the conduct and exercising discretion in the public interest. . . . Whether and to what degree police officers should maintain their discretion when relying on predictive technologies involves a host of policy considerations. . . . While predictive technologies are theoretically capable of injecting a degree of objectivity into crime-prevention and policing, they may also serve to amplify and perpetuate existing practices that further marginalize over-policed groups.[190]

Arrest without warrant. In order for an arrest without a warrant to be not arbitrary, the police officer must have 'reasonable grounds to believe' that the person has committed or is about to commit a crime (Criminal Code s. 495(1)(a)) 'Reasonable grounds to believe' refers to a subjective belief that must be based on objectively justifiable facts that would allow a reasonable person to believe that the individual is involved in a criminal act.[191] In the preceding part, we saw several issues inherent to the special way predictive AI tools work; those issues arise here also.

In substance, what we are looking for is the rationale for delegating to police officers the power to arrest. What justifies such a delegation, if not their capacity to share with the ordinary citizen a certain common form of reasoning. It is precisely the police officer's capacity to understand that society's need for crime protection 'requires that there be a reasonable balance achieved between the individual's right to liberty and the need for society to be protected from crime.'[192] In this sense, the 'objectivity' required to arrest a subject without a warrant must not be confused with a line of reasoning's 'scientific aura' or 'desubjectivization'. The objectivity requirement actually refers instead to a form of

---

[188] Citizen Lab, 125.
[189] *R.* v. *Mann*, 2004 CSC 52, [47].
[190] Michael Purcell and Mathew Zaia, 'Prediction, Prevention And Proof: Artificial Intelligence And Peace Bonds In Canada', (2020) 98-3 *Canadian Bar Review* 515, 541.
[191] *R.* v. *Storrey*, [1990] 1 S.C.R. 241.
[192] Ibid., 249–250.

human intersubjectivity (that takes the form of a dialogue between the police officer and citizens' collective will); this form of human intersubjectivity would be based on objective facts, understood as a reality shared by other humans and therefore also imbued with common social references. When we say that 'the existence of these reasonable and probable grounds must be objectively established', we are saying that '*a reasonable person placed in the position of the officer must be able to conclude* that there were indeed reasonable and probable grounds for the arrest' (our emphasis).[193]

The 'reasonable intersubjectivity' required by the law stands in opposition to the depersonalized, inflexible, mathematical, decontextualized objectivity of AI tools, whose reasoning does not necessarily refer to a shared reality, to the common social sense ('the reasonable person'). Police and criminal law cannot operate without referring to the human person, despite the limitations of human subjectivity; this is because reasonableness with regard to balancing freedom and security is properly social and dialogical. It results from a 'dialogue' between the police officer and citizens' collective will. The balancing of values must result from an attempt by the police officer to grasp this will, and it must be possible, if the police officer's attempt is contested, to submit the officer's interpretation to the adversarial procedure of the trial. This means that no mathematical formula will be able to provide a satisfactory answer in advance of the procedure by which values are balanced, since that process may also evolve over time.

3.4 Other forms of constitutional protection: Procedural equality, full and complete defence, right to remedy and sentence reduction

As we have seen, the exercise of many constitutional rights and the enjoyment of guarantees in criminal matters depend intrinsically on transparency and on disclosure of the source code of the algorithms underlying AI tools. Every accused has the right to a full and complete defence, which obliges the Crown to disclose all evidence to the accused (Charter s. 7).[194] In order for those who are accused to have the legality of their detention tested by habeas corpus (Charter s. 10(c)), to challenge their arrest or detention (Charter s. 9) or to exercise their right to a remedy (Charter s. 24(1)(2)), it seems necessary for them to have access to a certain insight into the functioning of the algorithm that is at the origin of their arrest, detention and prosecution.

Discriminatory or prejudicial conduct by police officers and, by extension, the use of an algorithmic tool that would steer toward discriminatory conduct could also lead to sentence reduction based on the concept of individualized proportionality established by the Supreme Court in *R.* v. *Nasogaluak*.[195] This interpretation of the fundamental principle of sentencing calls for taking into consideration, in order to determine the degree of severity of the sentence, the suffering that has already been inflicted on the accused when in the 'hands of the State' (the conduct of a prosecutor or police officer): 'A Charter breach

[193] Ibid., 250.
[194] *R.* v. *Stichcombe*, [1991] 3 SCR 326.
[195] *R.* v. *Nasogaluak*, 2010 SCC 6.

indicates that the state has offended these values and concerns and a sentence can and should communicate society's resulting condemnation if the breach has a sufficient link to the circumstances of the offence or the offender. . . . His sentence is justifiably reduced because he has already suffered harm at the hands of the state in response to his misconduct. When a judge decides how much and what form of punishment to inflict on the accused, the ways in which he has already suffered is salient.'[196]

[196] For the concept of 'individualized proportionality' in sentencing, which makes it possible to take into account all of the suffering already inflicted by the State on the accused, see Benjamin L. Berger, 'Sentencing and the Salience of Pain and Hope' 70 *Supreme Court Law Rev* 2d 337, who bases his interpretation on *R.* c. *Ipeelee*, 2012 SCC 13, [86] in particular: 'Who are courts sentencing if not the offender standing in front of them? If the offender is Aboriginal, then courts must consider all of the circumstances of that offender, including the unique circumstances described in Gladue.'

Acronyms and abbreviations

AIA - Algorithmic Impact Assessment tool

CAI - Commission d'accès à l'information du Québec

LCO – Law Commission of Ontario

CPIC – Canadian Police Information Centre

CPS - Calgary Police Service

CSA - Community Solutions Accelerator

CIOSC - Chief Information Officer Strategy Council

CIO - Chief Information Officers

RSD - Research and Statistics Division

EPS - Edmonton police service

FR – Facial recognition

GPS - Guelph Police Service

RCMP – Royal Canadian Mounted Police

ALFIT - Act to establish a legal framework for information technology (Québec)

League - Ligue des droits et Libertés

NIJ - National Institute of justice

OPP - Ontario Provincial Police

GPAI - Global Partnership on Artificial Intelligence

SPPAL - Saskatchewan police predictive analytics lab

SPS - Saskatoon Police Service

SPVM - Service de Police de la Ville de Montréal

SQ - Sûreté du Québec

TPS - Toronto Police Service

WRP - Waterloo Regional Police

VPD - Vancouver Police Department

Selected literature

Armony V, et al., 'Les interpellations policières à la lumière des identités racisées des personnes interpellées', (Final report to the SPVM, august 2019), 19–20

Benbouzid B, 'Quand prédire, c'est gérer, La police prédictive aux États-Unis', (2018) 211-5 *Réseaux* 221

Benyekhlef K and Déziel P-L, *Le droit à la vie privée en droit québécois et canadien*, (Montréal: Éditions Yvon Blais 2018)

Castets-Renard C, Guiraud É and Avril-Gagnon J, 'Cadre juridique applicable à l'utilisation de la reconnaissance faciale par les forces de police dans l'espace public au Québec et au Canada - Éléments de comparaison avec les États-Unis et l'Europe' (International Observatory on the Societal Impacts of AI and Digital Technology, Accountable AI in a Global Context Research Chair, 2020)

Christin A, 'Predictive algorithms and criminal sentencing', 283 in N. Guilhot and D. Bessner (eds*), The Decisionist Imagination* (Berghahn Books 2018)

Commission de la sécurité publique de Montréal, 'Rapport sur l'Utilisation par le SPVM de technologies de reconnaissance faciale et de systèmes de reconnaissance de plaques d'immatriculation' (Ville de Montréal, June 2021)

Conrod L-A, 'Smart Devices in Criminal Investigations: How Section 8 of the Canadian Charter of Rights and Freedoms Can Better Protect Privacy in the Search of Technology and Seizure of Information' (2019) 24 Appeal: Rev Current L & L Reform 115, 122; Citizen Lab, 78

Garapon A and Lassègue J, *Justice digitale* (Paris: Presses Universitaire de France 2018) 249

Information and Privacy Commissioner, *Investigation Report F12-01. Investigation into the use of facial recognition technology by the insurance corporation of British-Columbia*, [2012] B.C.I.P.C.D. No. 5

Konina A, 'The Privatization of Law Enforcement: Promoting Human Rights through Procurement Contracts', (2021) 1-1 *McGill GLSA Research Series* 1

Law Commission of Ontario, 'The Rise and Fall of AI and Algorithms in American Criminal Justice: Lessons for Canada', 2020

Ligue des droits et libertés, 'Mémoire : Étude des technologies de reconnaissance faciale et des lecteurs automatiques de plaques d'immatriculation' (LDL website, 30 octobre 2020) <https://liguedesdroits.ca/memoire-reconnaissance-faciale-lapi-csp-montreal-2020/#_ftnref10> accessed April 2022

Office of the Privacy Commissioner of Canada, 'Aboriginal Affairs and Northern Development Canada wrongly collects information from First Nations activist's personal Fa-

cebook page', (OPCP website, 29 October 2013) Aboriginal Affairs and Northern Development Canada wrongly collects information from First Nations activist's personal Facebook page - Office of the Privacy Commissioner of Canada accessed April 2022

Office of the Privacy Commissioner of Canada, "Police Use of Facial Recognition Technology in Canada and the Way Forward", 2021

Office of the Privacy Commissioner of Canada, *Examination of RCMP Exempt Data Banks*, (*OPCC website*, February 2008), <https://www.priv.gc.ca/fr/mesures-et-decisions-prises-par-le-commissariat/verifications/rcmp_080213/> accessed April 2022

Office of the Privacy Commissioner of Canada, *Joint investigation of Clearview AI, Inc*. (PIPEDA Findings #2021-001, 2 February 2021) available: <https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2021/pipeda-2021-001/> accessed April 2022

Purcell M and Zaia M, 'Prediction, Prevention And Proof: Artificial Intelligence And Peace Bonds In Canada', (2020) 98-3 Canadian Bar Review 515, 541

Robertson K, Khoo C and Song Y, *To Surveil and Predict: A Human Rights Analysis of Algorithmic Policing in Canada* (Toronto: Citizen Lab and International Human Rights Program, University of Toronto 2020)

Sanders C B, Weston C and Schott N, 'Police Innovations, "Secret Squirrels" and Accountability: Empirically Studying Intelligence-Led Policing in Canada', (2015) 55-4 *The British Journal of Criminology* 711

Thibierge C, ed., *La densification normative. Découverte d'un processus*, (Paris: Éditions mare & martin 2013), pp. 1123–1124

PREDICTIVE POLICING IN GERMANY

By Johanna Sprenger and Dominik Brodowski[*]

Abstract

In ever more areas, it becomes evident that the transformative power of information technology – and so-called 'artificial intelligence' in particular – affects the administration of criminal justice in Germany. The legal framing of issues relating to the use of 'AI technology' in criminal justice lags behind, however, and is of high complexity: In particular, it needs to take the European framework into account, and has to cope with the German peculiarity that the prevention of crimes by the police is a separate branch of law, which is regulated mostly at the 'Länder' (federal states) level, while criminal justice is regulated mostly on the federal level. In this report, we shed light on the practice, on legal discussions, and on current initiatives focusing on 'predictive policing'.

1 Introduction

German Law does not provide for a legal definition of the term 'predictive policing'. When focusing on the most characteristic function of 'predictive policing', the description 'prediction-based police-work'[1] seems most suitable because it stresses the prognostic element of predictive policing, while not strictly excluding forms that do not entail highly advanced or intelligent technology.[2] Further definitions used are, for example, 'tech-based analytical procedures aiming to predict the probability of future offences, offenders or crime scenes'[3] as well as 'computer-assisted method for spatially based probability calculations of crime'[4], which is focused on probable crime scenes, while the term 'automated generation of suspicion'[5] seems to be slightly better suited to describe techniques focused on probable offenders (even though not necessarily limited to them).

---

[*] Johanna Sprenger is a legal officer with the Federal Ministry of Justice; all views reflected in this article are her own. Dominik Brodowski is Professor of Criminal Law and Criminal Procedure, Saarland University, Saarbrücken, Germany, and is secretary of the German AIDP national group.

[1] Simon Egbert, 'Siegeszug der Algorithmen? Predicive Policing im deutschsprachigen Raum' [2017] A-PuZ 17, 19; Jörg Eisele and Kristine Böhm, 'Potential und Risiken von Predictive Policing' in Susanne Beck, Carsten Kusche and Brian Valerius (eds), *Digitalisierung, Automatisierung, KI und Recht* (Nomos 2020) 519; Tobias Knobloch, 'Vor die Lage kommen: Predictive Policing in Deutschland' (Stiftung Neue Verantwortung and Bertelsmann Stiftung 2018) 9 (translation to English by the authors).

[2] Hans-Heinrich Kuhlmann and Simone Trute, 'Predictive Policing als Formen polizeilicher Wissensgenerierung' [2021] GSZ 103, 104 (translation to English by the authors).

[3] Ines Härtel, 'Digitalisierung im Lichte des Verfassungsrechts – Algorithmen, Predictive Policing, autonomes Fahren' [2019] LKV 49, 54 (translation to English by the authors).

[4] Kai Seidensticker, 'SKALA – Predictive Policing in North Rhine-Westphalia' (2021) 21 European Law Enforcement Research Bulletin 47, 48.

[5] Alexander Baur, 'Maschinen führen die Aufsicht', [2020] ZIS 275, 277; Timo Rademacher, 'Verdachtsgewinnung durch Algorithmen. Maßstäbe für den Einsatz von predictive policing und retrospective policing' in Daniel Zimmer (ed), *Regulierung für Algorithmen und Künstliche Intelligenz* (Nomos 2019) 229, 231 (translation to English by the authors).

It is noteworthy that systems used by private companies as part of their compliance infrastructure are sometimes mentioned in the context of 'predictive policing' as well.[6] This concerns methods of automated fraud detection, or automated risk assessment systems regarding money laundering or insider trading[7] – compliance methods which can include, very generally speaking, risk detection models similar to the ones used by the police or regulatory bodies. For the purposes of this report, however, 'predictive policing' is understood as methods applied by state authorities.

2 Description of AI systems used in Germany for 'Predictive Policing'

2.1 Geospatial systems

Approaches on 'predictive policing' that aim to identify probable crime scenes are, or at least have been for the last years, the most prevalent type in Germany. These predictive software systems are used in order to determine the probability that certain offences, mostly residential burglaries, will be committed within a certain local area. They work theory-based, ie under the criminological assumption that some types of crime occur in certain patterns, that rules can be derived from these patters, and that these rules can then be applied to the available data through the respective software.[8]

2.1.1 PreCobs

'PreCobs' (Pre Crime Observation System) by 'Oberhausener Institut für musterbasierte Prognosetechnik GmbH' is a commercial predictive software and the first one that has been used in Germany. The software is said to be comparable to the US-American system 'PredPol'. It aims to predict the probability of residential burglaries and applies the near-repeat theory, ie the assumption that with regard to certain types of offences, crime events are often followed by a subsequent event of crime in temporal and local proximity, especially in case of professional offenders.[9] The main prognostic feature of 'PreCops' is its assessment whether a burglary has been committed professionally.[10] The software

[6] Lena Rutkowski, 'Predictive Policing am Arbeitsplatz' [2019] NZG 72.

[7] See the findings of a study by the Federal Financial Supervisory Authority (BaFin) regarding the use of Big Data Analysis and AI in regulatory compliance processes, 'Big Data trifft auf künstliche Intelligenz' (BaFin 2018) 76–78, 86–89 and advertising by software providers, eg Capgemini, 'Inventive FRC – Compliance' (2020) <https://www.capgemini.com/de-de/2020/09/inventive-frc-compliance-machine-learning/> accessed 9 August 2022.

[8] Thomas Wischmeyer, 'Predictive Policing, Nebenfolgen der Automatisierung von Prognosen im Sicherheitsrecht', in Andreas Kulick and Michael Goldhammer (eds), *Der Terrorist als Feind?* (Mohr Siebeck 2019) 193, 194; Franziska Lind, *Das raumbezogene Predictive Policing in Deutschland. Der aktuelle rechtliche Rahmen und seine Indikationen für Weiterentwicklungen des Einsatzes prädiktiver Analytik bei präventiv polizeilichem Handeln* (forthcoming).

[9] Silke Krasmann, and Simon Egbert, 'Predictive Policing. Eine ethnographische Studie neuer Technologien zur Vorhersage von Straftaten und ihre Folgen für die polizeiliche Praxis' (final project report University of Hamburg 2019) 27–29.

[10] Simon Egbert, 'Predictive Policing als Treiber rechtlicher Innovation?' (2021) 41 Zeitschrift für Rechtssoziologie 26, 33.

has been in regular use by police departments in Bavaria since 2015/2016, [11] subject to a series of test runs of several months respectively in Baden-Württemberg from 2015 to 2018[12], and has been subject of a pilot project in Saxony/Leipzig from September 2019 to September 2020.[13] Notably, Baden-Württemberg has decided against further implementation of PreCobs in 2019. Bavaria decided as well to end its use for police work in 2021.[14] In both cases, the reasons were similar: there has not – or, as in Bavaria, not anymore[15] – been enough data available for the system to work efficiently.[16] In the course of this, efforts in Bavaria to enhance the software's functions to other types of offences, which were based on an alternative and more complex theoretical approach,[17] have come to a halt as well.[18]

In various federal states ('Länder') of Germany, predictive software models have been developed 'in-house' by the respective police departments:

2.1.2 KLB-operativ

The prognostic system 'KLB-operativ' has been developed by authorities in Hessen and was implemented in 2017. The software is now in use throughout Hessen.[19]

2.1.3 KrimPro

Police authorities in Berlin (with some external support by Microsoft and Oraylis) have developed their system 'KrimPro' (KriminalitätsPrognose)[20] in 2016. KrimPro does not

[11] Knobloch (n 1) 14.

[12] Dominik Gerstner, 'Predictive Policing in the Context of Residential Burglary: An Empirical Illustration on the Basis of a Pilot Project in Baden-Württemberg, Germany' [2018] European Journal for Security Research 115.

[13] *Polizei Sachsen* (Saxon Police Force), 'Archiv abgeschlossener Forschungsprojekte' <https://www.polizei.sachsen.de/de/79682.htm> accessed 9 August 2022.

[14] *Bayerisches Landeskriminalamt* (Bavarian State Criminal Police Office), 'Predictive Policing bei der Bayerischen Polizei' (press release 27 October 2021) <https://www.polizei.bayern.de/aktuelles/pressemitteilungen/018804/index.html> accessed 9 August 2022.

[15] Bavarian State Criminal Police Office (n 14).

[16] Nils Mayr, 'Strobl entscheidet sich gegen PreCobs' *Stuttgarter Nachrichten* (Stuttgart, 3 September 2019) <https://www.stuttgarter-nachrichten.de/inhalt.aus-fuer-die-einbruchvorhersage-software-strobl-entscheidet-sich-gegen-precobs.19a18735-9c8f-4f1a-bf1b-80b6a3ad0142.html> accessed 9 August 2022.

[17] Krasmann and Egbert (n 9) 29.

[18] Bavarian State Criminal Police Office (n 14).

[19] *Hessisches Ministerium des Innern und für Sport* (Hessian Ministry of the Interior and Sport), 'Schwerpunkt-Fahndungsaktion: 564 festgestellte Straftaten und 84 Maßnahmen' (press release 25 November 2021) <https://innen.hessen.de/Presse/Schwerpunkt-Fahndungsaktion-564-festgestellte-Straftaten-und-84-Festnahmen> accessed 9 August 2022 and 'Jahresbilanz 2018', 12, 16 <https://innen.hessen.de/sites/innen.hessen.de/files/2021-10/jahresbilanz_2018_160119_web.pdf> accessed 9 August 2022.

[20] *Berliner Senatsverwaltung für Inneres und Sport* (Berlin Senate Department of Internal Affairs and Sport), LT-Drucks. (Berlin) 18/17562, 562.

only use police data, but can also access publicly available data regarding the demographic structure and infrastructure. It is now being used not only in Berlin but also in Brandenburg and Saxony-Anhalt.[21]

2.1.4 PreMAP

Lower Saxony began developing its software 'PreMAP' (Predictive Policing Mobile Analytics for Police) and started its use in 2017, successively expanding throughout the whole state of Lower Saxony.[22] Lower Saxony has since stopped its deployment, however, among other reasons due to its low cost/benefit ratio.[23]

KLB-operativ, KrimPro and PreMAP are based on the near-repeats hypothesis and focus on residential burglaries.

2.1.5 SKALA

In North Rhine-Westphalia, the respective software system 'SKALA' (*System zur Kriminalitätsanalyse und Lageantizipation* – system for analysis and anticipation of crime) is in operative use since 2018 (starting with individual police stations in urban areas, and successively expanding to rural areas). It stands out due to various reasons. First, it is applied to predict not only residential but also commercial burglary and vehicle-related crime; it is also under consideration regarding further types of crimes. Furthermore, it does not only rely on data regarding previous incidents of crime, but also on socio-economic data such as structural aspects regarding the population, rent and income structure, infrastructure and mobility opportunities within the respective area.[24] Additionally, its theoretical basis extends beyond the near-repeats hypothesis to further criminological and socio-scientific theories.[25]

2.2 Person-based 'Predictive Policing', individual risk assessments, and RADAR-iTE

While all of the above systems are still relatively similar to each other, the situation becomes much more complex regarding predictive methods that do not focus on probable local crime scenes but rather on probable offenders. 'Predictive policing' approaches aiming to apply individual risk assessments to natural persons are very rare in Germany.

---

[21] Stefan Löbel and Tino Schuppan, 'Potentiale und Herausforderungen einer neuen Datenorientierung im Kontext öffentlicher Aufgabenwahrnehmung' (2021) 16 Berichte des NEGZ 16–18.

[22] Kai Seidensticker, Felix Bode and Florian Stoffel, 'Predictive Policing in Germany' (University of Konstanz 2018) 4 <https://kops.uni-konstanz.de/handle/123456789/43114> accessed 9 August 2022.

[23] *Landeskriminalamt Niedersachsen* (Lower Saxony State Criminal Police Office), 'PreMAP – Predictive Policing (Vorausschauende Polizeiarbeit) in Niedersachsen' <https://www.lka.polizei-nds.de/startseite/kriminalitaet/forschung/premap/predictive-policing-in-niedersachsen-das-projekt-premap-114083.html> accessed 9 August 2022.

[24] Seidensticker (n 4) 52.

[25] Seidensticker, Bode and Stoffel (n 22) 5.

The coalition agreement of the parliamentary coalition forming the current German government indicates a very restrictive approach on such systems, as it states that the use of 'scoring' systems by state authorities shall be prohibited by EU law.[26]

As for now, the only system in Germany which is publicly known to focus on specific individuals and their respective risk potential appears to be 'RADAR-iTE' (*Regelbasierte Analyse potentiell destruktiver Täter zur Einschätzung des akuten Risikos – islamistischer Terrorismus* – rule-based analysis of potentially destructive perpetrators for an assessment of their acute risk – Islamist terrorism). RADAR-iTE is a risk-assessment tool developed by the *Bundeskriminalamt* (Federal Criminal Police Office) in cooperation with the Forensic Psychology Working Group of the University of Konstanz. RADAR-iTE serves to assess the risk that individuals – who have already been identified by the police authorities as potentially dangerous from previous law enforcement measures – are willing to commit acts of Islamist-motivated terrorism. It is used by police departments on the federal and Länder level since 2017 as a tool to assess the need for police interventions and to prioritise police resources. After evaluation, the system has been refined to its 2.0 version in 2019 according to scientific, ethical and legal aspects in cooperation with the University of Konstanz and the technical college for police in Saxony-Anhalt.[27] It evaluates both risk-increasing and risk-reducing factors. These factors are provided in a form with question and answers categories that are completed (manually by police officers) on the basis of information that has already been gathered.[28] The system then calculates a risk factor on this basis and assigns it to one of two pre-defined risk-levels, either 'moderate' or 'high'.[29] Even though the calculation itself is processed automatically, it relies on a rather simple model based on the software Microsoft Excel.[30] The legal basis for RADAR-iTE is § 18 (3) in connection with § 18 (1) No 4, § 16 BKAG (Federal Criminal Police Office Act[31]) which does not mention the use of technology but only refers to the 'further processing of personal data' in case of indications that a person concerned is likely to commit a serious crime in the future.

---

[26] Coalition agreement, lines 504–505 <https://www.wiwo.de/downloads/27830022/8/koalitionsvertrag-2021-2025.pdf> accessed 9 August 2022.

[27] Federal Criminal Police Office, 'RADAR (Regelbasierte Analyse potentiell destruktiver Täter zur Einschätzung des akuten Risikos)' <https://www.bka.de/DE/UnsereAufgaben/Deliktsbereiche/PMK/Radar/radar_node.html;jsessionid=9AB1BDE4A134C483F1820378A09EAF6A.live612#doc142872bodyText4> accessed 9 August 2022; for a more detailed description Celina Sonka and others, 'RADAR-iTE 2.0: Ein Instrument des polizeilichen Staatsschutzes, Aufbau, Entwicklung und Stand der Evaluation' [2020] Kriminalistik, 386.

[28] For example, information on social integration (friends and family), access to weapons or explosive devices, military experience, trips to war or crisis zones, affiliation to radical groups, BT-Drucks. 18/13422, 5.

[29] The first version of the system provided a third risk level category 'noticeable', cf BT-Drucks. 19/12859, 9; for further explanation of the risk levels see BT-Drucks. 19/5648, 5 and 66.

[30] BT-Drucks. 19/1513, 7.

[31] Gesetz über das Bundeskriminalamt und die Zusammenarbeit des Bundes und der Länder in kriminalpolizeilichen Angelegenheiten (Bundeskriminalamtgesetz – BKAG), as amended.

So far, RADAR-iTE has been applied with regard to terrorism risks from the Islamist spectrum. Currently, the Federal Criminal Police Office, in cooperation with the Centre of Criminology and the technical college for police of Saxony-Anhalt, is working on an additional version of RADAR-iTE which focuses on terrorism risks motivated by right-wing extremism. The new version is planned to be made available for operative use in the course of 2022.[32]

2.3 Other forms of person-based predictive policing

Notwithstanding the restrictive approach towards individual risk-assessments mentioned above, there are now more and more algorithm-based prediction systems aiming to recognise patterns or other indicators for potential threats or potentially criminal behaviour of individuals not yet known to the state authorities. Some of them may not only aim to identify specific individuals but also dangerous objects or situations such as social media information that indicates tendencies of radicalisation.

2.3.1 Passenger Name Records Data Analysis

The automated analysis according to § 4 (2) No 2 of the Act on the Processing of Passenger Name Record (PNR) Data to Implement Directive (EU) 2016/681 (PNR Act)[33] appears to be one of the most significant examples for person-focused predictive policing based on pattern recognition in Germany.

The PNR Act obliges air carriers to transfer PNR data collected in course of their business (comprising up to 20 categories of data, see the list in § 2 [1] PNR Act) to the Federal Criminal Police Office. The Federal Criminal Police Office processes such data for automated advance checks – either before arrival or departure of the relevant flight – in order to identify individuals previously unknown to the police authorities for whom there is reason to believe that they have committed acts of terrorism or other serious crimes or will do so in the foreseeable future. In the course of these automated advance checks the PNR data are tested against certain databases or so-called 'patterns'. In case this results in a 'match', the Federal Criminal Police Office must individually (ie by human officers) examine the results (§ 4 [2] 2 PNR Act) and may, if necessary, transfer the relevant data to other federal police or security authorities (§ 6 PNR Act).

Patterns indicating that an individual can be associated with terrorism or other serious crimes could include incriminating criteria such as certain itineraries, layovers, payment methods etc. § 4 (3) of the PNR Act sets out basic rules governing the establishment of the patterns by the Federal Criminal Police Office in cooperation with its data privacy officer and other security and police authorities. In order to keep the number of individuals matching these patterns low, the incriminating criteria shall be combined with ex-

[32] BT-Drucks. 19/32271; Federal Criminal Police Office (n 27).

[33] Unofficial translation available at <https://www.gesetze-im-internet.de/englisch_flugdag/index.html> accessed 9 August 2022.

onerating criteria. A person's race or ethnic origin, political opinions, religion or philosophical beliefs, trade union membership, health, sexual life or sexual orientation may not be used for the automated checks under any circumstances.[34] The Federal Commissioner for Data Protection and Freedom of Information shall review the production and use of the patterns at least every two years and report to the Federal Government every two years.

In 2018, the Federal Criminal Police Office began automated checks of PNR data, but limited to databases on persons or objects sought or under alert. In 2019, the German Government repeatedly announced that it did not yet operate advance checks against patterns and that it intended to do so at a later stage.[35] In 2020, all matches generated through advance checks of PNR data were still based on databases for persons or objects sought or under alert and not on checks against patterns.[36] By the time this report was finalised, it could not be confirmed whether automated checks against patterns had been put in place. Even though there is no further public information available on how the patterns are generated from a technical point of view, and how the respective advance checks will be operated in detail, the immense volume of data to be processed and the complexity of potential patterns makes the PNR system seem to be a typical use case for machine learning and big data analysis. This is also, as *Thüne* points out, indicated by the budget that has been assigned for the German PNR system alone (initial costs of € 78 million and yearly costs of € 65 million).[37] Furthermore, § 4 (4) PNR Act allows the analysis of PNR data in order to produce or update patterns; this provides for a legal basis to use such data as training data for the generation of patterns by use of machine learning.[38]

The PNR system is subject to much criticism. Scholars,[39] human rights organisations[40] and the German Federal Commissioner for Data Protection and Freedom of Information[41] have complained about the general and indiscriminate nature of the transfer, automated checking and retention of PNR data affecting people without any link to the crimes the PNR system aims to prevent or investigate, the possibility of large numbers of false positives, the long retention period of PNR data (five years) and that it is – with regard to the automated checks – completely up to the administrative bodies to decide about the design of the patterns.

---

[34] On the discriminatory potential, see **4.2** below.
[35] BT-Drucks. 19/10431, 3 and 19/12858, 3.
[36] Response of the Federal Ministry of the Interior dated 1 February 2021, BT-Drucks. 19/26440, 22.
[37] Martin Thüne, 'Predictive Policing' (2020) 144.
[38] Lucia M Sommerer, Self-imposed Algorithmic Thoughtlessness and the Automation of Crime Control (Nomos 2022) 81.
[39] Clemens Arzt, 'Einladung zur anlasslosen Rasterfahndung durch das BKA' [2017] DÖV 1023.
[40] *Gesellschaft für Freiheitsrechte* (Society for Civil Rights), 'NoPNR: Keine Massenüberwachung am Himmel' <https://freiheitsrechte.org/nopnr-de/> accessed 9 August 2022.
[41] Federal Commissioner for Data Protection and Freedom of Information, '28th Annual Activity Report' (2019) section 6.4, 51.

The legal basis for the PNR systems, meaning both the PNR Directive (on the EU level) as well as the PNR Act (on the national level), are currently being challenged in several civil and public administrative lawsuits and preliminary ruling procedures pending with the European Court of Justice (ECJ). The plaintiffs argue[42] that both the PNR Directive as well as the PNR Act are not in line with Articles 7, 8 and 52(1) of the Charter of Fundamental Rights of the European Union in light of the rulings of the ECJ in the Digital Rights Ireland[43] and the Tele2 Sverige and Watson[44] cases as well as the ECJ's Opinion 1/15 of 26 July 2017 on the EU-Canada Passenger Name Record Agreement.[45]

A recent ruling by the ECJ promises good prospects of success for the plaintiffs. Upon referral by the Belgian Constitutional Court, the ECJ determined strict requirements on how the PNR Directive needs to be interpreted in order to be in line with Articles 7, 8 and 21 and Article 52 (1) of the Charter of Fundamental Rights of the European Union. Among other quite significant aspects that might warrant adjustments to the German PNR Act in its current form, the ECJ sets important boundaries for the establishment of 'patterns' used for automated advance checks, in particular regarding machine learning: First of all, the ECJ held that member states may not operate AI systems using machine learning able to define or modify the criteria for the assessment without a human decision. In that regard, the ECJ warned against black-box effects. It stressed that any individual review of an automatically generated positive match depends on the possibility to understand the reason why the program generated a positive match. Furthermore, the ECJ sets out requirements for the pre-defined criteria in order to guarantee that the automated advance checks work in a non-discriminatory manner.[46] Given that it also recognises a high likelihood of false positives, the ECJ emphasised the importance of an individual, non-automated review of any positive match. According to the ECJ, member states are obliged to lay down clear and precise rules for such individual review and to ensure that the person concerned has an adequate understanding of the automated assessment in order to exercise their rights properly.[47]

2.3.2 Hessen-Data and similar data analytics systems

Another project of Hessen that has attracted significant attention is the analysis system 'HessenData' which has been in use since 2017. HessenData might not (yet) qualify as

[42] See for example the plaintiffs statement regarding Cases C-215/20 und C-220/20 <https://freiheitsrechte.org/home/wp-content/uploads/2020/09/GFF-Stellungnahme-an-den-EuGH-zur-Fluggastdatenspeicherung PNR-Richtlinie-Sept2020.pdf> accessed 9 August 2022.

[43] ECJ, joined cases C-293/12 and C-594/12 Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources and Others and Kärntner Landesregierung and Others ECLI:EU:C:2014:238.

[44] ECJ, joined cases C-203/15 and C-698/15 Tele2 Sverige AB v Postoch telestyrelsen and Secretary of State for the Home Department v Tom Watson and Others ECLI:EU:C:2016:970.

[45] ECJ, opinion 1/15 on the draft agreement between Canada and the European Union on the transfer of Passenger Name Record data ECLI:EU:C:2017:592.

[46] ECJ, case C-817/19 *Ligue des droits humains v Conseil des ministres* ECLI:EU:C:2022:491 para 193–201.

[47] ECJ, case C-817/19 *Ligue des droits humains v Conseil des ministres* ECLI:EU:C:2022:491 para 202–213.

'predictive policing', but clearly has the potential to be used for predictive purposes. It is based on the Palantir software 'Gotham' and rapidly processes information from various heteronomous sources of police data (such as data from different police databases, data requested from communication service providers, data from communication surveillance measures or extracted from electronic devices seized by law enforcement authorities) in order to identify and visualise ('mapping') links and patterns.[48] Its purpose is to provide the police authorities with rapid information that can be used in order to plan police operations and deployment strategies. In contrast to the predictive policing tools mentioned above, the software does not suggest any conclusions or assumptions (eg as to a potential risk or suspicion of crime) based on these findings. The Hessian authorities further stress that the software does not by itself automatically collect and integrate data from external sources such as social media; instead, such data is otherwise retrieved by the authorities and can then be accessed by the system.[49]

With § 25a of the Hessian Act on Public Security and Order (HSOG)[50], Hessen has introduced an explicit legal basis for HessenData limiting its use to the prevention of ('preventive fight against') those serious criminal offences listed in § 100a (2) of the German Act on Criminal Procedure (StPO)[51] or for the prevention of a danger of significant weight in justified individual cases. The provision explicitly allows to automatically identify affiliations or connections between individuals, groups, institutions, objects etc, to filter out irrelevant information and to statistically evaluate new findings and match them with known factual backgrounds. Furthermore, the decision to deploy or significantly change the software lies with the head of police, and before taking such a decision, the data protection officer needs to be consulted (without any veto rights, however).[52]

Meanwhile, other German Länder have also expressed their interest in the deployment of such a software, such as Hamburg which already introduced a legal basis identical to the one in Hessen,[53] or North Rhine-Westphalia which acquired the Palantir software and started using it for testing purposes in 2020.[54] Only recently, the legislator of North

[48] LT-Drucks. (Hessen) 19/6574, 17.

[49] LT-Drucks. (Hessen) 20/661, 3.

[50] *Hessisches Gesetz über die öffentliche Sicherheit und Ordnung – HSOG*, as amended. On the German differentiation between such 'police laws' on the one hand, and criminal procedure on the other, see Dominik Brodowski, 'Alternative Enforcement Mechanisms in Germany' in Matthew Dyson and Benjamin Vogel (eds), *The Limits of Criminal Law* (Intersentia 2018) 365, 385–90.

[51] Unofficial translation available at <https://www.gesetze-im-internet.de/englisch_stpo/index.html> accessed 9 August 2022.

[52] For explanatory remarks on the legislative draft, see LT-Drucks. (Hessen) 19/6502, 40.

[53] Section 49 of the Hamburg Act on Data Processing by the Police (Gesetz über die Datenverarbeitung der Polizei – PolDVG), for explanatory remarks on the legislative draft see LT-Drucks. (Hamburg) 21/17906, 26.

[54] 'NRW-Polizei verteidigt umstrittene Palantir-Software' (*Zeit Online*, 3 May 2021) <https://www.zeit.de/news/2021-05/03/nrw-polizei-verteidigt-umstrittene-palantir-software?utm_referrer=https%3A%2F%2Fwww.google.com%2F> accessed 9 August 2022.

Rhine-Westphalia adopted a legal basis for operational use of the software.[55] Unlike the provisions that have been introduced in Hessen and Hamburg, North Rhine-Westphalia's provision does not restrict the use of the software to 'justified individual cases' but instead requires that its use is necessary for the prevention or preventive fight against serious crimes or of a danger of significant weight. Furthermore, the provision is missing the requirement of the head of the police or any other higher-ranking representative having to decide on the deployment or significant changes to the software, nor does it require to involve the data protection officer prior to such decisions. In contrast to the other two provisions, it does, however, require the recording of each query.

It is very likely that other Länder will follow, as Bavaria has also acquired the Palantir software at the beginning of 2022 under the umbrella of a framework contract that is said to cover the use by other state or Länder authorities as well.[56]

Even though in none of these cases the software is used to calculate a risk score to individuals, it seems to provide a technically very suitable basis where such preventive functions could later be built upon.[57]

The use of the software is heavily criticised.[58] Even though it processes only information which is already provided (somewhere) in police databases, it is – by definition – characterised by a very broad scope, without even requiring a concrete threshold like a concrete threat or suspicion of a crime.[59] It processes personal data without any preliminary indications whether or not there is a link between such data and the individual case at hand. Quite the contrary, one of its main characteristic features is to rapidly access thousands of personal data across various databases only to find out whether and where such a link might exist. Such an approach shows elements of a 'fishing expedition' within the

[55] Section 23 (6) of the North Rhine-Westphalia Police Act (Polizeigesetz des Landes Nordrhein-Westfalen – PolG NRW).

[56] Werner Pluta, 'Bayerns Polizei bekommt Analyse-Software von Palantir' (*Golem*, 8 March 2022), <https://www.golem.de/news/big-data-bayerns-polizei-bekommt-analyse-software-von-palantir-2203-163691.html> accessed 9 August 2022; see also Clemens Arzt, 'Das Handeln von Polizei- und Ordnungsbehörden zur Gefahrenabwehr' in Matthias Bäcker, Erhard Denninger and Kurt Graulich (eds), *Handbuch des Polizeirechts* (7th edn, Beck 2021) mn 1305–1306.

[57] Markus Löffelmann, in his statement regarding the legislation draft for Section 25a HSOG describes predictive policing as its 'unspoken aim', page 107 of the committee document <https://hessischer-landtag.de/sites/default/files/scald/files/INA-AV-19-63-T1.pdf> accessed 9 August 2022; Sommerer, *Self-imposed Algorithmic Thoughtlessness and the Automation of Crime Control* (n 38) 81, describes HessenData as a 'precursor' to 'predictive policing', see also Krasmann and Egbert (n 9) 62–63 who predict a trend towards a 'one software fits all' approach including the 'platformisation' of data analytics, merging of different databases, and interoperability on several levels which allows police officers to pursue predictive work as well as data analytics for retrospective criminal prosecution.

[58] Marie Bröckling, 'Juristinnen kritisieren "Palantir-Paragraf" im geplanten Polizeigesetz' (*netzpolitik.org* 24 September 2019) <https://netzpolitik.org/2019/hamburg-juristinnen-kritisieren-palantir-paragraf-im-geplanten-polizeigesetz/> accessed 9 August 2022; Jannis Brühl, 'Palantir in Deutschland – Wo die Polizei alles sieht' *Süddeutsche Zeitung* (Munich, 18 October 2018) <https://www.sueddeutsche.de/digital/palantir-in-deutschland-wo-die-polizei-alles-sieht-1.4173809> accessed 9 August 2022.

[59] Arzt, 'Das Handeln von Polizei- und Ordnungsbehörden zur Gefahrenabwehr' (n 56) mn 1308.

relevant databases and therefore bears an undeniable indiscriminate effect. Furthermore, there is a lot of scepticism against cooperation with Palantir because of its links to US intelligence agencies and the political affiliations of the company's founder.[60]

The legal basis for HessenData, § 25a HSOG, and Hamburg's corresponding provision have been successfully challenged before the Federal Constitutional Court (see **5** below) and declared unconstitutional.[61] In October 2022, an additional constitutional complaint has been filed against the relevant law in North Rhine-Westphalia[62] – which had already faced harsh criticism in the course of the parliamentary debates.[63] A similar controversial debate has also started to unfold in Bavaria, whose Ministry of the Interior is still assessing whether or not it even recognises the need for a specific legal basis. In contrast, data protection advocates stress the need for specific regulations considering the intense infringements of fundamental rights the use of the software implies.[64]

2.3.3 Intelligent video surveillance

It might be questionable whether or not intelligent video surveillance should be defined as 'predictive policing', as it does not provide any predictions but rather identifies dangerous situations and behaviours in certain locations.[65] It seems, however, difficult to draw such a clear line, especially since it is also clearly based on assumptions as to which situations or behaviours can lead to further escalations.

The first project where intelligent video surveillance went into operational deployment was initiated by the City of Mannheim. It installed a number of cameras in certain local focus-points, and connected them to an AI-based software. The software has been developed by the *Fraunhofer Institut für Optronik, Systemtechnik und Bildauswertung* (Fraunhofer

[60] Pluta (n 56).

[61] The complaint written by *Tobias Singelnstein* which has been supported by a group of organisations from the human rights and data protection sphere (Gesellschaft für Freiheitsrechte e.V. [Society for Civil Rights], Humanistische Union, Datenschützer Rhein Main and Forum InformatikerInnen für Frieden und gesellschaftliche Verantwortung) is available at <https://freiheitsrechte.org/home/wp-content/uploads/2019/07/2019-07-01-VB-Hessen-finalohneAdressen.pdf> accessed (9 August 2022).

[62] *Gesellschaft für Freiheitsrechte* (Society for Civil Rights, ‚GFF erhebt Verfassungsbeschwerde gegen uferlose Big-Data-Methoden im Polizeigesetz von NRW: Der Einsatz von Big Data braucht strenge Voraussetzungen' (press release 6 October 2022), <https://freiheitsrechte.org/ueber-die-gff/presse/pressemitteilungen-der-gesellschaft-fur-freiheitsrechte/pm-stop-data-mining> accessed 21 October 2022; the full text of the complaint is available at < https://freiheitsrechte.org/uploads/documents/Freiheit-im-digitalen-Zeitalter/Polizeigesetz-NRW/2022-10-05-PolG_NRW_Palantir_Website_geschwaerzt_Punkte.pdf> accessed 21 October 2022.

[63] *Gesellschaft für Freiheitsrechte* (Society for Civil Rights), 'Stellungnahme' (28 March 2022) <https://freiheitsrechte.org/home/wp-content/uploads/2022/04/PolGNRW_Stellungnahme_GFF.pdf> accessed 9 August 2022.

[64] Elisa Harlan, Boris Kartheuser and Robert Schöffel, 'Analysetool der US-Firma Palantir: Schafft die Polizei den gläsernen Bürger?' (*Tagesschau*, 3 July 2022) <https://www.tagesschau.de/investigativ/br-recherche/polizei-analyse-software-palantir-101.html> accessed 9 August 2022.

[65] Kuhlmann and Trute (n 2) 107; arguing for a classification as predictive policing: Wischmeyer, 'Predictive Policing, Nebenfolgen der Automatisierung von Prognosen im Sicherheitsrecht' (n 8) 201.

Institute of Optronics, System Technologies and Image Exploitation)[66] and trained to identify dangerous behaviour and alarm law enforcement staff so that crimes can be prevented by early interventions. The first cameras were installed in 2018 at Mannheim's main station. By the end of 2021, 68 cameras were in place throughout three so-called 'hotspots', ie the central shopping street, a big city square as well as the forecourt of the main station. The video material is retained for 72 hours. For the time being, police officers are still watching the video in real time and decide whether or not to alarm their colleagues or paramedics.[67] The long-term aim for the software is to work on its own so that less police staff is required.[68] The legislator of the German 'Land' Baden-Württemberg has introduced a specific legal basis for the analysis of image recordings generated by video surveillance in 2017. The wording of the relevant provision is strictly limited to analyses regarding behavioural patterns that indicate the commission of a crime and therefore does not cover biometric face recognition (which is, in fact, not part of the surveillance system deployed in Mannheim).[69]

Other cities are considering the implementation of surveillance systems similar to the one in Mannheim, as well. The Bavarian legislator, however, abstained from introducing a new legal basis for intelligent video surveillance and biometric facial recognition in the course of a recent reform of its legislation governing police competences in 2018, because it was found that, based on practical experience, the necessary technology was not yet reliable enough.[70]

On the federal level, intelligent video surveillance has been tested in the course of the so-called pilot project 'Sicherheitsbahnhof' by the German Federal Police (*Bundespolizei*) in cooperation with the German railway company at the train station 'Südkreuz' in Berlin. The first part of the project was focused on biometric facial recognition. It started in 2017, and in its course, the systems 'BioSurveillance' by the company Herta Security, delivered by Dell EMC AG, 'Morpho Video Investigator (MVI)' by IDEMIA AG, and 'AnyVision' by AnyVision were used and tested.[71] The second part of the project was designed to focus on behavioural analysis, similar to technology used in Mannheim. It was supposed to start in July 2019 and to use software provided by IBM Germany GmbH, the Hitachi

[66] Kai Wendt, 'Zunehmender Einsatz intelligenter Videoüberwachung' [2018] ZD-Aktuell, 06122.

[67] Olivia Kaiser, 'Was brachte die intelligente Videoüberwachung bisher?' *Rhein-Neckar-Zeitung* (Heidelberg, 3 December 2021) <https://www.rnz.de/nachrichten/mannheim_artikel,-mannheim-was-brachte-die-intelligente-videoueberwachung-bisher-_arid,782203.html> accessed 9 August 2022.

[68] See the reasons put forward for its legal basis (§ 21 [4] Police Act Baden-Württemberg), LT-Drucks. (Baden-Württemberg) 16/2741, 9.

[69] See the reasons put forth for its legal basis (§ 21 [4] Police Act Baden-Württemberg), LT-Drucks. (Baden-Württemberg) 16/2741, 9.

[70] LT-Drucks. (Bavaria) 17/21887.

[71] *Bundespolizeipräsidium* (Federal Police National Headquarters), Final report 'Teilprojekt 1 "Biometrische Gesichtserkennung"', 22 <https://www.bundespolizei.de/Web/DE/04Aktuelles/01Meldungen/2018/10/181011_abschlussbericht_gesichtserkennung_down.pdf?__blob=publicationFile> accessed 9 August 2022.

Consortium (Hitachi, Conef, MIG), Funkwerk video systems GmbH, and G2K Group GmbH.[72]

Even though there is no public information on the results of the second part of the project, the former Federal Minister for the Interior repeatedly stressed the importance of both intelligent video surveillance and biometric facial recognition technology, and announced a significant expansion of video surveillance and investments of up to € 180 million for 3.000 new cameras with technology allowing high-definition pictures so that until 2024, every large train station throughout the country may be equipped with 'modern camera technology'.[73] On the other hand, critical voices argue that the results regarding facial recognition were not reliable and false positive rates were still too high. Therefore, they demand to abstain from the use of biometric facial recognition technologies.[74] This position seems to resonate with the coalition forming the current federal government which has expressly declared that video surveillance cannot substitute the presence of police officers, but that it can be used to support police work at crime hotspots. Currently, the Federal Ministry of Education and Research funds projects with a focus on intelligent video surveillance in the form of behavioural analysis, for example the development of a software program that is able to identify dangerous behaviour or medical emergencies on train stations or suspicious behaviour on airports through video-based pattern detection.[75] Aspirations regarding the use of biometric face recognition, however, appear to be at a halt as the coalition agreement states that the coalition opposes the ubiquitous use of video surveillance and any use of biometric technology for surveillance purposes.[76] Furthermore, the coalition agreement states with regard to the ongoing negotiations about the so-called Artificial Intelligence Act[77] on the EU level that the use of

[72] Federal Police, 'Test intelligenter Videoanalysetechnik' (press release 7 June 2019) <https://www.bundespolizei.de/Web/DE/04Aktuelles/01Meldungen/2019/06/190607_videoanalyse.html> accessed 3 April 2022.

[73] Federal Ministry of the Interior, 'Erhöhung der Sicherheit auf Bahnhöfen' (press release 12 September 2019), <https://www.bmi.bund.de/SharedDocs/pressemitteilungen/DE/2019/09/sicherheit-auf-bahnhoefen.html> and 'Bundesregierung und Deutsche Bahn beschließen weitere Maßnahmen für mehr Sicherheit an Bahnhöfen' (press release 13 December 2020) <https://www.bmi.bund.de/SharedDocs/pressemitteilungen/DE/2020/12/sicherheit-bahnhoefe.html> both accessed 9 August 2022.

[74] For a summary of the debate see Arzt, 'Das Handeln von Polizei- und Ordnungsbehörden zur Gefahrenabwehr' (n 56) paras 1155–1159; Johanna Sprenger, 'Verbrechensbekämpfung' in Martin Ebers and others (eds), *Künstliche Intelligenz und Robotik* (Beck 2020), paras 55–58.

[75] See the project descriptions by the Federal Ministry of Education and Research: <https://www.sifo.de/sifo/shareddocs/Downloads/files/projektumriss_apfel.pdf?__blob=publicationFile&v=1> and <https://www.sifo.de/sifo/shareddocs/Downloads/files/mustererkennung_d_adis.pdf?__blob=publicationFile&v=1> both accessed 9 August 2022.

[76] Coalition agreement, lines 3647–3649 <https://www.wiwo.de/downloads/27830022/8/koalitionsvertrag-2021-2025.pdf> accessed 9 August 2022.

[77] Commission, 'Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts' COM (2021) 206 final.

biometric face recognition technologies in public spaces as well as the use of 'scoring' systems by state authorities shall be prohibited by EU law.[78]

2.3.4 OSINT and SOCMINT

The German Government funds several research projects regarding software applications that strive to be able to automatically access external publicly available information in social media (open source intelligence/OSINT or social media intelligence/SOCMINT) in order to identify tendencies of extremism and radicalisation, and to prepare preventive strategies.[79]

One of these projects, X-SONAR (*Extremistische Bestrebungen in Social Media Netzwerken: Identifikation, Analyse und Management von Radikalisierungsprozessen* – extremist endeavours in social media networks: identification, analysis, and management of radicalisation processes) was conducted from 2017 to 2020 and focused on the development of an analytic tool that can assess discourses in publicly available online networks, platforms and blogs. The software crawls the relevant information in external sources (such as Facebook or Twitter), and then uses language analysis in order to identify patterns of radicalisation and indicators for early detection of radical tendencies.[80] Based on such identification, law enforcement authorities are supposed to be able to locate relevant discourses for further individual review. The software is said to work theory-based, ie (at least for now) without recourse to artificial intelligence or machine learning.[81]

A more recent example is the project ERAME (*Erkennung von Radikalisierungszeichen in Sozialen Medien* – detection of indications of radicalisation in social media). It aims to develop a software tool that helps with the assessment and analysis of content from video platforms (such as YouTube). Computer-linguistics shall be relied upon in order to create a catalogue which serves to identify and classify indicators for extremist content.[82]

---

[78] Coalition agreement, lines 504–505 <https://www.wiwo.de/downloads/27830022/8/koalitionsvertrag-2021-2025.pdf> accessed 9 August 2022.

[79] Wolfgang Kahl, 'PANDORA, RadigZ & X-SONAR' [2017] (2) Forum Kriminalprävention 35.

[80] See the description of the cooperative partner *Landeskriminalamt Niedersachsen* (Lower Saxony State Criminal Police Office), 'Forschungsprojekt "X-Sonar" – Extremistische Bestrebungen in Social Media Netzwerken: Identifikation, Analyse und Management von Radikalisierungsprozessen' <https://www.lka.polizei-nds.de/forschung/forschungsprojekt-x-sonar---extremistische-bestrebungen-in-social-media-netzwerken-identifikation-analyse-und-management-von-radikalisierungsprozessen-113539.html> accessed 9 August 2022.

[81] BT-Drucks. 19/7604, 10–11; Deutsche Hochschule der Polizei (German Police University), 'Forschungsbericht 2018', 97 <https://www.dhpol.de/Forschungsbericht_FIN_Versand.pdf> accessed 9 August 2022; Matthias Becker, 'Fundgrube für Fahndungsdaten – Wie die Polizei soziale Netzwerke nutzt' (*Deutschlandfunk, Online edition*, 26 May 2018) <https://www.deutschlandfunk.de/fundgrube-fuer-fahndungsdaten-wie-die-polizei-soziale-100.html> accessed 9 August 2022.

[82] Federal Ministry of Education and Research, 'Erkennung von Radikalisierungszeichen in Sozialen Medien (ERAME)' <https://www.sifo.de/sifo/shareddocs/Downloads/files/projektumriss_erame_bf.pdf?__blob=publicationFile&v=1> accessed 9 August 2022.

As for now, there is no specific (explicit) legal basis for the deployment of projects such as X-SONAR or ERAME. In relation to X-SONAR, individual rights shall be protected by anonymisation and pseudonymisation, meaning that no individuals are meant to be identified.[83] The project ERAME is described to lay special emphasis on the legal assessment of the development process in order to ensure that the functions of the software are in compliance with the law.[84]

2.4 Transaction-based 'Predictive Policing'

2.4.1 Advance risk assessments by Fiscal Authorities

Fiscal authorities are starting to use artificial intelligence in different areas of their responsibilities. One of these use cases is an automated analysis in order to identify cases of non-compliance with legal requirements, especially by evaluating certain risk indicators, such as irregularities etc.

The German Financial Intelligence Unit (FIU) handles reports on suspicious transaction concerning money laundering, terrorist financing and other criminal offences. These reports are filed through a software program which is an adapted version of the Software goAML that has been developed by the UN especially for use by all national Financial Intelligence Units.[85] The FIU handles these reports on the basis of a risk-based approach.[86] This means that in order to use its resources most efficiently on the vast number of suspicious transactions reported (144.005 in 2020 alone[87]) through goAML, the FIU decides on the most effective way to proceed with each individual report, according to its relevance for the prevention of money laundering and terrorist financing. The FIU reports that it has started to use an IT component based on artificial intelligence called 'FIU Analytics' since autumn 2020.[88] The software is said to help selecting cases that require further review by calculating risk scores between 1 and 100. The risk score can be subject to

---

[83] According to one of the scientists working on the project with *Fraunhofer-Institut für Sichere Informationstechnologie* (Fraunhofer Institute for Secure Information Technology), Martin Steinebach, in Becker (n 80).

[84] Federal Ministry of Education and Research (n 51).

[85] Arzt, 'Das Handeln von Polizei- und Ordnungsbehörden zur Gefahrenabwehr' (n 56) mn 1281.

[86] For a detailed description: Jens Bülte, 'Risikobasierte Arbeitsweise sowie Analyse- und Weiterleitungspflichten der FIU in den Grenzen des geltenden Rechts' (expert study commissioned by the German Federal Ministry of Finance, 2021), <https://www.bundesfinanzministerium.de/Content/DE/Downloads/Publikationen/gutachten-zu-risikobasierter-arbeitsweise-der-fiu-pdf.pdf?__blob=publicationFile&v=2> accessed 9 August 2022.

[87] FIU, 'Annual Report 2020' <https://www.zoll.de/SharedDocs/Pressemitteilungen/DE/Bargeld/2021/z85_fiu_jahresbericht.html> accessed 9 August 2022; it has to be noted, though, that it is highly controversial whether the risk-based approach is a viable and legitimate basis for the work of the FIU, see Steffen Barreto da Rosa, 'Zum "risikobasierten Ansatz" der FIU im Rahmen der operative Analyse von Meldungen nach dem Geldwäschegesetz', Der Kriminalist [2022], 23.

[88] FIU, 'Annual Report 2020', 11 and 'Annual Report 2021', 30 < https://www.zoll.de/DE/FIU/Fachliche-Informationen/Jahresberichte/jahresberichte_node.html> accessed 30 September 2022; see also BT-Drucks. 19/30278, 2; as to the rather vague and partly inconsistent communication regarding the functions

further changes, as new information reported through goAML is constantly matched with already existing data (as far as such data has been legally stored for such purposes). Therefore, a case that gained only a minor risk score in the beginning can be identified as part of a high-risk pattern at a later stage.[89]

The legal basis for the analysis of the incoming reports by the FIU is § 30 (2) of the Act against Money Laundering (GwG)[90] which does not specify any details of the analysis, such as the execution of advance checks or the use of technology.

The automated risk management systems run by tax authorities in Germany could also be defined as 'predictive policing'.[91] In Germany, tax reports are processed automatically in case there is no indication that a manual assessment is necessary. In order to identify cases that require such comprehensive review by tax officials or further investigations, tax authorities can use so-called automated risk management systems. Cases requiring comprehensive review can be both cases with irregularities or contradictions as well as 'high-risk' cases. According to § 88 (5) of the German Fiscal Code[92], automated risk management systems must, at a minimum, ensure (1) to select a sufficient number of cases randomly, ie in addition to those that are found to require comprehensive review, (2) that all the selected cases are actually reviewed, (3) that officials can manually select cases for comprehensive review as well, and (4) that regular reviews are conducted to determine whether risk management systems are fulfilling their objective. *Baur* argues that it follows from § 88 (5) 2 AO – which demands that the risk management systems take the principle of cost-effective administration into account – that petty cases are to be excluded from the selections.[93] § 88 (5) 4 AO expressly states that further details of the risk management systems do not have to be made public. Therefore, not much is known about the technologies tax authorities rely on. The Federal Ministry of Finance stated in 2021 that the systems used for tax assessment currently work theory-based, but that artificial intelligence technology could be implemented in upgrades.[94]

---

and use of 'FIU Analytics', see Steffen Barreto da Rosa 'Vorbemerkungen zu Abschnitt 5 – Zentralstelle für
Finanztransaktionsuntersuchungen' in Felix Herzog and Christoph Achtelik, 'Geldwäschegesetz' (Beck, 5th edn (forthcoming) mn 37.

[89] Publicly available information on the details on how FIU Analytics works are very rare, these clarifications stem from the protocol of an exchange with representatives of the Customs Directorate General and staff counsel representatives as well as representatives of the IT company Capgemini (BDZ Personalräte Kompakt 11/2019) <https://bdzovbremen.blogspot.com/2019/11/gzd-financial-intelligence-unit-automatisierte-vorbewertung-kuenstliche-intelligenz-ki.html> accessed 9 August 2022.

[90] Gesetz über das Aufspüren von Gewinnen aus schweren Straftaten.

[91] Rademacher (n 5) 235.

[92] Unofficial translation available at <https://www.gesetze-im-internet.de/englisch_ao/index.html> accessed 9 August 2022.

[93] Baur (n 5) 283; with doubts: Rademacher (n 5) 238.

[94] BT-Drucks. 19/30278, 4; see also Thomas Wischmeyer, 'Regulierungs- und Verwaltungshandeln durch KI' in Martin Ebers and others (eds), *Künstliche Intelligenz und Robotik* (Beck 2020) mn 26–27 who assumes that artificial intelligence is used for investigations on VAT carousels.

2.4.2 Risk assessments by Custom Authorities

In a similar vein, the German custom authorities are starting to use artificial intelligence to determine which goods to examine at custom controls. According to Article 46(2) of the (European) Union Customs Code, '[c]ustoms controls, other than random checks, shall primarily be based on risk analysis using electronic data-processing techniques, with the purpose of identifying and evaluating the risks and developing the necessary counter-measures, on the basis of criteria developed at national, Union and, where available, international level'. Public information on this risk analysis is scarce. Yet, it has been reported that German custom authorities employ – and intend to expand the use of – 'neural networks' and 'artificial intelligence' in a project called ZERBERUS.[95]

2.5 Objectives, effects and reception of 'Predictive Policing' in Germany

Even though the 'predictive policing' models described above vary significantly in both their functions and their concrete objectives, they all serve the general aim to link and analyse data more efficiently in order to rationalise the allocation of the relevant authorities' resources. In particular, they aim to use scarce resources in a more focused and efficient manner, and thereby allow authorities to fulfil their responsibilities more effectively. This is in line with the Artificial Intelligence Strategy of the German Federal Government: 'In the context of policing, the use of AI is an important strategic aspect of domestic security. For instance, it can help to significantly enhance existing capabilities and make police work more targeted and effective. […] In each specific use case, though, it must be examined whether and how AI can be deployed in a policing context in compliance with fundamental rights.'[96]

The perception among practitioners is documented mainly for location-based predictive policing systems, because among all of the different models, it is the one that has been the main subject to research projects and evaluations so far. For example, research projects on the use of KrimPro in Berlin and of PreCobs in Baden-Württemberg have shown that the perception among police officers is very ambivalent. Findings from Baden-Württemberg suggest that the view is more positive among higher ranks in the hierarchy and more pessimistic among patrol officers.[97] However, scepticism is not only expressed by patrol officers. Some analysts have stated that the software merely confirmed findings that they had previously been able to reach through classical police work – which some felt was now less appreciated.[98] Some officers have said they felt pressured to follow the system's advice or at least that doing so made it much easier to justify their operational

[95] BT-Drucks. 19/30278, 3.

[96] <https://www.ki-strategie-deutschland.de/files/downloads/Fortschreibung_KI-Strategie_engl.pdf> accessed 9 August 2022.

[97] Gerstner (n 12) 134.

[98] Löbel and Schuppan (n 21) 20 and 22.

decisions.[99] A distinctive challenge for the acceptance of 'predictive policing' software – or for preventive police work in general – seems to be best described by the almost proverbial phenomenon that 'there is no glory in prevention': While the software was found to have little to no effect on actual arrests, its preventive value is less evident and can only be deducted from statistical evaluations. Therefore, prevention can be felt to be less satisfying.[100] That being said, a lot of police officers also perceived 'predictive policing' software as a useful supplement for their work and stated that it had a significant effect not only on the planning of operations, but also on the actual pursuit, as they acted more cautiously in locations that had been flagged as high risks.[101] Many leading officers further reported that the software had a very useful effect on their work in that it made it much easier to successfully request additional police forces to a certain area.[102]

Reception of 'predictive policing' in the general public is very diverse. Journalists, critical voices in legal literature, non-governmental organisations and the Federal Commissioner for Data Protection and Freedom of Information warn against negative effects of 'predictive policing', such as excessive use of personal data, blind trust in technology, lack of quality of data, direct and indirect discriminatory effects, as well as so-called 'chilling effects' of automated policing.[103] This does not mean that the majority of these critics dismiss the idea of 'predictive policing' completely.[104] Furthermore, as seen above, the points of criticism differ depending on the specific system in question. What can be noted on a general level, however, is a call for stricter regulation of predictive policing systems, implying the need for specific and restrictive legal bases including effective legal safeguards, transparency and supervision requirements as well as thorough evaluation both prior to their introduction and continuously during the time of their use.[105]

---

[99] Albert Meijer, Lukas Lorenz and Martijn Wessels, 'Algorithmization of Bureaucratic Organizations: Using a Practice Lens to Study How Context Shapes Predictive Policing Systems' (2021) 81 Public Administration Review 837, 842.

[100] Löbel and Schuppan (n 21) 20; Gerstner (n 12) 134, Krasmann and Egbert (n 9) 51.

[101] Egbert, 'Predictive Policing als Treiber rechtlicher Innovation?' (n 10) 35–36.

[102] Meijer, Lorenz and Wessels (n 99) 841–842.

[103] See Sprenger (n 74) paras 40–44; Lind, *Raumbezogenes Predictive Policing in Deutschland* (n 8).

[104] Note, however, that in March 2022, 41 mostly European civil society organisations published an open letter in which they urge the Council of the European Union, the European Parliament, and all EU member state governments to prohibit AI predictive and profiling AI systems in law enforcement and criminal justice in the Artificial Intelligence Act; see Fair Trials International, European Digital Rights and others, 'Civil Society calls on the EU to prohibit predictive and profiling AI systems in Law Enforcements and Criminal Justice' (March 2022) <https://www.fairtrials.org/app/uploads/2022/03/Ban_Predictive_Policing_Criminal_Justice_Statement.pdf accessed> 9 August 2022.

[105] See Federal Commissioner for Data Protection and Freedom of Information, 'Einsatz von Künstlicher Intelligenz im Bereich der Strafverfolgung und Gefahrenabwehr' (thesis paper, 23 March 2022) <https://www.bfdi.bund.de/DE/DerBfDI/Inhalte/Konsultationsverfahren/KI-Strafverfolgung/KI-Strafverfolgung-Thesen-BfDI.html> accessed 9 August 2022; Lind (n 8).

2.6 Assessment of the reliability, impartiality and effectiveness of 'preventive policing' technology in Germany

As for the time being, most of the information on evaluations regarding 'predictive policing' systems that is publicly available pertains to location-based 'predictive policing' systems. These evaluations have in common that it was found to be simply impossible to prove a casual effect of the relevant prevention method on the development of crime or even to assess the accuracy of its individual predictions. The evaluations focused also on other aspects, such as practical and technical aspects on the handling of the relevant systems, its effect on the police work itself and perceptions among practitioners (see **2.5** above). The conclusion drawn from the evaluation of the PreCobs system in Baden-Württemberg seems to be exemplary in that regard: 'despite some positive findings, the impact on crime remains unclear and the size of crime reducing effects appears to be moderate. Within the police force, the acceptance of predictive policing is a divisive issue.'[106]

3 Normative framework

3.1 Law and soft law

3.1.1 Specific legal bases for use of person-focused 'predictive policing' systems

In contrast to location-based predictive policing systems and OSINT/SOCMINT, some of the above-mentioned examples of person-focused predictive policing already have a specific basis in law. The content of these provisions differs depending on the relevant methods. They do have in common, however, that they do not mention artificial intelligence explicitly but rather use more *technology-open wordings* such as 'automated analysis', 'automated comparisons' or 'automated systems'. In some cases, they do not even refer to automation at all but merely to the relevant task such as 'analysis' or 'further processing of personal data'. The relevant provisions define the use-cases of these automated measures. Partly, there are also rules in place on *substantial requirements* as to the characteristics of the relevant technology, or *procedural rules* regarding its use (eg human intervention), the decision regarding deployment or changes to the technology in use, and/or regular monitoring of the technology in question (see **2.2** and **2.3.2** above and **3.3.2** below for further details).

3.1.2 Lack of general legislation on predictive policing/use of AI

In contrast to these specific legal bases, there is – as of now – no general legislation on the use of artificial intelligence for 'predictive policing'. Whether or not such legislation might be adopted in the future also depends on the outcome of the negotiations on the so-called Artificial Intelligence Act, the European Commission's draft proposal to regulate artificial intelligence (AI) systems, including in the area of law enforcement and criminal justice.[107]

[106] Gerstner (n 12) 115.
[107] COM (2021) 206 final (n 77).

3.1.3 Compliance with EU law, constitutional law, and the data protection framework

All 'predictive policing' systems must, however, comply with

- constitutional law, in particular the fundamental right to *informational self-determination* following from Article 2 (1) in conjunction with Article 1 (1) of the German Basic Law[108] and the general principle of *equality*, especially the *ban on discrimination* according to Article 3 of the German Basic Law;
- EU primary law, especially the EU Charter of Fundamental Rights, in particular the rights to privacy (Article 7 of the Charter) and the right to protection of personal data (Article 8 of the Charter) as well as the right to non-discrimination (Article 21 of the Charter), whenever the Charter is applicable in accordance with Article 51 of the Charter; and
- European and German Data Protection Law, in particular the rules on the *automation of individual decisions* and on *impact assessments*.

3.1.4 Soft law

The *Artificial Intelligence Strategy of the German Federal Government* stresses compliance with fundamental rights and points to the recommendations made by the Data Ethics Commission, which call for a risk-adapted regulatory system.[109] On this basis, algorithmic systems with potential for harm should be regulated with instruments that may, depending on the severity of that harm, include 'formal and substantive requirements (eg transparency obligations, publication of a risk assessment) and monitoring procedures (eg disclosure obligations towards supervisory bodies, ex-post controls, audit procedures)', ex-ante approval procedures or – in cases with serious potential for harm – additional measures such as enhanced ('always-on') oversight and extensive transparency obligations. The Federal Government's Data Ethics Commission recommends to assess the 'use of algorithmic systems by state bodies' as 'particularly sensitive – entailing at the very least a comprehensive risk assessment.' It further stresses that 'decisions taken by the State on the basis of algorithmic systems must still be transparent, and it must still be possible to provide justifications for them. It may be necessary to clarify or expand the existing legislation on freedom of information and transparency in order to achieve these goals. Furthermore, the use of algorithmic systems does not negate the principle that decisions made by public authorities must generally be justified individually; on the

[108] Basic Law for the Federal Republic of Germany. Unofficial translation available at <https://www.gesetze-im-internet.de/englisch_gg/index.html> accessed 9 August 2022.

[109] Artificial Intelligence Strategy (2018) <https://www.bundesregierung.de/resource/blob/975226/1550276/3f7d3c41c6e05695741273e78b8039f2/2018-11-15-ki-strategie-data.pdf?download=1> accessed 9 August 2022.

contrary, this principle may impose limits on the use of overly complex algorithmic systems.'[110]

Only recently, the *Federal Commissioner for Data Protection and Freedom of Information* conducted a public consultation on the use of AI for preventive police work and criminal prosecution, and emphasised that more concrete regulatory standards are needed with regard to the use of AI for preventive police work and criminal investigations. The consultation entailed seven theses, starting with (1) the need for a broad public debate and comprehensive empiric review in order to clarify the benefits of AI applications in this area and its potential risks for individual rights, including potential discriminatory effects as well as its meaning for democratic and rule of law procedures. In that regard, the Federal Government should also provide an overall account of all police powers (especially surveillance measures). Furthermore, (2) the use of AI should always require a specific legal basis and must not be based on mere general clauses regarding police work. (3) The use of AI must be in compliance with the general rules on data protection and may not weaken individual remedies. (4) AI needs to be explainable; the quality of data, also of data used for training purposes, must be ensured. (5) The core area of private conduct of life and the guarantee of human dignity must not be affected. (6) Data protection authorities must be able to effectively supervise the use of AI; and (7) there must always be a privacy-impact assessment prior to the use of AI for the purpose of preventive police work and criminal prosecution.[111]

3.2 Case Law

While not all decisions of judicial bodies in Germany are published, the case law available to us does not address the use of AI-based systems for 'predictive policing' as such. Some guidance can be drawn, though, from the jurisprudence of the Federal Constitutional Court and the ECJ.

In the decisions relevant in the context of 'predictive policing', the Federal Constitutional Court assessed whether certain forms of processing of personal data by law enforcement authorities in order to prevent crime constitute an infringement of the right to informational self-determination derived from Article 2 (1) in conjunction with Article 1 (1) of the German Basic Law that could be justified because it is necessary and proportionate

[110] Opinion of the Data Ethics Commission (2019) <https://www.bmi.bund.de/SharedDocs/downloads/EN/themen/it-digital-policy/datenethikkommission-abschlussgutachten-lang.pdf;jsessionid=A6ACB701AD91D0CFE71972B454523A7E.2_cid364?__blob=publicationFile&v=4> accessed 9 August 2022.

[111] Federal Commissioner for Data Protection and Freedom of Information, 'Einsatz von Künstlicher Intelligenz im Bereich der Strafverfolgung und der Gefahrenabwehr' (report on the public consultation process, 23 March 2022) <https://www.bfdi.bund.de/SharedDocs/Downloads/DE/Konsultationsverfahren/2_KI-Strafverfolgung/Konsultationsbericht.pdf;jsessionid=A459A4FEDC511B17C2035C0FC5C5ADB9.intranet241?__blob=publicationFile&v=3> accessed 9 August 2022.

in order to serve a legitimate purpose. Some of the findings in that regard seem of particular relevance for 'predictive policing':

3.2.1 Infringement of the right to informational self-determination

As a starting point, all 'predictive policing' methods that process personal data constitute an infringement of the right to informational self-determination that requires justification. In its decision regarding automated number plate recognition, the Federal Constitutional Court (BVerfG) recently held (and explicitly overturned previous decisions to the contrary) that it even constitutes a relevant infringement of the right to informational self-determination when personal data is checked automatically with police data, the result is a 'no match', and the data is deleted immediately.[112] Furthermore, the Federal Constitutional Court also recognises an infringement of the right to informational self-determination when personal data that has already been collected by state authorities is used beyond the specific purpose of the data collection (further use).[113]

As to the weight of the infringement, the jurisprudence of the Federal Constitutional Court but also of the ECJ indicates that a broad personal scope of the relevant measure, ie a high number of persons potentially affected as well as the use of modern technologies allowing 'data mining' or 'complex forms of data cross-checking' increases the weight of the infringement.[114]

3.2.2 Justification

In order to be justified, an infringement must be necessary to serve a legitimate purpose, such as the prevention of crime or other threats. With regard to the processing of personal data, that means that there needs to be a link between that purpose and the data to be processed. This requirement has been highlighted in the Federal Constitutional Court's decision regarding the Federal Criminal Police Office Act. It stated that both the collection of personal data by state authorities as well as its 'further use' (beyond the purposes that were initially justified) require sufficiently specific grounds or another specific relation to its purpose, such as the targeting of specific risky activities or special sources of danger.[115] This is in line with the position taken by the ECJ regarding data retention for the purposes of prevention, investigation, detection and prosecution of serious crime, where it requires that objective criteria are met that establish a connection between the

---

[112] BVerfG, order of 18 December 2019 – 1 BvR 142/15 *Automatic number plate recognition II* ECLI:DE:BVerfG:2018:rs20181218.1bvr014215 = BVerfGE 150, 244.

[113] BVerfG, judgment of 20 April 2018 – 1 BvR 966/09, 1 BvR 1140/09 *BKAG* ECLI:DE:BVerfG:2016:rs20160420.1bvr096609 = BVerfGE 141, 220 para 289.

[114] BVerfG, judgment of 19 May 2020 – 1 BvR 2835/17 *Federal Intelligence Service – foreign surveillance* ECLI:DE:BVerfG:2020:rs20200519.1bvr283517 = BVerfGE 154, 152 para 192 and BVerfG, order of 10 November 2020 – 1 BvR 3214/15 *Counter-Terrorism Database Act II/Data-Mining* ECLI:DE:BVerfG:2020:rs20201110.1bvr321415 = BVerfGE 156, 11 para 109; ECJ, case C-817/19 *Ligue des droits humains v Conseil des ministres* ECLI:EU:C:2022:491 paras 98–111.

[115] BVerfG, judgment of 20 April 2018 – 1 BvR 966/09, 1 BvR 1140/09 *BKAG* ECLI:DE:BVerfG:2016:rs20160420.1bvr096609 = BVerfGE 141, 220 para 289.

data to be retained and the objective pursued. That means that there must be objective evidence for a link between the persons concerned with serious criminal offences, for example a connection to certain groups or areas with a high risk that such offences might be committed.[116]

There might be additional requirements depending on the weight of the infringement. For instance, in its decision regarding the extended use of data within the joint database for police authorities and intelligence services according to the Counter-Terrorism Database Act, the Federal Constitutional Court required at least the existence of a sufficiently identifiable danger or a suspicion based on specific facts that are supported by sufficiently concrete and tangible circumstances.[117] Depending on the weight of the infringement, the Federal Constitutional Court also sets restrictions as to which kind of crime to be prevented or the interests to be protected. With regard to automated number plate recognition, it ruled that '[g]iven the weight of its interference, automatic number plate recognition must serve to protect legal interests of at least considerable weight, or comparably weighty public interests'.[118] In a similar vein, the ECJ recently stressed that, in relation to data gathering without initial suspicion, there must be 'clear and precise rules governing the scope and application of the measures provided for', which must include 'safeguards, so that the persons whose data have been transferred have sufficient guarantees to protect effectively their personal data against the risk of abuse.' The legislation 'must, in particular, indicate in what circumstances and under which conditions a measure providing for the processing of such data may be adopted, thereby ensuring that the interference is limited to what is strictly necessary.' [119]

Notably, the Federal Constitutional Court expressly referred to the use of algorithms in its ruling on foreign surveillance by the Federal Intelligence Services, and stated that the legislator may have to lay down the modalities of their use, in particular to ensure that their use can generally be reviewed by the independent oversight regime.[120] Similarly, the ECJ stressed that the need for 'safeguards is all the greater where personal data are subject to automated processing.'[121]

---

[116] ECJ, joined cases C-203/15 and C-698/15 Tele2 Sverige AB v Postoch telestyrelsen and Secretary of State for the Home Department v Tom Watson and Others ECLI:EU:C:2016:970 para 111.

[117] BVerfG, order of 10 November 2020 – 1 BvR 3214/15 *Counter-Terrorism Database Act II/Data-Mining* ECLI:DE:BVerfG:2020:rs20201110.1bvr321415 = BVerfGE 156, 11; Golla argues that the findings of this decision are applicable to the legal basis for 'HessenData' (as well as its Hamburg equivalent) with the consequence that these provisions would not meet the constitutional requirements either, Sebastian Golla, 'Algorithmen, die nach Daten schürfen – "Data-Mining" zur Gefahrenabwehr und zur Strafverfolgung', [2021] NJW 667, 670–672.

[118] BVerfG, order of 18 December 2019 – 1 BvR 142/15 *Automatic number plate recognition II* ECLI:DE:BVerfG:2018:rs20181218.1bvr014215 = BVerfGE 150, 244.

[119] ECJ, case C-817/19 *Ligue des droits humains v Conseil des ministres* ECLI:EU:C:2016:970 para 117.

[120] BVerfG, judgment of 19 May 2020 – 1 BvR 2835/17 *Federal Intelligence Service – foreign surveillance* ECLI:DE:BVerfG:2020:rs20200519.1bvr283517 = BVerfGE 154, 152.

[121] ECJ, case C-817/19 *Ligue des droits humains v Conseil des ministres* ECLI:EU:C:2016:970 para 117.

3.3 Substantive guarantees

In addition to Constitutional and European Law as interpreted in the jurisprudence described above, data protection law and a few provisions governing specific methods of 'predictive policing' provide for some substantive guarantees.

3.3.1 Data protection law

§ 54 of the Federal Data Protection Act[122] implementing Article 11 of Directive (EU) 2016/680 sets out limits for '*decision(s) based solely on automated processing* which produces an adverse legal effect concerning the data subject or significantly affects him or her'. However, none of the 'predictive policing' systems explained above aim to generate automated decisions. Rather, it is regularly being emphasised that the relevant technologies serve as a mere means of support and the decision itself is still up to (a) human officer(s)[123] – although that decision may yet be 'anchored' in the suggestion made by technology.[124]

Therefore, § 67 of the Federal Data Protection Act implementing Article 27 of Directive (EU) 2016/680 seems of higher practical relevance as it requires to conduct, prior to the processing of personal data, a *data protection impact assessment* whenever data is to be processed by means of a new technology likely to result in a substantial risk to the legally protected interests of data subjects.

3.3.2 Method-specific provisions

§ 4 (3) of the PNR Act (see **1.3.1** above), which regards the patterns to be automatically matched against PNR-data, is one of the few examples setting out at least basic requirements as to both the design of the technology being used as well as procedural requirements for its establishment and further use. As to the design of the patterns, it prescribes the *combination of incriminating and exonerating criteria* in order to limit the amount of potential false-positives. Furthermore, and in order to *prevent discrimination*, it prohibits the use of information on a person's race or ethnic origin, political opinions, religion or philosophical beliefs, trade union membership, health, sexual life or sexual orientation. From a procedural point of view, the Federal Criminal Police Office must establish the patterns in *cooperation with its data privacy officer* and other security and police authorities. *Review by an independent body* is also guaranteed as the Federal Commissioner for Data Protection and Freedom of Information shall review the production and use of the patterns at least every two years, and report to the Federal Government every two years.

Another more detailed example is shown by § 88 (5) of the German Fiscal Code for automated risk assessment systems used by fiscal authorities (see **2.4.1** above). According

---

[122] An unofficial translation is available at <https://www.gesetze-im-internet.de/englisch_bdsg/index.html> accessed 9 August 2022.

[123] Arzt, 'Das Handeln von Polizei- und Ordnungsbehörden zur Gefahrenabwehr' (n 56) mn 1294; Lucia M Sommerer, *Personenbezogenes Predictive Policing* (Nomos 2020) 130.

[124] See **2.5** above.

to its second sentence, risk management systems have to take the *principle of cost-effective administration* into account.[125] The third sentence of this provision addresses the *reliability* of the use of these systems, as it requires (1) them to, at a minimum, select a sufficient number of cases randomly, ie in addition to those identified through the automated risk assessment, (2) that all the selected cases are actually reviewed, (3) that officials can also manually select cases for comprehensive review and (4) that *regular reviews* are conducted to determine whether risk management systems are fulfilling their objective. The provision also *limits transparency* explicitly as § 88 (5) 4 states that further details of the risk management systems do not have to be made public.

§ 25a HSOG and § 49 PolDVG (see **2.3.2** above) do not provide any requirements as to the design of the software, but at least define its tasks (automatically identify affiliations or connections between individuals etc), limit its use to the prevention of crimes or threats of significant weight in individual justified cases, and regulate the decision-making process on deployment or significant changes to the software. § 23 (6) PolG NRW, on the other hand, entails a rather generic description of the relevant tasks (comparisons, preparation or analysis of data), provides for less restrictive definitions of the purposes of its use and lacks further regulation of the decision-making process mentioned above (in contrast to the other two provisions it does, however, require recording of each query).

Other provisions, such as § 21 (4) of the Police Act Baden-Württemberg on intelligent video surveillance (see **2.3.3** above) merely regulate the use of 'automatic analysis' for specific purposes, but do not provide further details as to the design of the relevant technology or its monitoring.

Notably, § 30 (2) of the Act against Money Laundering merely refers to the assessment of incoming money laundering reports, but does not specify any details of the analysis, especially neither the execution of advance checks nor the use of technology (see **2.4.1** above).

4 General principles of Law

The potential effects of predictive policing on constitutional rights, proportionality and the rule of law are subject to a controversial debate. As stated above, the majority of the voices raising concerns do not seem to oppose the use of 'predictive policing' methods in general, but call for more legal safeguards and restrictions.

4.1 General limitations of predictive policing

Many commentators point to the inherent limitations of automated 'predictive policing' solutions and their potential negative side-effects. They point out that 'predictive policing' solutions base their assumptions on patterns and correlations rather than on an analysis of the root causes of crime and that therefore, its purpose will always be restricted

[125] Baur (n 5) 283; with doubts: Rademacher (n 5) 238.

to crime control through surveillance and short-term interventions. This might distract from the need for more complex but also more sustainable strategies against crime, such as efforts to remedy the social problems that can facilitate criminality.[126]

4.2 Discriminatory potential of predictive policing

Another characteristic limitation of automated 'predictive policing' is its dependency on data. Since every automated solution is only as good as the data it is trained on and provided with, all its accuracy relies on the data. Every imbalance, every error or incompleteness within the relevant data sets is likely to be reproduced in the assumptions and recommendations produced by the software in question.[127] This can have an adverse effect on groups within the population that are already vulnerable. If – for example – a certain community is already subject to high police attention, more crimes occurring within this group will be documented and more crime data relating to this group will be fed into the system.[128] Such negative feedback loops can increase discriminatory effects, even in case the relevant software does not process protected characteristics such as religion or ethnic background, but so-called proxies, ie circumstances that correlate with such characteristics (for example certain neighbourhoods, religious sites, travel routines, etc).[129] These concerns are also an issue of proportionality, as the consequences of biased 'predictive policing' are usually connected with police interference; therefore, the effects of false-positives can interfere with the right to liberty and security.

4.3 Potential remedies

Potential remedies against biased and false automation results are one of the most imminent topics within the current debate.

4.3.1 Exclusion of certain categories of data

In order to prevent discrimination, § 4 (3) of the PNR Act excludes protected characteristics from processing. This is in line with the approach of § 56 of the Federal Data Protection Act defining stricter conditions and safeguards for the processing of protected categories of data. These restrictions do, however, only focus on the protected characteristics, and therefore can only prevent direct discrimination. In order to identify the dis-

[126] Knobloch (n 1) 30; Sommerer, Self-imposed Algorithmic Thoughtlessness and the Automation of Crime Control (n 38) 85–87.

[127] Sommerer, *Self-imposed Algorithmic Thoughtlessness and the Automation of Crime Control* (n 38) 87–93; Hauke Bock and Katrin Höffler, 'Künstliche Intelligenz und Kriminalität' [2022] KriPoZ 257, 262.

[128] Henning Hofmann, *'Predictive Policing'* [Duncker & Humblot 2020], 281-283.

[129] Carsten Orwat, *Diskriminierungsrisiken durch Verwendung von Algorithmen* (Nomos 2019) 62–66 <https://www.antidiskriminierungsstelle.de/SharedDocs/downloads/DE/publikationen/Expertisen/studie_diskriminierungsrisiken_durch_verwendung_von_algorithmen.pdf?__blob=publicationFile&v=3> accessed 9 August 2022.

criminatory potential of other data or so-called proxies, some commentators even consider it to be necessary to process the protected characteristics and evaluate potential correlations to otherwise 'neutral' data.[130]

4.3.2 The relevance of human intervention – risks of automation bias

A huge majority of commentators agrees that automated 'predictive policing' systems must never replace human decisions and responsibilities – and that therefore individual users also should not follow automated findings blindly. It should always be up to a human officer to scrutinize the relevant results before responding to them with potential follow-up measures.[131] Critical voices raise doubts as to whether it can be realistically expected from relevant users to maintain a critical attitude towards such automated support tools, and even warn against inappropriate trust in automated recommendations ('automation bias').[132] As has been seen with regard to location-based 'predictive policing' systems, even a human operator might feel a strong incentive to follow automated results, either because of an inappropriate trust in the relevant technology, or because doing so might seem less controversial and easier to justify.[133] To make it worse, there are also indications that deployment of 'predictive policing' might even increase racial profiling among human operators. In this regard, *Egbert* points to empirical studies showing that patrol officers who are deployed to 'high-risk' locations are more likely to act suspicious towards individuals within these areas, especially towards individuals fulfilling certain visible stereotypes.[134]

4.3.3 Transparency and explainability

These concerns become even more relevant in case the police officers in question do not have an understanding of the way the relevant technology works, especially with regard to 'predictive policing' models that operate on machine learning technology which cannot be explained with recourse to a certain theory. Therefore, transparency and explainability seem to be crucial for an effective human control. Research on the question of how these can be provided even for machine learning systems – for example through ex-post validation[135] – is still in an early stage.[136]

4.3.4 Surveillance and chilling effects

On another note, critical voices complain that 'predictive policing' typically affects a broad range of individuals without any (or at least any clear) relation to the crimes which

[130] Rademacher (n 5) 265–266; Wischmeyer, 'Predictive Policing, Nebenfolgen der Automatisierung von Prognosen im Sicherheitsrecht' (n 8) 205.
[131] Hofmann (n 128) 292.
[132] Tobias Singelnstein, 'Predictive Policing: Algorithmenbasierte Straftatprognosen zur vorausschauenden Kriminalintervention' [2018] NStZ 1, 4.
[133] See **2.5** above.
[134] Egbert, 'Predictive Policing als Treiber rechtlicher Innovation?' (n 10) 39–42.
[135] Benedikt Kohn, Künstliche Intelligenz und Strafzumessung (Nomos 2021) 284.
[136] Sommerer, Personenbezogenes Predictive Policing (n 123) 206–221.

the software aims to prevent. 'Predictive policing' tools, which process personal data before any suspicion is established, are partially viewed as a disproportionate interference with the right to privacy. In this manner, the prevalence of 'predictive policing' could contribute to a feeling of 'overall surveillance'.[137] There is substantial concern that this can lead to chilling effects among the population, and impair the right to freedom of expression. Therefore, some consider the existing normative framework to be insufficient in order to safeguard compliance with the standards of the German Constitutional Law or the Law or Fundamental Rights in EU Law.

In order to monitor the overall proportionality of the interference with the right to privacy, many stress the need for a constant overview of all surveillance and similar measures ('Überwachungsgesamtrechnung').[138] The parliamentary coalition forming the current German government agreed to establish such an overview, and to conclude an independent scientific evaluation of all legislation on security matters, including their effects on freedom and democracy, as well as considering technological developments, until the end of 2023.[139]

5 Addendum: The German Federal Constitutional Court's 2023 Judgment on Automated Data Analysis

On 16 February 2023, the Federal Constitutional Court issued a landmark judgment on automated data analysis by police forces, in particular by making use of self-learning technologies.[140] A German NGO had filed a constitutional complaint against two Länder provisions regulating – and allowing – the automated analysis of large databases, namely § 25a HSOG and § 43 PolDVG Hamburg (see **2.3.2** above), for the purpose of preventing crimes and averting dangers resulting from the commission of crimes.

5.1 Infringement of the right to informational self-determination

In its judgment, the Federal Constitutional Court repeated its viewpoint that any automated data analysis encroaches into the fundamental right of data protection of all persons whose personal data are analysed, not only into the rights of those persons who are listed in the results of an analysis (see **3.2.1** above), and therefore requires justification in

[137] Bock and Höffler (n 127), 263 (with regard to video surveillance).

[138] Federal Commissioner for Data Protection and Freedom of Information, 'Einsatz von Künstlicher Intelligenz im Bereich der Strafverfolgung und der Gefahrenabwehr' (report on the public consultation process, 23 March 2022) <https://www.bfdi.bund.de/SharedDocs/Downloads/DE/Konsultationsverfahren/2_KI-Strafverfolgung/Konsultationsbericht.pdf;jsessionid=A459A4FEDC511B17C2035C0FC5C5ADB9.intranet241?__blob=publicationFile&v=3> accessed 9 August 2022; Rademacher (n 5) 268.

[139] Coalition agreement, lines 3638–3643 <https://www.wiwo.de/downloads/27830022/8/koalitionsvertrag-2021-2025.pdf> accessed 9 August 2022.

[140] BVerfG, judgment of 16 February 2023 – 1 BvR 1547/19, 1 BvR 2634/20 *Palantir* EECLI:DE:BVerfG:2023:rs20230216.1bvr154719. An English translation is available at https://www.bverfg.de/e/rs20230216_1bvr154719en.html.

law (see **3.2.2** above).[141] With regard to such justification, it highlighted that automated data analysis typically has a higher impact on fundamental rights than any 'normal' processing of data and also goes beyond the 'normal' re-purposing of personal data: According to the court, the use of such tools

> "enables large amounts of complex information to be processed. Depending on the analysis method used, the linking of existing datasets can generate new, otherwise inaccessible information that affects the personality rights of those affected. The measures in question thus intensify the generation of information from the data. Apart from extracting intelligence that was present in the data but had not yet been discovered because the datasets were not yet linked, this process can also come close to developing full profiles of the persons concerned […]. This is because the software can open up new possibilities of filling in the available information on a person by factoring in data and algorithmic assumptions about relationships and connections surrounding the person concerned. By combining personal and non-personal data, coupled where applicable with the fact that algorithms typically take into account mere correlations, new insights that would not otherwise be visible or detectable can be generated in ways that affect the personality rights of those concerned. The process vastly improves the effectiveness of conventional investigation methods, where authorities operate by gradually piecing together ever more information".[142]

5.2 Proportionality of legal bases justifying automated data analysis

Based on this assessment, the Federal Constitutional Court sets a higher bar for the proportionality test that any legal basis for the use of automated data analysis needs to pass but does not prohibit the use of such tools altogether. In particular, the Court accepts the argumentation brought forward by the government that the purpose – crime prevention – may be served by the use of automated data analysis, and that there are no alternatives available having fewer human rights implications.[143] Yet, in the adequacy test (proportionality *strictu sensu*), the peculiarities of automated data analysis need to be taken into account. In its analysis of the adequacy of the legal basis of automated data analysis, the court differentiates four dimensions:

Firstly, it points out that the severity of automated data analysis depends on the data sources. If personal data is acquired during normal police work, that already requires a specific justification and therefore prohibits overarching data gathering and analysis.[144]

---

[141] BVerfG, judgment of 16 February 2023 – 1 BvR 1547/19, 1 BvR 2634/20 *Palantir* EECLI:DE:BVerfG:2023:rs20230216.1bvr154719 headnote 1 and para 50.

[142] BVerfG, judgment of 16 February 2023 – 1 BvR 1547/19, 1 BvR 2634/20 *Palantir* EECLI:DE:BVerfG:2023:rs20230216.1bvr154719 para 52–53.

[143] BVerfG, judgment of 16 February 2023 – 1 BvR 1547/19, 1 BvR 2634/20 Palantir EECLI:DE:BVerfG:2023:rs20230216.1bvr154719 para 69; see also headnote 2.

[144] Cf. BVerfG, judgment of 16 February 2023 – 1 BvR 1547/19, 1 BvR 2634/20 *Palantir* EECLI:DE:BVerfG:2023:rs20230216.1bvr154719 para 80.

For data originating from special investigation measures – such as the secretive exfiltration of data from IT systems –, specific purpose-limitation requirements need to be met.[145] In contrast, "if the data analysis/interpretation platform is connected to the internet, this increases the severity of interference because it facilitates the processing of especially large amounts of data",[146] same as the exchange of data between law enforcement or even intelligence agencies.[147]

Secondly, it points out that systems generating object or area-related information (aggregated information) are less intrusive than systems trying to obtain intelligence about specific persons. The same holds true for systems obtaining intelligence about 'usual suspects', in contrast to systems focusing also on persons not yet in the view of the police.[148]

Thirdly, it refers to factors influencing the severity of the processing of data: For instance, it is of relevance whether data "files may not be included automatically but must be added manually for each data analysis/interpretation measure".[149] The processing of data is of low severity if the algorithm is nothing more than a simple search ("the process resembles a rudimentary cross-checking operation") which could be done by hand, although it might take very long.[150] The severity increases with the complexity of the algorithm, such as "when data analysis/interpretation is not based on a particular search term, at least not on a search term related to already known facts, but where the analysis/interpretation process is aimed entirely at identifying distinctive statistical features in the available data – distinctive features which, in additional steps, are (automatically) linked with information in other datasets and can then give rise to further intelligence that the police did not previously have any grounds to search for".[151] With regard to 'artificial intelligence' the Court points out:

> "The use of self-learning systems – i.e. artificial intelligence or AI – can interfere with fundamental rights in a particularly intrusive manner depending on the particular use in question. The advantages of such systems – as well as the specific dangers they pose – lie in the fact that they do not simply apply the criminologically sound profiles used by individual police officers, but rather that they automatically refine these profiles, or in some cases even create entirely new

[145] BVerfG, judgment of 16 February 2023 – 1 BvR 1547/19, 1 BvR 2634/20 Palantir EECLI:DE:BVerfG:2023:rs20230216.1bvr154719 para 81.
[146] BVerfG, judgment of 16 February 2023 – 1 BvR 1547/19, 1 BvR 2634/20 *Palantir* EECLI:DE:BVerfG:2023:rs20230216.1bvr154719 para 88.
[147] BVerfG, judgment of 16 February 2023 – 1 BvR 1547/19, 1 BvR 2634/20 *Palantir* EECLI:DE:BVerfG:2023:rs20230216.1bvr154719 para 79.
[148] BVerfG, judgment of 16 February 2023 – 1 BvR 1547/19, 1 BvR 2634/20 *Palantir* EECLI:DE:BVerfG:2023:rs20230216.1bvr154719 para 94, 96–8.
[149] BVerfG, judgment of 16 February 2023 – 1 BvR 1547/19, 1 BvR 2634/20 *Palantir* EECLI:DE:BVerfG:2023:rs20230216.1bvr154719 para 88.
[150] BVerfG, judgment of 16 February 2023 – 1 BvR 1547/19, 1 BvR 2634/20 *Palantir* EECLI:DE:BVerfG:2023:rs20230216.1bvr154719 para 91.
[151] BVerfG, judgment of 16 February 2023 – 1 BvR 1547/19, 1 BvR 2634/20 *Palantir* EECLI:DE:BVerfG:2023:rs20230216.1bvr154719 para 92–4.

> ones, and then continue to combine them during further stages of the analysis. Using complex algorithms, automated data processing software is thus capable of going beyond the mere identification of relationships and connections, and can begin autonomously producing further evaluations in the manner of 'predictive policing'. This enables particularly far-reaching insights and assumptions to be generated about a person. The verification of such information can be difficult in practice because, over the course of the machine learning process, complex algorithmic systems can increasingly detach themselves from the human programming that created them, with the machine learning process and the results generated becoming increasingly difficult to scrutinise […]. State oversight over the technology could then be rendered impossible. Furthermore, if software from private actors or foreign states is deployed, there is a risk that third parties could manipulate or gain access to data in undetected ways […]. Another specific challenge is to prevent the emergence and application of algorithmic discrimination. Self-learning systems may only be used in police work if special procedural safeguards are in place to ensure that sufficient levels of protection are guaranteed despite the reduced possibilities for exercising scrutiny."[152]

Fourthly, the court takes aspects such as transparency, the handling of errors, legal remedies, and administrative oversight into account.[153]

Noting further that important factors must be decided by the legislature itself and may not be passed on to the authorities, the Court stated that a broad and generic legal basis for the use of automated data analysis systems suffices only if there are sufficient safeguards in law. Such safeguards may relate to the data sources, on the intelligence to be obtained and/or the persons affected.[154] If the legislature does not exclude the use of machine learning technology, it must implement "special procedural safeguards [that] ensure that sufficient levels of protection are guaranteed despite the reduced possibilities for exercising scrutiny."[155] As the legal basis under constitutional review were far-reaching generic legal bases without sufficient safeguards, it declared them unconstitutional.[156]

---

[152] BVerfG, judgment of 16 February 2023 – 1 BvR 1547/19, 1 BvR 2634/20 *Palantir* EECLI:DE:BVerfG:2023:rs20230216.1bvr154719 para 100.

[153] BVerfG, judgment of 16 February 2023 – 1 BvR 1547/19, 1 BvR 2634/20 Palantir EECLI:DE:BVerfG:2023:rs20230216.1bvr154719 para 102–3.

[154] BVerfG, judgment of 16 February 2023 – 1 BvR 1547/19, 1 BvR 2634/20 *Palantir* EECLI:DE:BVerfG:2023:rs20230216.1bvr154719 para 103–22.

[155] BVerfG, judgment of 16 February 2023 – 1 BvR 1547/19, 1 BvR 2634/20 *Palantir* EECLI:DE:BVerfG:2023:rs20230216.1bvr154719 para 100.

[156] BVerfG, judgment of 16 February 2023 – 1 BvR 1547/19, 1 BvR 2634/20 *Palantir* EECLI:DE:BVerfG:2023:rs20230216.1bvr154719 para 123–73.

5.3 Assessment

This judgment highlights the legal complexity of assessing the constitutional and human rights implications of 'artificial intelligence'. If tens of factors influence the adequacy of a legal basis justifying automated data analysis, that increases, on the one hand, the legal uncertainty of whether a specific legal basis is within the boundaries of the constitution. On the other hand, this judgment sets out at least some barriers and offers guidance to the German legislatures by structuring the discussion on the lawfulness of automated data analysis and the use of self-learning systems in particular.

Selected Literature

Egbert S, 'Predictive Policing als Treiber rechtlicher Innovation?' (2021) 41 Zeitschrift für Rechtssoziologie 26

Lind F, *Das raumbezogene Predictive Policing in Deutschland. Der aktuelle rechtliche Rahmen und seine Indikationen für Weiterentwicklungen des Einsatzes prädiktiver Analytik bei präventiv polizeilichem Handeln* (forthcoming)

Rademacher T, 'Verdachtsgewinnung durch Algorithmen. Maßstäbe für den Einsatz von predictive policing und retrospective policing' in Zimmer D (ed) *Regulierung für Algorithmen und Künstliche Intelligenz* (Nomos 2019) 229

Sommerer LM, *Self-imposed Algorithmic Thoughtlessness and the Automation of Crime Control* (Nomos 2022)

Sprenger J, 'Verbrechensbekämpfung' in Ebers M and others (eds), *Künstliche Intelligenz und Robotik* (Beck 2020)

PREDICTIVE POLICING IN THE SPANISH LEGAL SYSTEM: A CRITICAL APPROACH

By Jordi Gimeno Beviá [*]

Abstract

Owing perhaps to the lack of specific regulations, Spain has limited experience in AI for predictive policing. However, the State Security Forces and Bodies use relevant instruments, Veripol and Viogen, in a generalized way. Additionally, initiatives that do not enjoy such widespread use have been initiated but subsequently abandoned or have remained at the level of pilots. These are, among others, the Geographical Information Systems (G.I.S) with a strong impact on fundamental rights and some initiatives promoted by private entities. This paper aims to critically analyze those instruments and their impact on the Spanish legal system. It also discusses the fit of the future uses of AI in predictive policing tools. Starting from the respect of fundamental rights our approach is not immobile, bearing in mind the advantages offered by AI in acceleration and greater efficiency of the criminal investigation.

1. National practices

In Spain, there is no uniform and univocal definition of the concept of predictive policing. One of the most relevant definitions has been given by the Organization for Security and Cooperation in Europe (OSCE) in 2017 as 'the systematic collection and evaluation of data and information, through a defined analytical process, which turns them into strategic and operational analytical products that serve as the basis for an improved, informed and documented decision-making process'.[1] An approximation of the national doctrine is offered by MIRÓ LLINARES who includes 'predictive policing' within Police

[*] Tenured Professor of Procedural Law UNED (SPAIN). This paper has been written through the research project 'Transición Digital de la Justicia' (IP. Dra. Sonia Calaza López) Proyecto estratégico orientado a la transición ecológica y a la transición digital del Plan Estatal de investigación científica, técnica y de innovación 2021-2023, en el marco del Plan de Recuperación, Transformación y Resiliencia, Ministerio de Ciencia e Innovación, financiado por la Unión Europea: Next Generation UE, con REF. RED 2021-130078B-100. I want also to thank Prof. Lelieur as a general rapporteur of Section III for all her great help and Prof. Miró Llinares and Nieto Martín from the Spanish Group AIDP.

[1] See OSCE Guide on Intelligence-Based Policing, 2017, p. 6: <https://www.osce.org/files/f/documents/6/4/455536.pdf> accessed 6 November 2023.

Artificial Intelligence and defines it as the 'application of quantitative techniques to identify targets of police interest for the purpose of reducing criminal risk by preventing future crimes or solving past crimes'.[2]

Predictive policing is a relatively new phenomenon in Spain and there is very limited experience – perhaps due to the absence of specific regulation – in Spanish application of AI for police prediction. Both in artificial intelligence and in other technological matters, Spain maintains a prudent approach. It seems that it will follow other Member States of the European Union and currently, as is known, the European Parliament is very cautious when implementing these mass surveillance systems.[3]

Thus, internally, the debate remains alive because the parliamentary group *Unidas Podemos* presented a Proposal on the use of AI in the tasks of surveillance and use of personal data of citizens by the State Security Forces and Bodies (FCSE).[4] They also have proposed the creation of an Algorithm Control Agency to ensure its transparency. The reason behind these initiatives is that, unfortunately, many of the 'AI solutions' tend to be implemented by private companies in public institutions, most of them through public contracts with very limited competition since there are very few national companies focused on the use of these new technologies. Beyond the above, there are instruments based on AI that the FCSE usually uses when carrying out their investigations.[5]

On the next pages, we will introduce these instruments which, for a better classification, we will group into two blocks. First, we will highlight the systems currently used by the FCSE, which are Veripol and Viogen. Second, we will mention those initiatives that do not enjoy such widespread use, have been initiated but subsequently abandoned, or have only been pilot experiences. Third, we will describe an interesting case of surveillance through a private company (Mercadona case).

[2] F. Miró Llinares, 'Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots', in *Revista Penal de Derecho y Criminología*, 3ª Época, nº 20, July 2018, pp. 87-130, p. 100.
[3] See the European Parliament resolution warning of the risks to our system of guarantees and freedoms of these technologies: <https://www.europarl.europa.eu/news/es/press-room/20210930IPR13925/uso-policial-de-la-inteligencia-artificial-el-pe-contra-la-vigilancia-masiva> accessed 6 November 2023.
[4] In Spanish, we say 'Fuerzas y Cuerpos de Seguridad del Estado' FCSE.
[5] For a first approach to this phenomenon in our country, see J. L. Gonzalez-Álvarez, J. Santos-Hermoso, and M. Camacho-Collados, 'Policía predictiva en España. Implementación y retos futuros.' In *Behaviour & Law Journal*, vol. 6, year 2020, pp. 26-41.

1.1 Systems currently used by FCSEs

1.1.1 Fight against gender violence through VioGen

As everybody knows, gender violence is not only a Spanish problem but a scourge suffered in many other countries. It is a very particular kind of violence that cannot only be addressed through a punitive response but also requires a preventive and comprehensive strategy. Therefore, we can consider whether new technologies and artificial intelligence can help in the fight against gender violence.

Perhaps in Spain, the most recognized tool in the application of AI is the VioGen program, aimed at preventing gender violence. VioGen stands for *Comprehensive Monitoring System in Cases of Gender Violence* and was created on July 26, 2007.

The main virtues of VioGen are the following: 1) It makes predictions of risk: VioGen is a type of predictive policing that, as we will see, facilitates estimates of the risk of recidivism in gender violence; 2) It allows monitoring and protection of the victims: Thanks to VioGen, the police can track and determine the appropriate protection measures for the victims in each situation; 3) It permits the integration of all the information in one system: thanks to VioGen, gender violence does not understand borders in Spain and the police upload the information to the system, which allows greater coordination; 4) Finally, it brings together different public organizations: Spain is divided into Autonomous Communities so we could assimilate it into a Federal State. Therefore, we have the Civil Guard and National Police but there are also Communities such as Catalonia or the Basque Country that have their own security forces and bodies (*Mossos de Esquadra* or *Ertzaintza*). There are also local police. Now all of them use VioGen to fight against gender violence.

Concerning the functioning of VioGen, while it seems obvious that it fits on a predictive police tool, we cannot ensure, because there is a lack of information on that point, that it uses AI. At least if VioGen uses some AI, is it clear that it does not rely on machine learning. As it was created in 2007, at that time there was no use of AI in criminal investigations. The Gender Violence area of the Secretary of State for Security (Ministry of the Interior), incorporated AI into VioGen in 2020 through the analytical platform of the software company SAS Iberia.[6] Nonetheless, the source code of the system is not public, so we do not know which kind of AI uses, if the system is based on logical functions, or if a

[6] See description of the application of A.I to tool in the own website of SAS Iberia: <https://www.sas.com/es_es/news/press-releases/locales/2020/viogen-secretaria-estado-seguridad-y-sas-unidos-lucha-contra-violencia-genero-analitica-avanzada-ia.html> accessed 6 November 2023.

differentiated weight has been given to risk indicators – it seems that this is the technique used. Therefore, having access to how VioGen was configured would be important to understand it and it would help to know the elements that have been decisive in producing the risk result according to each case.

At the operational level, utilizing VioGen is, *a priori*, quite simple. The system, in general, considers two major factors: the dangerousness of the aggressor and the vulnerability of the victim. The police officer must complete two questionnaires: first the Police Risk Assessment (VPR) and then the Police Assessment of Risk Evolution (VPER). Once the police agent has completed both questionnaires, she can confirm the risk assigned by VioGen or, attending to other factors – for example, body language or others not included in the questionnaires – modify the level of risk. Therefore, it is an assistance tool but never a decision-making tool because, finally, it is for the police to have the last word about the risk assessment. Thus, the requirement of 'a human in command' is met.

The first questionnaire, the VPR, is usually completed by the police officer when she receives a complaint of gender-based violence, either from the victim or from a family member or acquaintance. The VPR protocol is a mechanized procedure of information from four types of sources: (1) assessment of the violent incident reported – in order to take a first approach of the general risk –, (2) background of the aggressor; (3) vulnerability of the victim and (4) the victim's self-perception of the situation. For example, some of the questions asked in the block related to the history of violence are: Has there been any violence on the part of the aggressor? Has the aggressor used weapons? Has the victim received threats? Is there exaggerated jealousy in the last six months? Risk indicators are extracted from these factors, following the previous examples: what type of violence exerts (physical or mental) the use or access to weapons by the aggressor, etc.

After the police has completed the VPR questionnaire, VioGen assigns a level of risk: unappreciated, low, medium, high, or extremely high. Each risk level must be reviewed within a certain period: 3 months if 'not appreciated', 60 days if 'low', 30 days if 'medium', 7 days if 'high', and 72 hours if 'extremely high'. In addition, each level triggers different police protection measures: for example, when the risk is extremely high, the woman has permanent police surveillance at home.

However, gender-based violence is not static, but rather dynamic. Thus, in second place, once a level of risk and protective measures are assigned, it is important to analyze how the risk has evolved and whether the established measures have worked. This is possible through the VPER. This form consists of 43 indicators, also dichotomous, of which 34 are

about risk and about 9 protections, all grouped into 5 criminological dimensions: the four of the VPR and a new dimension of dynamic-relational indicators to monitor the risk and update the protection measures applied at first. The VPER is subdivided into two forms that depend on 1) If there has been an incident and 2) If there has been no incident since the measures were established. If there has been no incident, a lower risk level can be determined. In contrast, if there has been an incident, a personalized protection plan will be tailor-made.

It is very complex to measure or quantify the effectiveness of any instrument when it comes to evaluating its impact on a scourge such as gender-based violence. Anyway, the authorities consider that the percentage of reliability of VioGen is mainly satisfactory: since the launch of this tool in 2007, the recidivism of aggressions has decreased by 25% according to the latest data. While, in a generic way, recidivism in other neighboring countries reaches 35%, in Spain, it has decreased to 15%.[7] The assessment data is regularly updated on the website of the Ministry of the Interior, which allows an analysis of the reliability and effectiveness of the tool from a statistical perspective.[8]

Beyond the constant evaluation by the Ministry of the Interior, VioGen is also being evaluated externally. An example is the autonomous evaluation carried out by the non-profit organization *Eticas Foundation*. They exposed the difficulties of carrying it out because, according to its own words: 'This lack of transparency and explainability implies that we cannot know if VioGen tends to estimate a risk too high or too low in certain types of cases, such as when the complainants belong to a particular social group, such as immigrants who speak Spanish (or Catalan n or Galician) in a different way than those who have always spoken the language usually express themselves'.[9]

There are also other problems beyond the lack of transparency. The first problem is the lack of resources. It is difficult to offer optimal measures to all victims. In addition, in some cases, police officers choose to modify the level of risk to extremely high, which implies very expensive measures such as continuous police surveillance. Moreover, the

[7] Information obtained from an interview conducted by the newspaper La Vanguardia to one of the creators and Head of VioGen area Juan José López Ossorio in 2017. Available at the following link: <https://www.lavanguardia.com/tecnologia/20190519/462147339117/viogen-violencia-de-genero-violencia-machista-inteligencia-artificial-algoritmos.html> accessed 6 November 2023.
[8] The data of 2022 are available at the following link: <https://www.interior.gob.es/opencms/es/servicios-al-ciudadano/violencia-contra-la-mujer/estadisticas-sistema-viogen/> accessed 6 November 2023.
[9] See the information on their website: <https://eticasfoundation.org/es/viogen-un-algoritmo-para-predecir-el-riesgo-de-reincidencia-en-casos-de-violencia-de-genero/> accessed 6 November 2023.

lack of means sometimes hinders sudden changes in risk, for example from 'not appreciated' to 'extremely high' when an aggressor leaves prison.

The second problem is the lack of specialized training. Currently, in Spain, there are more than 40,000 VioGen users. However, the training they receive is far from adequate, and in many cases, the police officers, who do not have to be specialists in gender violence to be allowed to use VioGen, are not able to detect other risk factors. They usually retrace some kind of course but there is no evaluation on whether they can handle VioGen properly.

Thirdly, a double-checking system is missing. The police officer confirms or modifies the level of risk of the questionnaire, but it would be more appropriate if he was not a single officer but had the support of someone more qualified to review it. We can illustrate these problems with two court sentences. In the first case, the State was condemned to pay civil liability after the police had misapplied VioGen. The level of risk indicated by the tool was 'not appreciated' and consequently no specific measures were taken, however, the woman was finally murdered by her husband/partner. This happened because the agents did not modify the level of risk, which was very high. After all, the aggressor had a criminal record outside Spain, and this was not reflected in the tool.[10] Therefore, the agents should have modified the assessment and assigned a higher level of risk. In the second case, the Military Chamber of the Supreme Court convicted a Civil Guard because he refused to use the tool although its use is mandatory for all State security forces and bodies[11]

In conclusion, in a country with independence and decentralizing tensions, VioGen has allowed the authorities to act in a coordinated manner against gender violence. It has also allowed a more individualized follow-up of cases of gender-based violence and the control of the protection measures implemented.

1.1.2 Veripol

The Veripol system, launched in 2018, focuses on preventing false complaints, which are punished at art. 457 of the Spanish Criminal Code. In addition to being the first tool of its kind in the world, it has an accuracy of more than 90% and estimates the probability that a complaint for theft with violence and intimidation or pull is false. It deters, among

[10] Spanish National Court, (Audiencia Nacional) specifically the Contentious-Administrative Chamber, in the Judgment of September 30, 2020; is there a more specific reference?
[11] STS, Fifth Chamber, 73/2020, of October 28.

other actions 'spurious' complainants, for example, those who invent the theft of a mobile phone for the sole purpose of collecting the insurance previously contracted.

To do this, the tool feeds on a large amount of data (*big data*) and determines, based on the content of the information provided, the percentage of probabilities of falsity of the complaint, using natural language processing (NLP). For its implementation, the application passed different tests of operation, nourishing itself from a databank of more than 1000 complaints for robbery with violence and intimidation that were presented in Spain during the year 2015. Approximately 50% of these complaints were true and the other 50% were false. The model, in which several officials worked for more than two years, allows us to appreciate the differences that may exist between the narration of complaints that have turned out to be true and false, based on the information provided by the complainant, morphosyntax and a wide amount of detail.

Despite the positive aspects of the tool, some authors expose important shortcomings. Thus, in the words of Jaume Palasí, 'Body language also matters in the complaint and here it does not appear. This system creates ideal types. It does not describe reality, but artificially establishes a mechanized description of reality. Reality is more dynamic than just a few words'. In the same way, that some qualify the percentage of 91% of reliability/correctness as a success,[12] is seen by others, such as Baeza Yates, in their own words 'That 9% is wrong implies that the system wrongly accuses nine out of every 100 people. And this is a very serious ethical conflict'.[13] Likewise, ethics experts miss specific regulations, as is the case in other countries (Japan, Finland, etc.) that have already faced this reality.[14]

Moreover, from a procedural law perspective, it seems to violate the position of the victim, whose statement is questioned by an agent, encouraged by the application. It implies, therefore, an exchange of roles in which the victim of a crime automatically passes

[12] This has been clearly stated by the creators of the tool on which have been participated researchers from University Carlos III, University Complutense, Univertisty from Rome La Sapienza and the Ministry of Internal Affairs. They say that 91% success rate is 15 points higher than experienced agents on this kind of crime: <https://www.ucm.es/otri/veripol-inteligencia-artificial-a-la-caza-de-denuncias-falsas> accessed 6 November 2023.

[13] In the same sense, Alonso Salgado indicates that 'although, obviously, the estimation of VeriPol does not compromise the decision of the Security Forces and Bodies, there is no doubt that it establishes a starting bias...' in C. Alonso Salgado, 'Acerca de la inteligencia artificial en el ámbito penal: especial referencia a la actividad de las fuerzas y cuerpos de seguridad', Us et Scienctia, vol. 7, pp. 25-36, 2021.

[14] See expert views in this article: <https://elpais.com/tecnologia/2021-03-08/veripol-el-poligrafo-inteligente-de-la-policia-puesto-en-cuestion-por-expertos-en-etica-de-los-algoritmos.html> accessed 6 November 2023.

to the position of alleged perpetrator of another. And, although the final decision rests with the agent, in most cases he will not depart from the forecast/suggestion of the software.

1.2. Systems initiated and/or abandoned by FCSEs

As for the systems whose use has not been widespread in the FCSEs, it is worth highlighting mainly G.I.S or geographic information systems, which are usually used to prevent crime in high-risk places through a kind of 'digital crime maps' and the creation of hot spots where criminal activity is concentrated.

1.2.1 EuroCop PredCrime

Since 2011, in the field of citizen security, different Public Administrations, mainly Local Entities – local police –, have been raising the possibility of equipping themselves with the EuroCop PredCrime software. The software, as defined on the web, consists of 'the experimental development of an Integrated System for the treatment of massive data linked to crimes and misdemeanors already committed, based on the use of mathematical models and algorithms, which allows the prevention and resolution of a crime not yet produced'.[15] It integrates and processes massive data linked to crimes and bases its operation on a spatiotemporal model and geographic information of heat-maps through models and mathematical algorithms for the prevention, through the forecast/prevision of crimes that could be committed in the future.

Despite the 'Euro' word at the beginning, it is a tool created by a private company and it does not receive European funds. Municipalities for the protection of their towns contracted the software. However, it is not possible to know the scope of the contracts, and it seems that they might be contrary to the criteria of the RGPD. Some of the EuroCop Pred Crime systems were 'temporarily abandoned' by the City Councils that had signed it, such as Rivas Vaciamadrid (Madrid).[16] We intuit that due to lack of guarantees or legal

[15] <https://www.eurocop.com/catedra-eurocop/proyectos-en-marcha/eurocop-pred-crime-sistemas-para-la-prediccion-y-prevencion-del-delito/> accessed 6 November 2023.
[16] See, in the following link, its use by Rivas Vaciamadrid: <https://www.rivasciudad.es/noticias/organizacion-municipal/2015/12/10/un-sistema-pionero-en-prevencion-de-delitos/862600041423/> and its abandonment a few months later in this other: <https://rebelion.org/el-estado-policial-espanol-2-0-tecnologias-de-empresas-privadas-para-vigilar-a-los-ciudadanos/> accessed 6 November 2023.

basis for their use. However, it is unknown if they finally implemented it or if once implemented, they had to abandon it due to its impact on fundamental rights or the lack of sufficient regulation.

The main problem lies in private participation not only in public security – which usually happens in not a few enclosures – but in the management of data and sensitive information usually collected in police databases. Indeed, the tools that arise from a public-private partnership can lead to profound problems of legality. Hence, at the time of writing this paper, it is not possible to speak of a generalized use by the FCSE but rather the opposite, because given the doubts there is no evidence that they are currently used.[17]

1.2.2 Predictive Police Patrolling (P3-DSS)

A pilot study was developed in 2017 by the National Police Corps (CNP) in the central district of Madrid, entitled *Predictive Police Patrolling* (P3-DSS). It allows, through algorithms, to forecast crimes, knowing their typology as well as improving the efficiency of police patrol shifts. It is a predictive policing tool, but it does not use AI. The project was devised by the policeman and mathematician Miguel Camacho, and part of it can be seen in his doctoral thesis entitled *Statistical Analysis of Spatio-Temporal Crime Patterns: Optimization of Patrolling Strategies*, defended in 2016.[18]

This application, focused on violent assaults and robberies, refers to crime prevention and improvement of efficiency in patrolling. It can do that through the development of Geographic Information System (GIS) techniques, which allows the police to manage in a reasonable time spatiotemporal data that helps to identify concentrations of criminal acts. Therefore, a predictive patrolling model provides greater efficiency in the distribution of patrols according to criminal risk.[19] For the use of the pilot tool used in the Central

[17] Ekaitz Cancela and Aitor Jiménez, journalists from El Salto who, after a thorough investigation, warn of the risks posed by Ekaitz Cancela and Aitor Jiménez, are very critical of the risks it poses. This tool. Thus, the following questions arise, which we reproduce literally: 'What compromised and private data can a company that lends and manages the critical digital infrastructure of police agencies have access to? Don't citizens have the right to know the inside of these black boxes? Do we want a private corporation to be in a position to offer "a solution that covers the integral management of the police, both in the operational aspect (automating all its operational, administrative, judicial tasks, etc., from anywhere and at any time), and in the tactical and strategic aspect in order to achieve maximum efficiency in police work?…' The result of the information, very critical with these predictive policing systems adopted by local police, available at the following link: <https://www.elsaltodiario.com/tecnologia/estado-policial-espanol-2.0-empresas-privadas-eurocop-vigilar-ciudadanos> accessed 6 November 2023.

[18] The thesis is open at the following link: <https://hera.ugr.es/tesisugr/26134081.pdf> accessed 6 November 2023.

[19] J. L. Gonzalez-Álvarez, J. Santos-Hermoso and M. Camacho-Collados, op. cit., p. 30.

District of the city of Madrid, criminal records were collected regarding the crime of theft (105,755 incidents) between 2008 and 2012. In turn, they used the Geographic Information Systems (GIS) of the CNP that integrates criminal events on a geographical map of the city, in addition to the location of police patrols.[20] Anyway, the forecast takes place in a misleading way: it is effective because the police act in a specific area and due to this effectiveness, it will keep sending police to this specific area. Consequently, the main problem is the creation of 'hot spots' that criminalize neighborhoods.

1.3 Surveillance through private companies (Mercadona case)

The private sector has timidly tried to establish predictive policing mechanisms but given the absence of a legal basis, its use has not become widespread. As a relevant example, we can bring up the system of facial recognition of Mercadona supermarkets. It works detecting people with firm convictions and precautionary measures to drive them out of the supermarkets (usually because of theft). This 'solution' was developed by the Israeli company AnyVision. The reality is that it did not offer much information about important issues, e.g. where they extracted the data and images of the condemned persons as well as the time it took to delete the images of other customers, … For all these reasons, Mercadona was fined 2.5 million euros by the Spanish Data Protection Agency and this software eventually ceased to be used.

Even the courts ruled against its use because they considered that it did not protect the public interest but only corporate interests. The court said: 'Not everything goes in terms of fundamental rights. These technologies can be truly intrusive and require a calm ethical and legal debate, as they can have very adverse effects on fundamental values and human integrity'. This is because facial recognition is not protecting public interests, but the private interests of the legal person, and 'the appropriate guarantees for the protection of the rights and freedoms of the persons concerned would be violated, not only of those who have been punished and whose access is forbidden to them but of other persons who access the supermarket".[21]

So, after the Mercadona case, no similar system has been implemented by private entities.

---

[20] M. Jiménez Hernández, "El big data como herramienta de prevención de la delincuencia ", página 28: <https://rua.ua.es/dspace/bitstream/10045/115934/1/EL_BIG_DATA_COMO_HERRAMIENTA_DE_PREVENCION_DE_Jimenez_Hernandez_Miguel_Angel.pdf> accessed 6 November 2023.

[21] Vid. Auto 72/2021 of 15 February, Audiencia Provincial de Barcelona, Sección 9ª, Rec 840/2021.

2. The fit of AI in the Spanish legal system

Once exposed the tools of predictive policing that timidly use AI in Spain, it is possible to consider whether a more intense AI with a more widespread use would fit in the Spanish legal system. For this, it becomes essential, among other basic premises, to start from the following.

- *Access to justice and presumption of innocence (art. 24 C.E):* As previously explained, through the Veripol system the roles of perpetrator and victim may be reversed. This is why, in case of a false positive, the victim could be deprived of her/his access to a judge. Going further, the presumption of innocence could be endangered because if the Veripol system detects a high probability of a false complaint, the person who comes to a police station as a victim could leave the place as a suspect. As we said, in 91% of cases it won't be like this but in the other 9%, they will be under investigation. So, according to Art. 24 Spanish Constitution, both access to justice – as a victim – and the presumption of innocence – as a suspected person – will be affected. Thus, even if the model is ruled by a "human-in-command" because the police agent has the last word on the decision, the rights at stake make necessary the maximum diligence in this task, even if we consider a more intense use of AI.
- *Right to equality and non-discrimination*: police prediction techniques have been questioned because they can collide with equal and non-discrimination rights. The Spanish Charter of Digital Rights, even it does not have normative force, provides in its right XXV 'Rights before artificial intelligence', specifically in its section 2. a) that 'The right to non-discrimination must be guaranteed regardless of its origin, cause or nature, in relation to decisions, use of data and processes based on artificial intelligence'. This relevant – but not legally binding – Charter was published 14th July 2021 with the aim of creating a frame of reference for all public authorities and to serve as a guide for future legislative projects. Consequently, although it does not directly address predictive policing tools, it provides the keys to set these tools in a way that respects both fundamental rights.[22]

[22] This concern also occupies doctrine. Thus, Nieva Fenoll warns about the use of big data in police research because data is randomly stored from people, neighborhoods, etc., despite this randomness, will have been selected according to the damages of the algorithm configurator, which implies that the results are not neutral, see J. Nieva Fenoll, Inteligencia artificial y proceso judicial, Ed. Marcial Pons, 2018, p. 151. In the same sense, Miró Llinares summarizes the problem posed: 'the predictive tools we are talking about come only to do what was already done and is done today in a traditional and manual way and probably with the same biases or more, adding, in some cases, a more systematic or scientific methodology' because, -continues the author-'... What we know so far tells us that algorithms, which accurately reflect our

- *AI must be public and accessible:* One of the main risks that we face as a society is that the AI used in the prevention of crime is configured, controlled, and executed by a few, moreover, of the private sector, and that could obey the interests of certain lobbies. Therefore, on the one hand, AI should be regulated and, on the other, it should be public and accessible to any citizen.[23]
- *Evaluation and review by independent persons or entities:* AI is constantly evolving, so it becomes essential that its application in Spain is periodically evaluated. Therefore, the tools that use this technology should be reviewed, mainly by a group of independent experts, if possible appointed by a public entity, but taking into consideration some help of private groups or entities – better if they are non-profit. Moreover, the question of labeling AI systems should be discussed in Spain.
- *Towards a relevant role of the Public Prosecutor's Office as guarantor of the proper functioning of AI supervising police operations:* It must be recalled that the Public Prosecutor's Office is, in accordance with its principles of action, an impartial party that must be both for the conviction of the guilty and for the acquittal of the innocent. Moreover, the Spanish Constitution, in Article 124, gives it a leading role in the defense of citizens' rights. Hence, it must be the guardian of the correct use of Artificial Intelligence in the judicial process and during police operations, monitoring its correct functioning and denouncing the infractions and violations of rights that may cause a negligent use of this technology.[24]
- *Training and information for police officers:* The widespread application of AI in the criminal process without even knowing the basis and operation of this technology can imply undesirable consequences. With this, it is not intended, much less, the acquisition of an expert level, from a computer-scientific perspective, of the knowledge and management of this technology. At least, the police relying on these tools must have sufficient technological knowledge for a good use of them.
- *Greater pedagogy and information towards society:* If it is important to bring police institutions closer to citizens, with greater reason they should clearly explain the role that AI will have in predictive policing tools.[25] There are already interesting

world, seem to reflect our prejudices as well', see F. Miró Llinares, "Artificial intelligence and criminal justice: beyond the harmful results caused by robots", op. cit., p. 126.

[23] In this sense, T. Armenta Deu, op. cit., p. 319.

[24] Nieva Fenoll, op. cit., p.150.

[25] See index of the European Union EU Justice of the year 2021: <https://ec.europa.eu/info/sites/default/files/eu_justice_scoreboard_2021.pdf> accessed 6 November 2023.

studies, which show how low the acceptance of this technology and its application to criminal justice and police investigation is. Today citizens distrust decisions that rely on algorithmic predictions.[26] Therefore, beyond being involved in a generalized digital transformation of society, we must be very clear, and pedagogical when explaining both the benefits of the application of this technology in the administration of justice and in criminal investigations.

3. Conclusion

The predictive policing tools currently used in Spain do not reflect a widespread application of AI. It is necessary to use it prudently and in line with what has happened in other countries and according to the law of the European Union, whose regulation on AI is still to come.[27]

Beyond the fact that the Spanish Government is enthusiastic about digital transformation and allocates a large amount of funds in its Justice 2030 plan, the truth is that there is a lack of specificity about the purpose of the application of AI in criminal investigations, beyond a generic approach of intelligent justice oriented to data.[28] If the objectives are not clearly defined at this key point of digital transformation, we run the risk that, on the one hand, innovation and entrepreneurship will be hindered and, on the other, we may encounter tensions with private initiatives or solutions that may fail in their attempt to be used by public authorities.

Thus, it is necessary to subscribe to a prudent approach, fleeing from extremes, without neglecting the advances and opportunities presented by this technology, but at the same time, ensuring respect for fundamental rights and guarantees.[29] A *totum revolutum* cannot be proposed, but the application of AI to criminal investigations must be carried out

[26] Vid, for all, A. Morales Moreno, 'Algoritmos en el estrado, ¿realmente los aceptamos? Percepciones del uso de la inteligencia artificial en la toma de decisiones jurídico-penales' Revista *Ius et Scientia* vol.7, nº2, 2021

[27] The necessary harmonisation in the context of the Union European, see, by all, M. De Hoyos Sancho, 'El uso jurisdiccional de los sistemas de inteligencia artificial y la necesidad de su armonización en el contexto de la Unión Europea' *Revista General de Derecho Procesal* nº55, 2021. Sobre el pronóstico en la aprobación del Reglamento, acierto la autora al considerar que es muy poco probable que entre en vigor antes de 2023, pág. 23.

[28] The work plan on digital efficiency can be seen at the following link: <https://www.justicia2030.es/eficiencia-digital> accessed 6 November 2023.

[29] In full line with the approach of Simón Castellano, which maintains an "ambivalent" position situated 'in the center of the extremes and that it tries to take advantage of the advantages of technical progress while warning of the ends and edges that it deploys, setting certain red lines' P. Simón Castellano, Precautionary justice and artificial intelligence, Ed. Bosch, 2021, p. 98.

gradually and calmly, establishing scientific reviews before implementation and after it.[30] Therefore, its implementation should not only be carried out separately by police officers and legal experts, on the one hand, and by computer scientists, on the other. Criminologists should play an important role taking into account that the tools must be evaluated and scientifically.

Finally, we should accept a result different from the one predicted by the AI. That does not mean that AI has been wrong – or even that it is misconfigured – and should not lead us to a hasty conclusion about the malfunction of this technology. A strong and widespread AI will be a reality sooner rather than later, and although in an assistive way, it will have an increasing presence in predictive policing. Let us not turn our backs on a technology that, although unknown, is fascinating, and let us prepare today for the police of tomorrow.

Selected literature

Alonso Salgado C, "Acerca de la inteligencia artificial en el ámbito penal: especial referencia a la actividad de las fuerzas y cuerpos de seguridad", *Us et Scienctia*, vol. 7

Armenta deu T, *Derivas de la justicia*, Ed. Marcial Pons, 2017

De Hoyos Sancho M, "El uso jurisdiccional de los sistemas de inteligencia artificial y la necesidad de su armonización en el contexto de la Unión Europea" *Revista General de Derecho Procesal* nº55, 2021

Gonzalez-Álvarez J.L., Santos-Hermoso J. y Camacho-Collados M, "Policía predictiva en España. Aplicación y retos futuros." En *Behaviour & Law Journal*, vol. 6., año 2020

Miró Llinares F, "Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots", en *Revista Penal de Derecho y Criminología*, 3ª Época, nº 20, julio 2018

Miró Llinares F, "Predictive policing: utopia or dystopia? On attitudes towards the use of Big Data algorithms for law enforcement", Revista IDP, nº 30, Marzo 2020

Morales Moreno A, "Algoritmos en el estrado, ¿realmente los aceptamos? Percepciones del uso de la inteligencia artificial en la toma de decisiones jurídico-penales" Revista *Ius et Scientia* vol.7, nº2, 2021

Nieva Fenoll J, *Inteligencia artificial y proceso judicial,* Ed. Marcial Pons, 2018

Simón Castellano P, Justicia cautelar e inteligencia artificial, Ed. Bosch, 2021

[30] In this sense, see, for all, MIRÓ LLINARES, whose positioning can be seen in detail in MIRÓ LLINARES, F. "Predictive policing: utopia or dystopia? On attitudes towards the use of Big Data algorithms for law enforcement", *IDP Magazine*, nº 30, March 2020, pp.1-8L

NATIONAL REPORTS ON PREDICTIVE JUSTICE

PREDICTIVE JUSTICE IN FRANCE

By Emmanuelle Gindre [*]

Abstract

This report attempts to take stock of predictive justice in criminal matters in France, in a context where questions and research on this subject are multiplying as fast as technologies using artificial intelligence are progressing. Without claiming to be exhaustive, it first looks at the concept of predictive justice in France, for which there is still no consensus due to the lack of a legal defini-tion. It then identifies the practices and tools implemented in the courts, often on an experimental basis, before analysing the reception given to these technologies, both by the doctrine and by the practitioners and professionals using them. Finally, the report draws up an inventory of the reg-ulations and other standards governing the use of artificial intelligence in the administration of justice, and highlights the various problems posed by these technologies, particularly with regard to the fundamental rights associated with criminal procedure.

1 General questions

1.1 Definition of 'predictive justice' in France

As the concept is currently understood, it seems that 'predictive justice' has existed in some countries since the 1950s (in particular in the United States, where it is called 'juri-metrics').[1] In France, it was introduced by the Digital Republic Act of 7 October 2016[2] instituting open-data judicial decisions, and studies concerning it have proliferated since. However, that Act did not define predictive justice, which is merely a consequence of using open data, and neither has any subsequent normative text or case law.

Predictive criminal justice has an older meaning, however, based on nineteenth-century Italian positivist doctrines and the idea of anticipating criminal activity through a 'prob-abilistic calculation of recidivism.'[3] Mireille Delmas-Marty described it as an application of the precautionary principle: 'With the "predictive" function, dangerousness replaces guilt, and punishment is replaced by risk prevention or even precaution in the face of uncertain risks (the risk of risks).'[4]. That definition of predictive criminal justice seems more circumscribed than the contemporary concept, as the aim is specifically to predict

[*] 1. Univ. Polynésie française, GDI EA 4240, Tahiti, Polynésie française; 2. UPPA, IFTJ, EA 7504, Centre de recherche sur la justice pénale et pénitentiaire, Pau, emmanuelle.gindre@upf.pf.

[1] S. Lebreton-Derrien, 'La justice prédictive, Introduction à une justice "simplement" virtuelle,' [2018] Arch. phil. droit no. 60, 3.

[2] Loi n° 2016-1321 du 7 octobre 2016, JORF 8 October 2016.

[3] J.-M. Brigant, 'Les risques accentués d'une justice pénale predictive', [2018] Arch. phil. Droit no. 60, 46.

[4] M. Delmas-Marty, 'Vers une justice pénale prédictive', in Mélanges en l'honneur de Geneviève Giudi-celli-Delage, Humanisme et Justice (Dalloz 2017), 58.

recidivism in order to prevent crime. It also implies a paradigm shift, with predictive justice being the opposite of retributive justice.

But that is not the meaning reflected in discussions about predictive justice in France today. Even when the term is applied to criminal justice, it refers more to predicting the results of proceedings based on previous results in similar cases rather than predicting future events. Some authors, therefore, prefer to use the term 'foreseeable justice'[5] or 'algorithm-based prediction of legal outcomes'.[6] The French Human Rights Advisory Commission (Commission Nationale Consultative des Droits de l'Homme, or CNCDH) even advises using the generic phrase 'algorithmic decision-support systems' when referring to AI systems.[7]

Despite the overall lack of legal research in this area in France,[8] legal commentators developed a taste for this subject after the Digital Republic Act was passed and several new definitions have since been posited. However, the fact that the term 'predictive justice' made its way into the 2018–2019 Lexique des termes juridiques [Glossary of Legal Terms] by Serge Guinchard and Thierry Debard[9] did not harmonize those definitions or the terminology, which different authors use in markedly different ways when explaining what predictive justice is and what it is used for.

1.1.1 What is predictive justice?

Most authors define predictive justice as a tool or a set of tools. Some liken it to 'tools that analyze case law and the parties' writings,'[10] and others to 'a computer tool'[11] that works with a database of case law, sorting algorithms, or even neuronal networks. In keeping with the idea that predictive justice is a tool, the taskforce on making judicial decisions available to the public, presided by Loïc Cadiet, defined it as 'a set of tools, developed by analyzing large volumes of data collected through the legal system,'[12] that use probabilities. Other authors define predictive justice as a method: 'a method for resolving legal disputes that relies on algorithmic processing of masses of data collected from case law'.[13] Lastly, Anaïs Coletta uses the phrase 'prediction of legal outcomes by

[5] J.-M. Brigant, 'Les risques accentués d'une justice pénale predictive', n. 3 above, 47. Foreseeability is a requirement of Article 7 of the European Convention on Human Rights. See also A. Coletta, La prédiction judiciaire par les algorithmes, [2022] Dissertation under the supervision of G. Cerqueira, Université de Nîmes (France); T. Cassuto, 'La justice à l'épreuve de sa prédictibilité', [2017] AJ Pén., 334.

[6] A. Coletta, 'La prédiction judiciaire par les algorithmes', n. 5 above, paras. 4 to 8.

[7] CNCDH, Avis relatif à l'impact de l'intelligence artificielle sur les droits fondamentaux, [2022] A-2022-6, 4.

[8] Ibid., para. 11 and 12.

[9] S. Guinchard, Th. Debard, Lexique des termes juridiques 2018-2019 (Dalloz 2018).

[10] B. Dondero, 'Justice prédictive: la fin de l'aléa judiciaire?', [2017] D., 532.

[11] R. Boucq, 'La justice prédictive en question,' [2017] Dalloz actualités, 14 June : <https://www.dalloz-actualite.fr/chronique/justice-predictive-en-question>accessed on 14 Nov. 2022.

[12] Mission d'étude et de préfiguration sur l'ouverture au public des décisions de justice, L'open data des décisions de justice (Report to the minister of Justice from the taskforce presided by L. Cadiet, 2017), 14.

[13] S. Guinchard, Th. Debard [2018] n 9 above.

algorithms' in her dissertation to refer to both the processing itself and the techniques used.[14] Regardless of how it is defined, 'predictive justice' refers to decisions based on algorithms, and primarily AI algorithms.

1.1.2 What is the purpose of predictive justice?

Predictive justice is also defined by the purposes it serves. For some authors, it has a very generic function: 'to predict what the case law will be in the future.'[15] For others, it serves the more precise purposes of predicting 'the outcome of litigation (to the extent possible),'[16] 'the chances of success of various legal arguments,[17] and 'how a court will rule in a given case.'[18] There is one function most authors agree is prohibited, however: decision-making may not be delegated to an AI system. But that is precisely the function described by the Guinchard and Debard definition: 'a method by which the courts and/or prosecutors resolve disputes.'[19]

Some commentators use the terms 'analytical justice'[20] or 'algorithmic justice' in order to emphasize the fact that algorithms, rather than judges or prosecutors, are doing the predicting: 'predictive justice, therefore, does not exist as such, only the predictive algorithmic tool exists and, therefore, the result(s) of the calculation.'[21] Still, others call it a decision support tool, or a 'statistical tool for quantifying the risks involved in a dispute,'[22] because it makes it possible to calculate a party's probability of success, the average amount of compensation usually awarded, and even, according to some sales pitches, to identify the most persuasive arguments. That would make 'predictive justice' a decision support tool that helps lawyers and their clients rather than judges.

From these attempts to define predictive justice, Marie-Cécile Lasserre has derived two categories: quantitative and cognitive. Quantitative predictive justice involves the use of tools that 'use data to deliver a legal response, but one that is devoid of independent human reasoning.'[23] For example, by using AI, one can determine probabilities and statistical trends, or put a figure on harm.[24] Cognitive predictive justice is still a subject of fiction, as it 'refers to machines that use a human reasoning process developed by AI to

---

[14] A. Coletta, [2022] n 5 above, para. 6.
[15] B. Dondero [2017] n 10 above.
[16] Mission d'étude et de préfiguration sur l'ouverture au public des décisions de justice, L'open data des décisions de justice (Report to the minister of Justice from the taskforce presided by L. Cadiet, 2017).
[17] R. Boucq (2017], n 11 above.
[18] A. Coletta, [2022] n 5 above, para. 7.
[19] S. Guinchard, Th. Debard [2018] n 9 above
[20] C. Guillard, 'La justice prédictive et l'IA dans le procès pénal: risques et opportunités,' [2020] OJP: <https://www.justicepenale.net/post/la-justice-prédictive-et-l-ia-dans-le-procès-pénal-risques-et-opportunités> accessed on 14 Nov. 2022.
[21] S. Lebreton-Derrien, [2018], n 1 above, 4.
[22] Ibid. 5.
[23] M.-C. Lasserre, 'L'intelligence artificielle au service du droit: la justice prédictive, la justice du futur?' [2017] LPA 30 June (130), 6.
[24] E. Rottier, 'Quelle prévisibilité pour la justice?' [2018] Arch. phil. Droit no.60, 189.

resolve legal disputes. Predictive justice tools would thus become 'robo-judges' that would replace human decision makers.'[25]

1.2 National practice with respect to using predictive justice tools in criminal cases

By providing that all judicial decisions will be available to the public by 2025 (decisions in criminal cases being the last to become available), the Digital Republic Act promotes the development of startups and other so-called 'legal tech' companies that offer tools to help analyze and use judicial decisions, but computer tools are not new to the legal world.

The long, tedious process of computerizing the criminal courts[26] led to the development, by local courts for their own use or by the Justice Ministry for national use, of software that helps judges write decisions and even helps judges make decisions. The superior courts (tribunaux de grande instance), now called judicial courts (tribunaux judiciaires), were using national criminal law applications such as Micro-pénale, Mini-pénale and EPWIN, or locally developed applications (INSTRU, WINSTRU, WINEURS),[27] all of which have gradually been replaced since 2009 by Cassiopée,[28] which processes data collected in the criminal justice system. Cassiopée 'makes it possible, in particular, (i) to manage hearings and write the courts' decisions and related documents, (ii) to manage appeals and requests for presidential pardons, petitions, evidence and other items in custody, sentence enforcement, schedules, the notice/reminder systems, and the document printing systems, (iv) to carry out electronic archiving, and (v) to conduct intra- and inter-court searches and read other courts' decisions'[29] (emphasis added). Because information is recorded from the moment a person enters the criminal justice system till the moment they leave it, Cassiopée enables judges and others authorized to use this software to do research, conduct statistical analyses, and make comparisons, which means that they, as well as their decisions, are better informed. The software was developed in

[25] M.-C. Lasserre, [2017] n 23 above.

[26] Audit Court (Cour des comptes), Améliorer le fonctionnement de la justice, Point d'étape du plan de transformation numérique du Ministère de la Justice (Communication to the French Senate Finance Committee, January 2022).

[27] French Senate, Rapport général sur la Justice (no. 74 by M. Roland of LUART, written in the name of the Finance Committee, 2004), esp. 130, L'informatique pénale.

[28] Chaîne Applicative Supportant le Système d'Information Opérationnel pour le Pénal et les Enfants [Application Chain Supporting the Operational Information System for Criminal and Children's Matters], automated processing of personal data, including the application called 'automated national office of legal proceedings and processing,' the purpose of which may be to use the data collected for statistical research purposes, French Crim Pro Code art. 48-1 and R. 15-33-66-4 and ff.

[29] CNIL, Decision no. 2009-170 of 26 March 2009 constituting an opinion on a proposed decree in the State Council (Conseil d'État, the highest administrative court] related to the automated national office of legal proceedings and processing called 'Cassiopée'.

late 2003 by Atos and began being installed in prosecutors' offices in 2008, then in investigation chambers and the superior courts. It was not installed in the courts of appeal until 2019.[30]

Processing of such data may be expanded as part of the Justice Ministry's digital transformation plan for 2018–2022.[31] Enhancing these information systems, in particular Cassiopée, or APPI[32] for criminal matters, will make it possible to refine the ministry's statistics on how criminal cases are handled.

France is nonetheless far behind other countries when it comes to computerized processing of justice system data,[33] and given the implementation problems currently encountered by applications such as Cassiopée, it is doubtful that predictive justice applications will be used in criminal matters any time soon[34] (the applications currently used are software solutions developed under Justice Ministry supervision).

A similar application, called DataJust,[35] processes personal data for the purpose of developing an algorithm related to personal injury awards (including those granted as the result of a civil case brought in connection with a criminal offense).[36] The decree that created DataJust authorized the Justice Ministry to carry out this type of automated processing for two years in order to conduct retrospective and prospective assessments of public policies concerning civil and administrative liability; develop guidelines for personal injury awards; promote out-of-court settlements by informing parties to disputes and helping them determine how much the victim may claim in damages; and provide information and documentation to judges ruling on personal injury claims. To accomplish these purposes, the algorithm inventoried the amount of damages claimed and offered by parties to previous disputes, the appraisals proposed during out-of-court settlement proceedings, and the amounts awarded victims for each type of injury, as well as

[30] G. Thierry, "2019: l'année Cassiopée,' [2019] Dalloz actualité, 23 January: <https://www.dalloz-actualite.fr/flash/2019-l-annee-cassiopee> accessed on 22 mars 2022.
[31] Audit Court, [2022] n 26 above.
[32] Application des Peines, Probation et Insertion [Enforcement of Penalties, Probation, and Réinsertion], automated processing of personal data, the purpose of which includes using data collected for statistical research purposes and which may be compared with Cassiopée.
[33] Audit Court, [2022] n 26 above ; French National Assembly, Rapport d'information sur les carences de l'exécution des peines et l'évaluation de l'application Cassiopée (no. 3177, presented by E. Blanc, 2011).
[34] French National Assembly, [2011] ibid.; G. Thierry, [2019] n 30 above; L. Le Devin, 'Chez les magistrats, Cassiopée frôle la nullité', Libération, (Paris, 10 November 2017) : <https://www.liberation.fr/france/2017/11/10/chez-les-magistrats-cassiopee-frole-la-nullite_1609375/> accessed on 14 March 2022; see below paras 1.3.1 et 1.3.2.
[35] Décret n° 2020-356 du 27 mars 2020 portant création d'un traitement automatisé de données à caractère personnel dénommé "DataJust,' JORF no. 77, 29 March 2020.
[36] Ibid. Article 2 provides that the necessary data will be extracted from decisions issued on appeal between 1 January 2017 and 31 December 2019 by the administrative courts and the civil chambers of the judicial courts, solely in litigation concerning personal injury claims.

the numerous types of data and informations listed in the decree instituting the algorithm.[37] Its content would be accessible to Justice Ministry employees who are assigned to the department in charge of IT developments within the ministry's general secretariat and individually appointed by the secretary general, and agents of the office of the law of obligations individually appointed by the director of civil affairs and the seal. Although the State Council (Conseil d'Etat, the highest administrative court) approved this automated processing,[38] the Justice Ministry did not extend it, apparently because it was too complex.[39]

Lastly, legal tool is currently being used to help enforce the French Vehicle Code: the Anti-Road Violence Act of 2003[40] authorized the agency that automatically processes vehicle code violations (Agence Nationale du Traitement Automatisé des Infractions) to develop automated radar surveillance and automatic delivery of fines.[41] In 2019, AI was used to develop a program called 'AI Flash', which was to be incorporated into the radar systems in 2020 to make the automated processing more reliable. For example, an image-recognition algorithm detects license plates that don't match the vehicle model recorded in the national register, so if a license plate has been stolen and used on another vehicle, the system will not issue a fine.[42]

Meanwhile private legal tech firms are developing software applications for lawyers and companies, but they tend not to be used much in criminal cases. Even though such use is not prohibited, most of the applications that can be used in criminal cases use French open-data case law, and there is very little such case law available.

These applications are essentially search engines, but may also include services such as automatic monitoring of changes in legislation, regulations, and case law, as well as legal analysis of documents. Legal analysis is performed by scanning the documents (e.g. the opposing party's writings) for references to laws, regulations, and case law and displaying them to the person submitting the search request. Some applications also calculate the probabilities of various outcomes and estimate the damages that may be awarded based on a statistical analysis of relevant case law (e.g. Predictice and Case Law Analytics).[43]

---

[37] Ibid. Article 2

[38] State Council, 30 December 2021, no. 440376.

[39] E. Marzolf, 'Le ministère de la Justice renonce à son algorithme DataJust,' Acteurs publics, 14 January 2022.

[40] Loi n° 2003-495 du 12 juin 2003 renforçant la lutte contre la violence routière.

[41] Art. 529-11, French Crim Pro Code.

[42] Application developed as part of the Entrepreneur of General Interest program: <https://eig.etalab.gouv.fr/defis/ia-flash/> accessed on 14 Nov. 2022; See also <https://www.permisapoints.fr/securite-routiere/intelligence-artificielle-venir-aide-radars-automatiques#:~:text=L'intelligence%20artificielle%20va%20venir,usurpations%20de%20plaques%20d'immatriculation> accessed on 14 Nov. 2022

[43] See the list of softwares and on-line solutions prepared by L. Tavitian, 'Justice prédictive où en est-on? (2016]: <https://www.village-justice.com/articles/Justice-predictive-est-jurimetrie,22683.html> accessed 12 March 2022.

These tools analyze numerous judicial decisions to build mathematical or statistical models that highlight the criteria judges relied on in making their decisions. Some commentators stress that 'because it simplifies reality,' a model is false by definition.[44] It is impossible, of course, to account for all the criteria in play in a judicial decision, some of which may be entirely unrelated to the law or the facts of the case.[45] The limits to these tools must therefore be made known.

For example, Supra Legem, an application that analyzes administrative case law, was said to use predictive algorithms. Developed in 2016, it used AI to analyze administrative court decisions and determine possible outcomes given the decision's subject matter and the type of plaintiff and defendant. It also provided statistics and graphics showing how each judge tended to rule in a particular type of dispute. The designer claimed that the application thus made it possible to know in advance how a judge would rule, and therefore helped increase impartiality.[46] The software could be used in 'criminal matters' broadly speaking,[47] in particular when prison administration decisions are subject to appeal, or in connection with complaints regarding indecent detention conditions. However, in 2019 access to the application's website was blocked and lawmakers passed a law[48,] that added provisions to the French Judicial Organization Code and Administrative Justice Code prohibiting the reuse of the identity data of judges and clerk's office employees, in particular for profiling or ranking of which the purpose or effect is to assess, analyze, compare, or predict their actual or supposed professional practices.[49]

The government has, however, encouraged greater use of AI, including in the judicial system. In 2016 the Department of Legal and Administrative Information created the 'DILA – open law – case law' award to reward innovation, in particular the development of applications, services, or products for visualizing legal data or that facilitate the reuse of such data[50] (Prédictice was the 2016 winner).[51] In addition, the Ministry of Transformation and Public Office's Public Interest Entrepreneur (Entrepreneur d'intérêt général)

[44] J. Levy-Véhel, 'L'office du juge : un éclairage via la modélisation mathématique,' [2020] Cahiers de la Justice, 4, 744.

[45] S. Danziger, J. Levav, & L. Avnaim-Pesso, 'Qu'a mangé le juge à son petit-déjeuner ? De l'impact des conditions de travail sur la décision de justice,' [2015] Les Cahiers de la Justice, 579.

[46] See the presentation of the application on the government website <https://www.data.gouv.fr/fr/reuses/supra-legem/>.

[47] Translation note: 'criminal matters' is used in the broad sense throughout this article to include proceedings related to prison administration and sentence enforcement as well as pre-conviction proceedings.

[48] Loi n° 2019-222 du 23 mars 2019 de programmation et de réforme pour la justice.

[49] Art. L. 10 of the French Administrative Justice Code and L. 111-13 of the Judicial Organization Code.

[50] Arrêté du 4 November 2016 relatif à la création et dotation du prix de la direction de l'information légale et administrative "DILA - le droit ouvert - jurisprudence", JORF no. 268, 8 November 2016.

[51] https://www.dila.premier-ministre.gouv.fr/actualites/toutes-les-actualites/open-case-law-2016-remise-des-prix-le-droit-ouvert-jurisprudence.

program led to the development of an AI tool to make the automated processing of vehicle code violations more reliable.[52]

In addition, even though criminal justice is not yet concerned with the applications they develop, legal tech firms are increasingly partnering with well-known legal or teaching and research institutions to give their applications better name recognition. For example, Case Law Analytics has partnered with the legal publisher Dalloz and dispenses training at École Nationale de la Magistrature, École Nationale du Barreau, and in some universities.[53] Similarly, Prédictice has offered to support innovative educational projects by making its platform available at no charge to interested students and university professors.[54]

[52] See n 42 above.

[53] <https://www.caselawanalytics.com/wp-content/uploads/2021/06/Catalogue-de-formations-2021-Case-Law-Analytics.pdf>.

[54] <https://blog.predictice.com/le-programme-predictice-pour-lenseignement-et-la-recherche>.

Principal AI Applications in Criminal Law

Name	Date created	Intended users	Features	Areas of the law covered	Sources and types of data	Technology used
Juri'Prédis (SAS Juri'Prédis)	2018	Students, law firms, legal departments, in-house lawyers, legal departments of local governments or in the banking/ insurance sector, notaries, bailiffs, and auditors	Case law searches based on a legal issue, additional applications for monitoring legal developments and analyzing case law and documents (Juri'détect) for lawyers	All	Decisions that use open data, case law from the French judicial courts (Court of Cassation, courts of appeal, first instance courts), administrative courts, and the Constitutional Council (Conseil Constitutionnel)	AI (machine learning)
Doctrine.Fr (Forseti SAS)	2016	Lawyers, companies	Search engine, monitoring of legal developments, legal analysis of documents (Analyzer)	All	French case law	AI (legal intelligence)
Judicial open-data search engine	2021	All users	Search engine	All	Court of Cassation case law	AI

Principal AI Applications in Criminal Law

Supra-legem (M. Benesty) No longer accessible	2016	Citizens	Search engine, predictive analysis to determine a lawsuit's chances of success, as well as each judge's stance and degree of impartiality	Administrative law (may include prison-related disputes)	Administrative case law	Predictive algorithms, machine learning, statistical calculations
Jurisdata Analytics (LexisNexis)	2016	Legal professionals	Search and analysis engine, decision support tool, search for comparable decisions to develop legal strategies, assess the amount of damages	All but criminal law to date	Indexed judicial decisions from the Jurisdata analysis	Active data-visualization and correlation analysis

1.3 How Predictive Justice is Perceived

1.3.1 Commentators' perceptions

Predictive justice only recently became a topic of research in France and little has been written about it, especially concerning its use with respect to criminal matters.[55] In general, however, French commentators tend to distrust algorithmic predictions of judicial decisions, especially those predictions concern criminal matters.

Predictive justice is primarily seen as a fantasy,[56] as 'simply' potential rather than real judicial decision-making,[57] but criminal law is currently not concerned by these technological developments in France.[58] Commentators nonetheless discuss very real events oc-curring elsewhere (primarily in the US). Some authors favor formal justice and see AI as providing an opportunity to make the law more foreseeable and make decisions more consistent.[59] Others believe that algorithms can make trials more efficient through auto-mation, disembodiment, and speed (and even predictability).[60] Favorable opinions such as these are rare, however, and rapidly overshadowed by those expressing doubt.

Various authors are skeptical of the idea that French legal rules can be systematized the way that Common Law rules can, given our code-based system. Although 'the law seems to be a "logical" system'[61] that makes it possible to foresee the legal consequences of one's

---

[55] For a list, see A. Coletta, [2022], n 5 above, para. 11. See also, esp. in criminal law, the dissertations being prepared since 2018 by Sarah Cherqaoui, L'intelligence artificielle en matière pénale, under the supervision of O. Decima, Bordeaux, and since 2019 by Emily Mongaillard, Étude de la contribution de l'intelligence artificielle à l'évolution du droit: l'exemple du droit pénal des affaires, under the supervision of C. Mascala, Toulouse. The Ecole Nationale de la Magistrature (ENM) funds academics research programs: E. Vergès, G. Vial, 'L'impact des algorithmes sur les décisions de justice des magistrats au pénal et au civil', (2022); the same researchers are interested in the practice of judges and evidential reasoning.
International or national congresses have also been able to address the subject in criminal matters: J.-B. Hubin, H. Jacquemin, B. Michaux (dir.), 'Le juge et l'algorithme: juges augmentés ou justice diminuée ?' (Larcier 2019) or lastly a on-line congress under the supervision of P. Mistretta and J. Alix, 'Intelligence artificielle et justice pénale', 12 march 2021 <https://lexradio.fr/emission/1-27-COLLOQUE-INTELLIGENCE-ARTIFICIELLE-ET-JUSTICE-PENALE-EN-LIGNE-LE-12-MARS-2021>.
See also essays written by academics researchers : S. Desmoulin-Canselier, D. Le Métayer, 'Décider avec les algorithmes, quelle place pour l'Homme, quelle place pour le droit ?', (Dalloz, 2020); F. G'Sell, 'Justice numérique, (Dalloz, 2021); or essays written by judges : E. Poinas, 'Le tribunal des algorithmes: juger à l'ère des nouvelles technologies', (Berger-Levrault, 2019).

[56] Dondero B., 'Justice prédictive: la fin de l'aléa judiciaire?' [2017], D., 532.

[57] S. Lebreton-Derrien, 'La justice prédictive, Introduction à une justice "simplement" virtuelle,' [2018] Arch. phil. droit no. 60, 21.

[58] J.-M. Brigant, 'Les risques accentués d'une justice pénale predictive', [2018] Arch. phil. Droit no. 60, 46.

[59] Guillaume Zambrano, 'Précédents et prédictions jurisprudentielles à l'ère des big data: parier sur le résultat (probable) d'un procès,' [2015] (hal-01496098).

[60] J-B Duclercq, 'Les algorithmes en procès,' [2018] RFDA, 131.

[61] Boucq R., 'La justice prédictive en question,' [2017], Dalloz actualités, 14 June : <https://www.dalloz-actualite.fr/chronique/justice-predictive-en-question>.

actions, there is still some uncertainty because the relationships the law governs are human, and factual situations cannot be reduced to the mere statement of an abstract, general rule.[62] As not all the concepts applied in a legal decision are defined by statute, judges have discretion to craft solutions to fit the political and social situations they are dealing with, which may not be covered by any code, and to decide what the 'rule for applying the [legal] rule' is.[63] It seems that a judgment therefore cannot be rendered without a judge who interprets and applies the legal rule.[64]

Some authors doubt that algorithms are reliable or of high enough quality because they are based on only a small portion of the information normally taken into account by a judge. That is especially true in criminal cases, since trial court decisions are still largely unavailable to the public and algorithms do not take the growing number of prosecution and sentencing alternatives into account.[65]

Some authors therefore emphasize that the algorithms are the result of arbitrary choices made by their designers, and are therefore necessarily limited and to be used cautiously: 'Applied to judicial decisions, the mathematical models have trouble accounting completely for the reality they claim to describe and can only incompletely serve predictive or actuarial purposes.'[66]. A judicial decision is much more complex than a simple syllogism, and algorithms, which can only make correlations or lexical connections, do not provide the kind of fine analysis a judge does. Applications developed on this basis therefore offer only distorted explanations of judicial decisions, 'a sort of nearsighted memory of justice, devoid of any close analysis of the factors underlying the judicial decisions it claims to render'[67] because a certain amount of unforeseeability remains. This criticism is precisely why some authors support the profiling of judges, believing it will ensure that judicial decisions are truly predictable and, in turn, that the application of law is foreseeable.[68]

The statistics obtained may also be distorted if the decisions used are not assigned a hierarchy, such as between the decisions issued by the Court of Cassation and the courts of appeal or the courts of first instance.

---

[62] V. Vigneau, 'Le passé ne manque pas d'avenir, Libres propos d'un juge sur la justice prédictive,' [2018] D., 1095.

[63] A. Garapon, 'Les enjeux de la justice prédictive,' [2017] JCP G no. 1-2, 9 January, doctr. 31.

[64] B. Dondero, 'Justice prédictive: la fin de l'aléa judiciaire?' [2017] D., 532; See Vigneau, [2018] n 62 above, citing Gérard Cornu: 'Judges are required by statute to create law, which will only become reality through judicial contributions' (G. Cornu, Cours de doctorat 1970-1971. L'apport des réformes récentes du code civil à la théorie du droit civil, p. 167.)

[65] J.-M. Brigant, [2018] n 58 above.

[66] Y. Meneceur, C. Barbaro, 'Intelligence artificielle et mémoire de la justice : le grand malentendu,' [2019] Les Cahiers de la Justice, 277.

[67] B. Dondero [2017], n 64 above.

[68] A. Coletta, 'La prédiction judiciaire par les algorithmes', [2022], Dissertation under the supervision of G. Cerqueira, University of Nîmes (France), paras. 20, 24.

Several fears have been expressed about predictive justice. First, Antoine Garapon highlights the danger of performativity, that is, that the algorithmically predicted outcome will become standard.[69] Discretion, and therefore judicial freedom, would be replaced by conformism induced by pressure to rule in line with the algorithmic standard. A judge who wants to deviate from the standard will have to provide extensive justification for their decision and may be subject to a new standard of liability. The homogenized case law obtained this way may, in turn, reproduce social stigmas and stop the law from evolving away from outdated solutions because the algorithms rely on prior case law.[70] Moreover, with respect to individualized punishment, AI could create conflict between humanist individualization of punishment (Saleilles, Ancel) and scientific individualization based on a statistical assessment of an individual's dangerousness (Italian Positivist School).

Several fears have also been raised with respect to the general principles of law, with some commentators claiming that arguments and decisions dictated by algorithms threaten the adversarial principle: 'Using algorithms in connection with a trial raises the issue of this practice's compliance with international treaties and the French Constitution, especially from the point of view of the right to a fair trial and the principle of judicial independence.'[71] Trials will be of lesser quality because judges and prosecutors will be less independent from other judges and prosecutors, from the parties, from any experts called to give an opinion, and from the machines. Quality will also be threatened by the reduced social acceptability of trials that will result if the parties are 'heard' by a machine rather than a judge.[72] And when the predicted outcome is unfavorable, the advice may be to avoid a trial.[73]

Some commentators averse to predictive justice would therefore refuse, in the name of judicial freedom and independence, to allow the Court of Cassation's case law to be used as the foundation for predictive algorithms because trial judges are not required to follow that court's non-precedential case law and their decisions are not dictated by rigid concern for consistency and foreseeability.[74]

Other commentators temper those fears, however, openly welcoming the possibility of another type of justice. Predictive justice cannot thrive without human intervention, which means there is room to be creative to protect ourselves from digital domination.[75] For example, a new procedural guideline could emerge: '. . . a principle of candor from the judge, who should make every effort to look at the parties with a fresh eye, devoid

[69] A. Garapon, [2017], n 63 above.
[70] B. Dondero, [2017] n 64 above
[71] J-B Duclercq, 'Les algorithmes en procès,' [2018] RFDA, 131.
[72] Ibid.
[73] S. Lebreton-Derrien, 'La justice prédictive, Introduction à une justice "simplement" virtuelle,' [2018] Arch. phil. droit no. 60, 13.
[74] V. Vigneau, [2018] n 62 above.
[75] S. Lebreton-Derrien, [2017] n 73 above., 12.

of all prejudice and free from predictive pressures.'[76] Or it could promote mediation, and therefore 'better acceptance of the outcome' because the parties will have participated 'in contractualizing the proceedings.'[77]

French legal commentators therefore suggest solutions that preserve the humanity inherent in legal proceedings. A first necessary step seems to be training for legal professionals so they understand the various aspects of the predictive justice tools and algorithms that have been developed. Ethical standards must also be developed so that those who use algorithms are liable for the results.[78]

Lastly, some commentators decry the 'capitalistic entrepreneurial attitude'[79] that underlies these developments, which are profoundly changing the profession of lawyer and pose a 'serious threat to current forms of law, legal professionals, and the courts.'[80] In addition, the report by the taskforce on making judicial decisions available to the public highlighted the possibility that the legal firm market will be affected, as French legal techs may run into competition from more powerful foreign firms.[81] It also foresaw changes in the work of legal professionals that would require a legal and ethical framework.[82]

1.3.2 Practitioners' perceptions

To date, judges rarely use predictive justice tools, and that use is limited to experimenting with software that processes case-law databases. To get feedback from judges on Predictice's an AI tool and analyze both its repercussions on how they reach their decisions and their perceptions of the activity of judging, a study was conducted in spring 2017 in connection with the testing of the Predictice software by the Rennes and Douai courts of appeal.[83] Fifteen pilot law firms also tested the software at that time.

Predictice is presented as a decision support platform based on data collected from court of appeal case law that has become open data. It analyzes the decisions in the database according to the criteria selected by the person submitting the search request and indicates the outcomes in similar cases and, where applicable, the amount of compensation that was awarded.

---

[76] A. Garapon and J. Lassègue, Justice digitale, (Paris PUF 2018), 259.
[77] S. Lebreton-Derrien, [2017] n 73 above, 14.
[78] Ibid., 17.
[79] Dondero B., 'Justice prédictive: la fin de l'aléa judiciaire?' [2017], D., 532.
[80] A. Garapon, « Les enjeux de la justice prédictive », [2017] JCP G n°1-2, 9 january, doctr. 31.
[81] Mission d'étude et de préfiguration sur l'ouverture au public des décisions de justice, L'open data des décisions de justice (Report to the minister of Justice from the taskforce presided by L. Cadiet, 2017), 28.
[82] Ibid., 30.
[83] C. Licoppe and L. Dumoulin. 'Le travail des juges et les algorithmes de traitement de la jurisprudence. Premières analyses d'une expérimentation de "justice prédictive" en France,' [2019] Droit et société, vol. 103, no. 3, 535.

It received a lukewarm-to-hostile welcome from the judges due to distrust of an application that threatened their independence, and due to technical problems caused by the fact that the courts' information systems are often obsolete. The experiment thus did not proceed as the developer had imagined—collectively in a setting of 'participative innovation,'[84]—as each participating judge continued to work alone with the software, which is common among judges in France. (Prosecutor's offices have developed more collaborative working methods in the past few years, but since criminal cases were not included in the experiment, no prosecutors took part.)

The lukewarm welcome may also be due to the preexistence of 'homemade' decision support tools. The 2017 study refers to computerized judgment templates for simple, repetitive cases, compensation tables for dismissal cases, and a now national, albeit criticized, reference table for personal injury awards.[85] In criminal matters, Cassiopée helps trial judges write decisions, for example by providing decision templates and preestablished lists of grounds that can be selected when writing a first instance decision.[86]

There thus seems to be tension between 'the independence demanded of judges (and their de facto independence in the sociology-of-labor sense) and the need for consistency among judicial decisions.'[87]. Judges have built their own tools, which they can control, but resist standardization and software such as predictive justice tools developed by private firms, which they cannot control.

Judges also seem to disagree on whether their tools, and in particular the compensation tables, should be shared with lawyers. Some fear they will be bound by the result returned by the tool or will give the impression that the decision has already been made. Others, however, think the tables merely give lawyers indications that enable them to state more realistic claims. In the end, the compensation table was made public in order to enable parties to settle disputes out of court and thus reduce the amount of litigation.[88]

The 2017 study also emphasized that some judges used Predictice to compare their decisions to the average. However, the ability not only for judges but also for lawyers to analyze a particular judge's decisions raises the fear that judgments will become standardized, that judges will be tempted to conform to the average, and therefore that judicial independence will suffer.

[84] Ibid.
[85] Ibid. and the citations therein.
[86] G. Thierry, "2019: l'année Cassiopée,' [2019] Dalloz actualité, 23 January: <https://www.dalloz-actualite.fr/flash/2019-l-annee-cassiopee> accessed on 22 March 2022.
[87] Ibid.
[88] Ibid.

1.4 Assessment of reliability, impartiality, equality, and adaptability

According to the information available, the judges who participated in the experiment described above were generally disappointed, expressing concerns about the application's reliability more than its impartiality. The experiment did not include criminal appeals.

Although they liked how up to date the tool was, the judges did not find that it outperformed the search engines they already had.[89] They also said the software should be improved and the analysis refined, because it sometimes produced aberrant results. For example, the calculation of dismissal compensation could be distorted by the fact that the software did not distinguish between managers and other employees.[90] After that initial experiment, the tool stopped being used.

Other studies have been conducted or are in progress, however. For example, the Predictice blog mentions studies being carried out at Université de Paris Dauphine on the application's performativity.[91]

In addition, as a result of the call for projects issued by the Mission de recherche droit et justice[92] for November 2020–February 2023, a study called 'Law and Artificial Intelligence: Can Market Regulation Produce Trustworthy Predictive Justice Tools?' is being conducted under the supervision of Agnès Delaborde, Aurore Hyde, Christian Licoppe. The study is examining how predictive justice tools are developed, the conditions under which the results they produce may become sources of law, the type of public or private regulation needed to govern their use, and the reliability and impartiality of the tools selected to illustrate the research.[93]

Due to the lack of testing to date, there is no relevant data on which to base an assessment of these tools' reliability or impartiality in criminal matters.

---

[89] Reply from the minister of Justice, JO Sénat 28/12/2017 p. 4694, to Written Question no. 01823 from Jérôme Durain, JO Sénat 02/11/2017 page 3392: <https://www.senat.fr/questions/base/2017/qSEQ171101823.html>.

[90] 'L'utilisation de l'outil Predictice déçoit la cour d'appel de Rennes, Interview de X. Ronsin, premier président de la cour d'appel de Rennes,' Dalloz Actualité, 16 October 2017 : <https://www.dalloz-actualite.fr/interview/l-utilisation-de-l-outil-predictice-decoit-cour-d-appel-de-rennes>. Cf. C. Licoppe and L. Dumoulin, 'Le travail des juges et les algorithmes de traitement de la jurisprudence. Premières analyses d'une expérimentation de "justice prédictive" en France,' [2019] Droit et société, vol. 103, no. 3, 535, minimizing the failure reported in the press.

[91] <https://blog.predictice.com/le-programme-predictice-pour-lenseignement-et-la-recherche-souffle-sa-première-bougie>. The rapporteur has not found a report on this test, however.

[92] Now called Institut des Études et de la Recherche sur le Droit et la Justice.

[93] <http://www.gip-recherche-justice.fr/publication/droit-et-intelligence-quelle-regulation-du-marche-pour-des-outils-de-justice-previsionnelle-dignes-de-confiance/>.

2 Legislation, regulations and soft law

France does not yet have a specific legal framework concerning the use of AI-based systems for predictive justice purposes. As indicated above, research is currently being done on whether such a framework is needed and what an appropriate framework would be.[94] Other legislative and regulatory provisions may apply to such systems, however, in particular with respect to the automated processing of personal data. The National Commission on Information Technology and Civil Liberties (Commission Nationale de l'Informatique et des Libertés, or CNIL) has indicated that many of the issues and questions raised by AI were raised in 1978 with respect to computerizing the government, and current legislation contains responses to them. The CNIL also encourages considering the possibility of regulation by the actors involved, along with government regulation. Ethical charters may also be a means of regulation.[95]

2.1 Applicable laws and regulations

2.1.1 The Information Technology and Civil Liberties Act

The Information Technology (IT) and Civil Liberties Act[96] sets forth the principles guiding algorithms' use of personal data: 'Information technology must serve all citizens. It must be developed in the context of international cooperation. It must not be detrimental to or infringe human identity, human rights, privacy, or civil liberties or individual rights' (Article 1).

It governs the creation and use of personal data processing systems, and in particular the conditions for data collection and retention, and assigns the CNIL the role of supervisor. However, it does not apply to processing that reuses the data from judicial decisions available as open data since the Digital Republic Act was passed:

- Article 44 5° of the IT and Civil Liberties Act excludes from the Act's scope 'processing pertaining to the reuse of public data appearing in the decisions mentioned in Article L. 10 of the Administrative Justice Code and Article L. 111-13 of the Judicial Organization Code, provided that neither the purpose nor the effect of such processing is the reidentification of data subjects . . .'; and
- Article 46 of the IT and Civil Liberties Act allows those who reuse the data appearing in open-data judicial decisions to process data related to criminal matters: 'Personal data related to criminal convictions, offenses, or related

[94] Ibid.
[95] CNIL, Comment permettre à l'Homme de garder la main? Les enjeux éthiques des algorithmes et de l'intelligence artificielle [2017] 44.
[96] Loi no. 78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés, JORF 7 January 1978.

> security measures may be processed only by . . . 5° Those who reuse the public data appearing in the decisions mentioned in Article L. 10 of the Administrative Justice Code and Article L. 111-13 of the Judicial Organization Code, provided that neither the purpose nor the effect of such processing is the reidentification of data subjects . . . .'

In addition, Article 42 3° excludes from the Act's scope personal data processing 'by the competent authorities for purposes of (i) preventing and detecting, investigating, or prosecuting criminal offenses, or (ii) carrying out criminal sentences, including to protect against threats to public safety and prevent such threats.' This provision thus seems to authorize judges and prosecutors to use predictive software that calculates the probability of reoffending. That authorization is limited by Article 47, however, which provides that 'no judicial decision involving the assessment of a person's behavior may be based on automated personal data processing designed to assess certain aspects of the data subject's personality.'

In other words, the IT and Civil Liberties Act requires a certain amount of human intervention when it comes to making decisions that have significant consequences for the people involved, and in particular judicial decisions. For example, it has always prohibited the use of profiling algorithms, particularly by the courts. Similarly, 'no decision that produces legal effects with regard to a person or that significantly affects that person may be made solely on the basis of automated personal data processing, including profiling,' subject to a few exceptions governed by the rest of Article 47.

These provisions also refer to the European Union (EU) regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (GDPR),[97] which applies directly (without implementing legislation) in EU member states and limits the use of certain data with predictive justice tools. For example, the processing of personal data related to criminal convictions and offenses is governed by Article 10 of the GDPR, which provides that such processing must be carried out 'under the control of official authority.' Similarly, the IT and Civil Liberties Act refers to GDPR Article 22, which governs profiling.

The Act also sets forth the procedures for data subjects to exercise their rights concerning their data and the right to be informed of how the algorithm works.[98]

As some authors have emphasized, 'the IT and Civil Liberties Act does not explicitly prohibit the courts from relying on profiling algorithms that process non-personal data' or 'from relying on other types of algorithms, which are legion.'[99]

---

[97] Regulation (EU) 2016/679 of the European Parliament and of the Council of April 27, 2016.

[98] Loi no. 78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés, JORF 7 January 197. art. 48 and following.

[99] J.-B. Duclercq, 'Les algorithmes en procès,' [2018] RFDA 131.

2.1.2 *Judicial Organization Code (COJ) and Administrative Justice Code (CJA)*

The Judicial Organization Code and the Administrative Justice Code govern the judicial and administrative court systems. The judicial courts include those that have jurisdiction over criminal cases, and the administrative courts may be called on to resolve disputes that involve application of criminal law, such as disputes with the prison administration regarding enforcement of a sentence or challenging the legality of a regulation adopted pursuant to a criminal statute.

Articles L 111-13 COJ and 10 CJA set forth the terms for making judicial and administrative court decisions available to the public, provided they have been anonymized. Those provisions also prohibit any 'reuse of the identity data of the judges and members of the clerks' offices that has the purpose or the effect of assessing, analyzing, comparing or predicting their actual or supposed professional practices. Violation of that prohibition is punishable by the penalties set forth in articles 226-18, 226-24 and 226-31 of the Criminal Code,[100] without prejudice to the measures and penalties provided for by the IT and Civil Liberties Act. This provision therefore prohibits the development of applications such as SupraLegem.[101]

2.1.3 *Code of relations between the public and the government (CRPA)*

As its name indicates, the CRPA governs relations between the public and the government, in particular as regards administrative documents, the communication of information, and access to personal information. Because it governs administrative proceedings other than litigation, it may also address criminal matters when a government agency is responsible for applying criminal laws and regulations.

CRPA articles L 321-1 and following set forth the rules for reusing public information, and such information is subject to the IT and Civil Liberties Act.

More specifically, CRPA Article L. 311-3-1 sets forth the rules applicable to the use of algorithmic processing, stating that any decision concerning an individual that is made based on algorithmic processing must include an explicit indication informing the individual of that fact. That indication must contain, as required by Article R 311-3-1-1, the purpose of the processing and an explanation of how to refer the matter to the commission for access to administrative documents, if necessary. The individual concerned by the decision may then ask the government agency in question to provide them with the rules governing such processing and the 'principal characteristics of its implementation': the extent to which and how the algorithmic processing contributed to decision making, the data that was processed and their sources, the processing parameters applied to the person's situation and, where applicable, the weight assigned them, and any operations

[100] Misdemeanor punishable by a custodial sentence of 5 years and a fine of €300,000.
[101] See para 1.2 above.

carried out by the processing. Article R 311-3-1-2 requires that such information be understandable by the recipient.

2.2 Non-Binding Sources

After studying the technology developed by legal techs, on October 9, 2020, the French Bar Council (Conseil National des Barreaux, or CNB) adopted a Charter on Transparency and Ethical Use of Judicial System Data.[102] Case Law Analytics and Doctrine were the first legal techs to sign it (October and December 2020).

According to the preamble, the Charter contains 'a set of principles designed to guarantee that actors will self-regulate with respect to both the algorithms used to exploit the judicial decision databases and the reuse of the data those databases contain.' It targets legal tech actors and encourages them to propose tools that comply with its principles.

By committing to uphold principles of 'doing good' and 'doing no harm,' the designers of predictive justice algorithms commit to protecting fundamental rights and freedoms rather than seeking performance. The fairness principle requires them to publicize any conflicts of interest they may have, while the principle of explainability requires them to provide clear information about how their algorithms work, what their nature and function is, their decision-making logic, and their purpose. The transparency principle requires that users be fully informed of any biases that were detected when the tool was used. The expertise and equality principle allows for including legal professionals on design teams and ensuring that teams are diverse and have equal numbers of men and women, to avoid reproducing biases. The protection principle aims to guarantee that everyone has equal access to the technology, and is furthered by the accessibility principle, which aims to ensure that technology is inclusive: general with options that can be activated to adapt to a target population. The designer accountability principle ensures that designers cannot avoid liability. The foreseeability and assessment principle aims to assess the tool's actions, measure the effects, and prevent risks. The mitigation of damages, remediation, and set-off principle is triggered in the event a defect results in negative effects for the user. The principle of technological neutrality and security aims to prevent performativity and protect the data.

These principles aim, in particular, to guarantee that users are fully informed about the algorithms that are used, to avoid reproducing biases by making sure that algorithm design teams are appropriately diverse, and to institute regular software assessments.

Even though it is non-binding, the European Ethical Charter on the use of AI in judicial systems adopted by the European Commission for the Efficiency of Justice (CEPEJ) of the Council of Europe (COE) in December 2018 may have a strong influence. Its five ethical principles are incorporated into the CNB's charter: respect for fundamental rights;

[102] Attached to the report of the legal tech working group, 'Legaltechs du domaine de la jurimétrie, préconisations d'actions,' 9 October 2020.

non-discrimination; quality and security; transparency, impartiality and fairness; and 'under user control.'

2.3 Case Law

To the rapporteur's knowledge, the French criminal courts have not yet been confronted with AI-based systems used for predictive justice purposes, as the existing systems are not used in criminal matters.

However, the State Council has confirmed that the decree instituting the DataJust automated processing[103] complies with the French Constitution, the European Convention on Human Rights and Fundamental Freedoms, the Charter of Fundamental Rights of the European Union, the GDPR, and the IT and Civil Liberties Act. In its decision of December 30, 2021, it found that personal data protection rights are adequately protected by DataJust, the purpose of which is to develop an algorithm that will analyze the compensation awarded for personal injuries by the administrative and judicial courts, and which is merely experimental and not intended to be made available to judges or parties to disputes. The State Council's flexible interpretation, in particular as to data minimization may be explained by the fact that it is difficult to precisely determine which data are strictly necessary since the purpose of the processing is to develop an algorithm.[104] The CNIL had also issued a favorable opinion concerning the creation of DataJust,[105] but the experiment nonetheless ended early, in January 2022, because implementation was deemed too complex.[106]

In addition, the CNIL has authorized the Justice Ministry to implement personal data processing for the purpose of gaining statistical knowledge of the penal response to racist offences by analyzing judgments in which at least one offense was perpetrated because of the victim's actual or supposed origins, nationality, religion, or race[107] and compiling statistics in that regard. Access to the data collected for these purposes is restricted and secure. This is therefore not a tool developed for predictive justice purposes, but rather to assess (and perhaps adjust) criminal justice policy regarding a category of offenses. However, the database constituted for this purpose would be the same for a predictive justice application.

---

[103] Décret n° 2020-356 du 27 mars 2020 portant création d'un traitement automatisé de données à caractère personnel dénommé "DataJust,' JORF no. 77, 29 March 2020.

[104] N. Belkacem, 'Secteur public, affaires régaliennes et intelligence artificielle - décisions de justice et développement d'un algorithme,' [2022] Communication Commerce électronique no. 2, February, comm. 14, noting the State Council's necessarily flexible interpretation.

[105] Decision no. 2020-002 of January 9, 2020 constituting an opinion on a proposed decree in the State Council creating automated personal data processing called 'DataJust' (request for an opinion no. 19020148, JORF no. 77, March 29, 2020).

[106] See below para 1.2; see also. CNCDH, Avis relatif à l'impact de l'intelligence artificielle sur les droits fondamentaux, [2022] A-2022-6, para. 30; E. Marzolf, 'Le ministère de la Justice renonce à son algorithme DataJust,' Acteurs publics, 14 January 2022.

[107] Decision no. 2017-186 of June 15, 2017.

Lastly, note that no CNIL authorization is required for algorithmic processing developed for predictive justice purposes to the extent such processing does not use personal data, but instead reuses anonymized data that has already been made public.

2.4 Substantive guarantees

The reliability, impartiality, equality, and adaptability of predictive justice tools are not guaranteed by any specific legislative framework. In its report, the taskforce on making judicial decisions available to the public recommends 'preventing, by an appropriate legal framework, the constitution of databases of judicial decisions that do not comply with the requirements, constraints, and guarantees recommended in this report.'[108] These principles are, however, guaranteed by the COE's and the CNB's non-binding ethical charters discussed above,[109] which may serve to regulate the use of predictive justice tools, and even provide a basis, if necessary, for a regulatory framework designed to protect these principles.

Transparent functioning and the use of algorithms was also one of the taskforce's concerns. In its report, it recommends 'regulating the use of new so-called 'predictive' justice tools by establishing an obligation of transparency for algorithms, implementation of flexible supervisory mechanisms by the government, and the institution of quality certification by an independent organization.'[110]

A principle of transparency with regard to the public, which may make it possible to detect any lack of reliability, impartiality, or equality, is therefore set out in articles L. 311-3-1 and R. 311-3-1-1 -2 of the CRPA.[111]

If an AI-based system used for predictive justice purposes does not use personal data, French law does not require authorization for it to be sold, 'provided neither the purpose nor the effect of the processing is to enable the reidentification of data subjects.'[112] Similarly, the designers of such tools are under no technical or technological obligation as regards the design method or the need to partner with legal professionals. Only the CNB's ethical charter recommends such a partnership in connection with the expertise principle.[113] Note, however, that the French companies that sell predictive justice tools count legal professionals, and lawyers in particular, among their partners or employees.

---

[108] Mission d'étude et de préfiguration sur l'ouverture au public des décisions de justice, L'open data des décisions de justice (Report to the minister of Justice from the taskforce presided by L. Cadiet, 2017), Recommendation no. 19.

[109] See para 2.2 above.

[110] Mission d'étude et de préfiguration sur l'ouverture au public des décisions de justice, n 108 above, Recommendation no. 20.

[111] See para 2.1.3 above.

[112] Art. 44 and 46 of Loi no. 78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés, see para 2.1 above.

[113] See. para 2.2 above.

Furthermore, although the taskforce recommended certification,[114] France has not yet proposed or required any kind of certification or labels for predictive justice tools.

And as for algorithm designer liability or user training, only the CNB's ethical charter provides these substantive guarantees, which are therefore not binding.

3 General legal principles

Until somewhat recently, discussions of general legal principles, and more specifically criminal law principles, were left primarily to legal commentators, but a few bodies have started addressing that issue. For example, in its report *Comment permettre à l'homme de garder la main* [How to Keep Humans in Charge], the CNIL addresses certain 'ethical' issues involving general legal principles,[115] and the Rights Defender has published a study on AI-related discrimination.[116] Meanwhile, the CNCDH issued an opinion on the impact of AI on fundamental rights,[117] in which it suggests changes to the proposed European regulation on AI.[118] Among other things, it notes that to be considered legitimate, any limitation of civil liberty by an AI system must be 'appropriate, necessary, and proportionate.'[119] It therefore recommends assessing the impact an AI-based application will have on fundamental rights before it is put on the market, in consultation with the stakeholders. The opinion sets forth in detail how such a study should be carried out and what it should contain, including identification of the fundamental rights that may be affected

[114] See esp. A. Louvaris, 'La justice prédictive entre être et devoir-être' in La justice prédictive (Paris, Dalloz thèmes et commentaires 2018) 36, and the report by the Mission d'étude et de préfiguration sur l'ouverture au public des décisions de justice, n 108 above, Recommendation no. 20.

[115] CNIL, Comment permettre à l'homme de garder la main, Les enjeux éthiques des algorithmes et e l'intelligence artificielle, Summary of public discussions [2017].

[116] Rights Defender, Algorithmes, prévenir l'automatisation des discriminations, [2020]. See also Technologies biométriques: l'impératif respect des droits fondamentaux [2021], addressing the issue of facial recognition, and, with Equinet (European Network of Equality Bodies), Pour une IA européenne protectrice et garante du principe de non-discrimination [2021], setting forth recommendations and fundamental principles for the future European legislation on artificial intelligence.

[117] CNCDH, Avis relatif à l'impact de l'intelligence artificielle sur les droits fondamentaux, n 106 above.

[118] Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts, com/2021/206 final.

[119] CNCDH [2022], n 106 above para. 25. While the European Court of Human Rights checks whether a restriction on civil liberties is provided for by statute, French law imposes this already via Article 4 of the Declaration of the Rights of Man and of the Citizen of 1789, which has constitutional value. The Constitutional Council and the State Council therefore review the proportionality of the restriction in light of the three-pronged test borrowed from German law. See CC, February 21, 2008, Loi relative à la rétention de sûreté et à la déclaration d'irresponsabilité pénale pour cause de trouble mental, no. 2008-562 pt. 13; M. Guyomar, 'Le passeport biométrique au contrôle: empreintes et cliché,' [2012] Actualité Juridique Droit Administratif, 35; J.-M. Sauvé, 'Le principe de proportionnalité, protecteur des libertés,' Institut Portalis, [2017], Aix-en-Provence (https://www.conseil-etat.fr/publications-colloques/discours-et-interventions/le-principe-de-proportionnalite-protecteur-des-libertes).

and the measures to be taken to mitigate the expected risks. The CNCDH also encourages continued monitoring for so long as these applications are used.[120]

3.1. Equality and the Fight Against Discrimination

The CNIL draws attention to the risk of bias and discrimination that may be inherent in an algorithm's design, citing the COMPAS (Correctional Offender Management Profile for Alternative Sanctions) application as an example. COMPAS produces a score representing the risk that an offender will reoffend, but the results are racially biased.[121]. The CNIL suggests broadening the fairness principle posited by the State Council in 2014 as follows:[122] 'a fair algorithm should not result in eliciting, reproducing, or reinforcing any discrimination whatsoever, even without the knowledge of the algorithm's designers.'[123] In addition, all normative provisions prohibiting all forms of discrimination apply, based on the constitutional principle of equality before the criminal law protected by Article 6 of the Declaration of the Rights of Man and the Citizen of 1789.[124]

Similarly, the Rights Defender recommends close monitoring to detect and penalties to punish discriminatory decisions resulting from algorithmic processing.[125] However, it calls attention to the ineffectiveness of existing protections given the systems' lack of transparency and the fact that their biases are often invisible.

3.2. Right to a Fair Trial

The right to a fair trial is guaranteed by Article 6(1) of the European Convention on Human Rights (ECHR) as a right to access to justice: 'Everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal established by law.' That protection is supplemented by the case law of the European Court of Human Rights, which in particular has established the principle of 'adversarial' proceedings (in which both parties are present or represented), the right to an effective appeal, and the right to equality of arms.

The risk that predictive justice (decision support) tools may infringe on these fundamental rights is highlighted by legal commentators[126] as well as independent administrative

---

[120] CNCDH, n 106 above, esp. Recommendations no. 9, 10, 11.

[121] CNIL, Comment permettre à l'homme de garder la main, n 115 above, 32.

[122] State Council, Le Numérique et les droits fondamentaux, [2014], 273, 278, and 281.

[123] CNIL, Comment permettre à l'homme de garder la main, n 115 above, 49.

[124] Esp. Article 225-1 of the Penal Code, which prohibits all discrimination between individuals based on 'their origins, sex, family situation, pregnancy, physical appearance, particular vulnerability resulting from their economic situation that is apparent or known to the perpetrator, their last name, place of residence, state of health, state of dependency, disability, genetic characteristics, morals, sexual orientation, gender identity, age, political opinion, union activities, ability to express themselves in a language other than French, their actual or supposed membership or non-membership in a particular ethnic group, Nation, alleged race, or religion.'

[125] Rights Defender, Algorithmes: prévenir l'automatisation des discriminations, [2020].

[126] S. Amrani Mekki, 'Le point de vue d'une universitaire,' in La justice prédictive (Ordre des avocats au Conseil d'État and Cour de cassation, eds.), Paris, Dalloz, Thèmes et commentaires, 2018, 49; See also S.-

agencies.[127] Because they provide statistics or probabilities on trial outcomes, predictive justice tools may lead to out-of-court dispute resolution.[128] One author emphasizes the limits of algorithms that may lead a person to waive their right to a trial based on biased statistics, whereas the waiver of a right protected by the ECHR must be the subject of fully informed and freely given consent.[129]

Similarly, the lack of algorithm neutrality threatens the principles of judicial independence and impartiality. Commentators[130] as well as practitioners[131] have underscored the danger of performativity posed by predictive software tools, namely that judges may simply follow the majority and thus increase that majority.[132] Despite the recommendation to establish a principle of algorithm neutrality,[133] this risk of prejudice harms judicial independence and impartiality and is not easily solved by recusal, or even by referral to another court based on legitimate suspicions if the judges in the other court have access to the same predictive justice tools.[134]

The CNCDH also doubts that judges will remain impartial, as they will be tempted to almost systematically reproduce the result reached by the algorithm-based software, 'given their current workload.'[135]

Commentators emphasize that an algorithm can hardly be considered a court within the meaning of Article 6(1) ECHR, at least as far as its ability to provide all the guarantees associated with that concept is concerned. That observation, coupled with the fact that a syllogistic algorithm does not reflect the reality and complexity of a judicial decision, raises doubts about the compliance of predictive justice tools with fair trial rights.[136]

M. Ferrié, 'Les algorithmes à l'épreuve du droit au procès équitable,' [2018] La Semaine Juridique Edition Générale no. 11, 12 March, doctr. 29.

[127] See esp. CNCDH, Avis relatif à l'impact de l'intelligence artificielle sur les droits fondamentaux, n 106 above.

[128] S. Amrani Mekki, 'Le point de vue d'une universitaire,' n 126 above, 58.

[129] S.-M. Ferrié, 'Les algorithmes à l'épreuve du droit au procès équitable,' n 126 above.

[130] See para 1.3.1 above, and A. Garapon, 'Les enjeux de la justice prédictive,' [2017] JCP G no. 1-2, 9 January, doctr. 31.

[131] See esp. J.-M. Sauvé, presentation at 'La justice predictive,' a colloquium held in connection with the bicentenary of the Order of lawyers practicing before the State Council and the Court of Cassation, 12 February 2018: <https://www.conseil-etat.fr/publications-colloques/discours-et-interventions/la-justice-predictive>accessed on 6 April 2022.

[132] S.-M. Ferrié, 'Les algorithmes à l'épreuve du droit au procès équitable,' n 126 above.

[133] Mission d'étude et de préfiguration sur l'ouverture au public des décisions de justice, L'open data des décisions de justice (Report to the minister of Justice from the taskforce presided by L. Cadiet, 2017), Recommendation no. 20.

[134] Articles 668 and 662 Crim. Pro. Code. See J.-M. Brigant, ''Les risques accentués d'une justice pénale predictive', [2018] Arch. phil. Droit no. 60, 57.

[135] CNCDH, Avis relatif à l'impact de l'intelligence artificielle sur les droits fondamentaux, [2022] n 106 above para.29.

[136] S.-M. Ferrié, 'Les algorithmes à l'épreuve du droit au procès équitable,' n 126 above.

3.3. Right to Access to a Human Judge and Right to Appeal an Algorithm-Based Decision

The CNIL highlights that even though the IT and Civil Liberties Act prohibits using automated personal data processing as the sole basis for a decision that has legal effects on individuals, decisions are increasingly automated today and the Act is being interpreted more loosely.[137] The CNIL, therefore, proposes that human intervention should not be required for each individual decision, which would cancel out the optimization gained by using an algorithm, but for groups of decisions: 'One could, for example, ensure that human, adversarial deliberation governs and supports the use of algorithms by examining and questioning the configuration as well as all of the system's effects, both direct and indirect. Such supervision could then be exercised from time to time to relatively numerous series of decisions rather than each individual decision.'[138]

The CNCDH, meanwhile, recommends that individuals who are the subject of decisions based wholly or partially on algorithmic processing be systematically informed of that fact. It also recommends that such individuals have a right to human review of any individual decision based entirely or even partially on algorithmic processing if the decision has significant consequences for the individual.[139]

The issue of appeals has not yet been the subject of specific discussions, however. The right to appeal a decision based on automated personal data processing provided for by the IT and Civil Liberties Act seems adequate in its current state. In criminal cases, for example, automated processing may concern vehicle code violations established by radar systems, which may be appealed to the police court.[140]

With respect to whether there is a second degree of jurisdiction when the appellate court uses the same predictive software as the trial court, the issue does not seem to have been raised, as this situation is still merely a potential, especially in criminal matters.

3.4. Constitutional Principles

Predictive justice tools also threaten constitutional principles other than equality.[141]

First, the principle of legality, established by Article 7 of the Declaration of the Rights of Man and the Citizen, is threatened by the advent of algorithms, which are based on judicial precedents more than on legal rules themselves.[142] The fact that algorithms are designed by private firms may also endanger the role lawmakers play in criminal law.

---

[137] CNIL, Comment permettre à l'homme de garder la main, Les enjeux éthiques des algorithmes et e l'intelligence artificielle, Summary of public discussions [2017], 52.
[138] Ibid.
[139] CNCDH, Avis relatif à l'impact de l'intelligence artificielle sur les droits fondamentaux, [2022] n 106 above, recommendations no. 17 and 18.
[140] Art. 529-11 Crim. Pro. Code.
[141] See para 3.1 above.
[142] J.-M. Brigant, 'Les risques accentués d'une justice pénale predictive', [2018] Arch. phil. Droit no. 60, 52.

With respect to the principle of strict interpretation of criminal statutes, a corollary of the legality principle, although it prohibits reasoning by analogy, it does not prevent the courts from reasoning teleologically, looking for a purpose that enables them to adapt the text to the social context of the offense. This human approach to law is lacking from algorithms and mathematical reasoning, such that if decisions based on them are reproduced systematically, case law can never be reversed.[143]

The same is true for the principle that punishment must be necessary (Article 8 of the Declaration of the Rights of Man and the Citizen). The corollary of this principle is the principle of individualized punishment set forth in Article 132-1 of the Penal Code: 'All sentences pronounced by a court must be individualized. To the extent allowed by law, courts decide on the nature, quantum, and regime of the penalties pronounced, according to the facts of the offense and the offender's personality and economic, family, and social situation, in accordance with the purposes and functions of punishment set forth in Article 130-1.' However, 'automatic calculation of the penalties that may be ordered for the offenses committed'[144] not only violates the principle of individualization, it also fails to satisfy the requirement to state the reasons for all penalties ordered today.[145]

Lastly, even though using actuarial tools to assess the risk of recidivism is being studied, in particular in light of other countries' experience with these tools, French commentators have not addressed the issue of their respect for the presumption of innocence.[146]

Selected literature

Amrani Mekki S, 'Le point de vue d'une universitaire,' in *La justice prédictive (Ordre des avocats au Conseil d'État and Cour de cassation*, eds.), [2018] Paris, Dalloz, Thèmes et commentaires

Audit Court (Cour des comptes), Améliorer le fonctionnement de la justice, Point d'étape du plan de transformation numérique du Ministère de la Justice (Communication to the French Senate Finance Committee, January 2022)

Barraud B, 'Un algorithme capable de prédire les décisions des juges : vers une robotisation de la justice ?', [2017] *Les Cahiers de la justice* no. 1, 121

[143] Ibid., p. 53, noting that algorithmic justice would probably not have given the employee immunity for stealing documents to their employer's detriment.
[144] Ibid. p. 54.
[145] Art. 132-19 and 132-20, French Penal Code with respect to misdemeanors and infractions; Art. 365-1 French Penal Code with respect to felonies.
[146] See eg P.-L. Déziel, 'L'utilisation de renseignements personnels dans le contexte de la justice prédictive : le cas des outils actuariels d'évaluation des risques de récidive,' [2018] Arch. Phil. Droit 60, 253.

Belkacem N, 'Secteur public, affaires régaliennes et intelligence artificielle - décisions de justice et développement d'un algorithme,' [2022] *Communication Commerce électronique* no. 2, February, comm. 14

Boucq R, 'La justice prédictive en question,' [2017] *Dalloz actualités*, 14 June: <https://www.dalloz-actualite.fr/chronique/justice-predictive-en-question> accessed on 14 Nov. 2022

Brigant J-M, 'Les risques accentués d'une justice pénale predictive', [2018] *Arch. phil. Droit* no. 60, 46

Cassuto T, 'La justice à l'épreuve de sa prédictibilité', [2017] *AJ Pén*., 334

CNCDH, Avis relatif à l'impact de l'intelligence artificielle sur les droits fondamentaux, (A-2022-6, 2022)

CNIL, Comment permettre à l Homme de garder la main ? Les enjeux éthiques des algorithmes et de l intelligence artificielle, [2017] 44

Coletta A, La prédiction judiciaire par les algorithmes, [2022] Dissertation under the supervision of G. Cerqueira, Université de Nîmes (France)

Danziger S, Levav J, & Avnaim-Pesso L, 'Qu'a mangé le juge à son petit-déjeuner ? De l'impact des conditions de travail sur la décision de justice,' [2015] *Les Cahiers de la Justice*, 579

Delmas-Marty M, 'Vers une justice pénale prédictive', in *Mélanges en l'honneur de Geneviève Giudicelli-Delage, Humanisme et Justice* (Dalloz 2017), 58

Le Devin L, 'Chez les magistrats, Cassiopée frôle la nullité', *Libération*, (Paris, 10 November 2017): <https://www.liberation.fr/france/2017/11/10/chez-les-magistrats-cassiopee-frole-la-nullite_1609375/> accessed on 14 March 2022

Desmoulin-Canselier S, Le Métayer D, *Décider avec les algorithmes, quelle place pour l'Homme, quelle place pour le droit ?*, (Dalloz, 2020)

Déziel P-L, 'L utilisation de renseignements personnels dans le contexte de la justice prédictive : le cas des outils actuariels d'évaluation des risques de récidive', [2018] *Arch. Phil. Droit* 60, 253

Dondero B, 'Justice prédictive : la fin de l aléa judiciaire ?', [2017] *D*., 532

Duclercq J-B, 'Les algorithmes en procès,' [2018] *RFDA*, 131

Faget J, 'La fabrique de la décision pénale. Une dialectique des asservissements et des émancipations', Champ pénal/Penal field, V 2008, p. 1-17

Ferrié S-M, 'Les algorithmes à l'épreuve du droit au procès équitable,' [2018] *La Semaine Juridique Edition Générale* no. 11, 12 March, doctr. 29

French Senate, Rapport général sur la Justice (no. 74 by M. Roland of LUART, written in the name of the Finance Committee, 2004), esp. 130, 'L'informatique pénale'

French National Assembly, Rapport d'information sur les carences de l'exécution des peines et l'évaluation de l'application Cassiopée (no. 3177, presented by E. Blanc, 2011)
State Council, Le Numérique et les droits fondamentaux, [2014]

Garapon A, 'Les enjeux de la justice prédictive,' [2017] *JCP G* no. 1-2, 9 January, doctr. 31

Garapon A and Lassègue J, *Justice digitale*, (Paris PUF 2018)

Guillard C, 'La justice prédictive et l'IA dans le procès pénal: risques et opportunités,' [2020] *OJP*: <https://www.justicepenale.net/post/la-justice-prédictive-et-l-ia-dans-le-procès-pénal-risques-et-opportunités> accessed on 14 Nov. 2022

Guyomar M, 'Le passeport biométrique au contrôle : empreintes et cliché,' [2012] *Actualité Juridique Droit Administratif*, 35

G'Sell F, *Justice numérique*, (Dalloz, 2021)

Hubin J-B, Jacquemin H, Michaux B (dir.), *Le juge et l'algorithme : juges augmentés ou justice diminuée ?* (Larcier 2019)

Lasserre M-C, 'L'intelligence artificielle au service du droit : la justice prédictive, la justice du futur ?' [2017] *LPA* 30 June (130), 6

Lebreton-Derrien S, 'La justice prédictive, Introduction à une justice "simplement" virtuelle,' [2018] *Arch. phil. droit* no. 60, 3

Levy-Véhel J, 'L'office du juge : un éclairage via la modélisation mathématique,' [2020] *Cahiers de la Justice*, 4, 744

Licoppe C and Dumoulin L, 'Le travail des juges et les algorithmes de traitement de la jurisprudence. Premières analyses d une expérimentation de "justice prédictive" en France,' [2019] *Droit et société*, vol. 103, no. 3, 535

Louvaris A, 'La justice prédictive entre être et devoir-être' in *La justice prédictive* (Paris, Dalloz thèmes et commentaires 2018)

Marzolf E, 'Le ministère de la Justice renonce à son algorithme DataJust,' *Acteurs publics*, 14 January 2022

Meneceur Y, Barbaro C, 'Intelligence artificielle et mémoire de la justice : le grand malentendu,' [2019] *Les Cahiers de la Justice*, 277

Mission d'étude et de préfiguration sur l ouverture au public des décisions de justice, L'open data des décisions de justice (Report to the minister of Justice from the taskforce presided by L. Cadiet, 2017)

Ordre des avocats au Conseil d'Etat et à la Cour de cassation (dir.), *La justice prédictive* (Dalloz, coll. Thèmes et commentaires, 2018)

Poinas E, *Le tribunal des algorithmes : juger à l'ère des nouvelles technologie*s (Berger-Levrault, 2019)

Report of the mission chaired by the MP and mathematician C. Villani, Donner un sens à l'intelligence artificielle (submitted to the Prime Minister on 8 mars 2018)

Rights Defender and CNIL, Algorithmes, prévenir l'automatisation des discriminations, (2020)

Rights Defender, Technologies biométriques : l'impératif respect des droits fondamentaux (2021)

Rights Defender and Equinet (European Network of Equality Bodies), Pour une IA européenne protectrice et garante du principe de non-discrimination (2021)

Rottier E, 'Quelle prévisibilité pour la justice ?' [2018] *Arch. phil. Droit* no.60, 189

Rotily C, Archambault L, 'Données biométriques issues d'expérimentations de reconnaissance faciale sur le territoire français : un défi à l'aune du droit 2.0 ?', (2020) *Dalloz IP/IT*, 54

Rouvière F, 'La justice prédictive, version moderne de la boule de cristal', (2017) *RTD civ*. 527

Tavitian L, 'Justice prédictive où en est-on ? (2016]: <https://www.village-justice.com/articles/Justice-predictive-est-jurimetrie,22683.html> accessed 12 March 2022

Thierry G, '2019: l'année Cassiopée,' [2019] *Dalloz actualité*, 23 January: <https://www.dalloz-actualite.fr/flash/2019-l-annee-cassiopee> accessed on 22 mars 2022

Vigneau V, 'Le passé ne manque pas d'avenir, Libres propos d'un juge sur la justice prédictive,' [2018] *D.*, 1095

Zambrano G, 'Précédents et prédictions jurisprudentielles à l'ère des big data : parier sur le résultat (probable) d un procès,' [2015] (hal-01496098)

PREDICTIVE JUSTICE IN ITALY

By Mitja Gialuz and Serena Quattrocolo

Abstract – The report builds upon the general questionnaire to outline the current state of affairs regarding the usage of AI solutions in criminal proceedings in Italy. With specific regard to quantitative legal prediction techniques, the report presents a situation of extremely limited application in the Italian legal system, which has not yet taken a clear position on the matter.

1 National practices

In the Italian legal system, there is no normative definition of 'predictive justice', nor is it under discussion in terms of drafts or proposals.

As a consequence, there is no software officially used at the moment for predictive purposes, nor are there trials or experiments openly set forth on behalf of the Ministry of Justice. It is likely that research groups, in the academic context, are testing and experimenting with computational models for the specific purpose of 'predictive justice', but the Ministry has not ordered or engaged in possible research of this kind.

This is not due to a specific and explicit decision about the reliability and desirability of 'quantitative legal prediction' methods[1] (as happened in France),[2] but rather the delay in the discussion of the matter. However, one important reference to recidivism risk assessment is necessary here.[3] In Italy, according to Art. 220 § 2 of the Code of Criminal Procedure, the psycho-criminological expertise on the defendant's character is allowed only after sentencing, in the correctional phase: it cannot be used either for adjudicating on guilt or for sentencing. Traditionally, a persisting mistrust in psychology is said to be the

[1] QLP, Quantitative Legal Prediction, is a computational approach. Based on data-driven AI, it implies the use of computational modelling to predict many different aspects of legal cases, or potential legal cases, moving from whole collections of existing data. According to the scholar who mainly promoted it 'QLP-based are designed to remedy or supplement the shortcomings of human reasoners' (Daniel M. Katz, 'Quantitative legal prediction or how I learned to stop worrying and start preparing for the data-driven future of the legal services industry', [2013], Emory Law Journal, 928). When referring to predictive justice, literature usually encompasses software to specifically predict the outcome of case or, in criminal law, the risk of recidivism.

[2] For a specific focus, Edouard Rottier, 'La justice prédictive et l'acte de juger: quelle prévisibilité pour la justice?', [2018] *Archives de Philosophie du Droit*, La justice predictive; Jean Marie Brigan, 'Les risques accentués d'une justice pénale predictive', [2018] *Archives de Philosophie du Droit*, La justice predictive, Dalloz; Pascale Deumier, La justice prédicitive et les sources du droit: la jurisprudence du fond, [2018] *Archives de Philosophie du Droit*, La justice predictive, Dalloz

[3] Georgia Zara, David P. Farrington, Criminal Recidivism: explanation, prediction and prevention (2016, Rutledge).

main rationale for the norm. However, it seems that the strongest reason for this is, rather, the reluctance to value character in the decision of the case, as the defendant's personal attitudes have no evidentiary value as to the facts of the file. Based on the assumption reported above, an Italian court could not make use of any risk-assessment tool, such as COMPAS or SAVRY, in the guilt and sentencing phase, given the prohibition of art. 220 § 2 ItCCP, as risk assessment reports should be considered and treated as psycho-criminological expert testimonies. Only the judge presiding over the correctional phase could rely on a risk-assessment tool, in order to decide, e.g. on parole or other prison benefits. Incorporated into a file, a digital risk assessment tool delivers reports that are the result of the application of a specific scientific theory, elaborated by scholars or clinicians, tested, verified and criticised by a community of peers. Although a digital tool does not imply the expert's personal presence in court, it delivers a result that is based on a scientific theory, according to the traditional paradigm of expert testimony: the adjudication on a matter of the case entails the application of technical or scientific knowledge, of which the judge is deprived.

Given this legal restriction, the psycho-criminological research on judicial risk assessment tools has been limited to the correctional area and the digital turn did not change the situation, due to the absence of a tradition in the field[4] and the unlikeliness of a change in the normative framework.

1.1 Quantitative legal analysis

As to the other component of predictive justice,[5] that is to say tools foreseeing judicial decision, the Italian institutional scheme does not encourage the research. Strongly rooted in the roman tradition, the jurisdiction is not based on stare decisis, and consistency – even in the Supreme Court's decision – is difficult to achieve, due to the wide range of means of appeal provided by the system, in opposition to the most important consequence of the rule of stare decisis, i.e. the binding commitment for the courts to respect the precedent, even if it appears wrong or unjust. A diffused organisation of the higher courts is a distinctive feature of many continental systems, such as e.g. in France, Italy and Germany.[6] Usually deprived of the power of certiorari, or to select the cases to be reviewed, these courts decide cases in their thousands each, per year, and deploy greater numbers of judges, organised in different sections or chambers, adjudicating independently from each other. This hampers the possibility of such courts to focus on their original mission, granting the uniformity of the law. For these reasons, quantitative legal prediction, based on retrieval of precedent decisions, may turn out to be less attractive than in common law countries.[7] By attractive, we mean less useful, due to the limited

[4] However, among others, see Georgia Zara, Valutare il rischio in ambito criminologico. Procedure e strumenti per l'assessment psicologico, (2016, Il Mulino), passim.

[5] Jordi Nieva Fenoll, *Inteligencia artificial y proceso judicial*, (2018, Marcial Pons).

[6] Michele Taruffo, Institutional Factors Influencing Precedents, in D.N. MacCormick, R.S. Summers, Interpreting Precedents (1997, Aldershot), 451 ff.

[7] Robert S. Summers, Precedent in the US, in D.N. MacCormick, R.S. Summers, Interpreting precedents, (1997, Aldershot), 358.

binding role of the precedent and also to the reduced reference that the Italian Supreme Court (and the whole continental tradition) owes to the merits of the fact, rather than to the principle of law. This has an impact on the capability of a predictive software to establish useful and accurate correlations. When referring to a precedent, either in common law or in civil law, the judge applies a form of analogy, comparing a certain number of variables, on the basis of which a precedent can prove relevant for the solution of the pending case.[8] Within a computational model, the number of factors, variables, involved in the decision-making process is crucial: it can be established ex ante, and once for all, e.g. in expert systems; or, in machine learning, it can be left to the training set, during which the system will be fed with a sufficient number of decisions, to allow it to recognise the relevant variables. The factors selected during the training set, will be applied in the predicting set. In criminal cases, the variables are many, both substantive and procedural (with specific regard to evidence admission and evaluation) and this strongly affects the possibility to release accurate 'predictive software'.[9] AI solutions are used in the most popular case-law data base, such as that of the Italian Supreme Court (Centro Elaborazione Dati Corte suprema di cassazione [www.italgiureweb.it]) and DeJure, by the publisher Giuffré, which are, however, traditional – although sophisticated – data bases retrieved with different sets of keywords. Quantitative legal prediction tools are unavailable in criminal justice both to the judiciary and the public, so there are no trends of alternative dispute resolution in this context.

The attention of both scientific literature and media to quantitative legal prediction, or predictive justice is scarce. As to the literature, the topic has recently (in the last two years) started to be treated as a matter of academic research. As to the media, they tend to ignore this aspect, given the absence of a practical application, while they tend to highlight any possible achievement in predictive policing.

1.2 Assessment of reliability, impartiality, equality, adaptability

Given the framework above, the answer to this sub-section is generally negative. At the time being, there is a huge ongoing reform of criminal justice. The process started in Spring 2021, in connection with the Next Generation EU fund. The Government's engagement in a profound reform of the Italian justice system has been crucial in the negotiations. In particular, with regard to criminal justice, one main endeavour is the great reduction in the length of proceedings, especially appeals. The Ministry of Justice appointed an expert committee to draft a 'delegation law' (piece of secondary law, passed by the Parliament, delegating the Government to implement the principles established by the law itself), and a second committee, in charge of drafting the implementation. Both rapporteurs were appointed in both committees and are now working on the implementation draft.

---

[8] François Ost, (Re)Learning to Think About Law from Cases, in S. Glanert (ed.), *Comparative legal reasoning. Essays in Honour of Geoffrey Samuels* (2018, Wildy, Simmond&Hill Publishing), 67 ff.

[9] Danièle Bourcier, 'L'acte de juger est-il modélisable ? De la logique à la justice', 2011] *Cahiers de Philosophie du Droit*, L'E-justice, Dalloz.

One point in the implementation draft is 'digitalisation of criminal justice'. However, in the principles established by the delegation law, there is no specific reference to predictive justice, rather to a more basic profile of digitalisation. This means, in the first instance, transforming the file into a digital one, something that has not happened yet, in Italy.[10] It will imply using digital resources for summoning parts and depositing acts in the file, at any stage of the proceedings, transforming notifications and deposits into a system of online, certified, exchanges. Although other countries may have experienced such a transition years ago, this will be a huge challenge for the Italian system. Firstly, such a transition depends on effective and reliable digital infrastructures, allowing both the judiciary and private parties to access and take advantage of the digital systems. Italy did not invest enough in digital infrastructures in the past decades and does not deploy the same effective resources in every part of the country: there are areas that are not covered by reliable, fast and regular digital services 24/7. Secondly, a digital file implies secure IT systems, guaranteeing both authenticity and privacy of communications: adequate subsidiaries must be provided for, in case of the misfunctioning of IT systems, so as not to jeopardise the chain of acts of the proceedings. This is the basic level of engagement at which the commission is working at the moment, with no attention to further, specific aspects of predictive justice. However, it is possible that, in the future, there will be an engagement of the Ministry of Justice in the endeavour of predictive justice. It is our opinion, based on the ongoing experience at the Ministry of Justice, that the topic will be studied and analysed at a governmental level. As the topic is gaining momentum in public-funded research calls, it is likely that, in the future, the Ministry of Justice will monitor the research process in order to understand the effective qualities of quantitative legal prediction (in terms of consistency, neutrality, equality, as mentioned by this section of the questionnaire) and the advantages, if any, that this may bring to the Italian justice system. In these terms, a full public engagement in the discussion is desirable: not only publicly funded research but publicly monitored research, because of the dramatic impact of a digitalised judicial decision-making process on the basic features of society. Massive usage of an automated decision-making process in criminal matters may affect every aspect of the institutional structure of justice, from the independence of the judiciary to the concept of predictability and certainty of criminal law: for these reasons, there should be no room for private investments and research in this field, rather a publicly regulated and monitored system. In this sense, the limited experience in other fields of public administration, suggests the need for a transparent process of commitment, development, and application of AI solutions in public administration, in order to fully realise the purposes of a good and fair public administration, which is enshrined in art. 97 of the Italian Constitution.

[10] Benedetta Galgani, Forme e garanzie nel prisma dell'innovazione tecnologica, (2022, Wolters Kluwer), 115 ff.

2 Normative framework

2.1 Law and soft law

As it has already been said, at the moment, in Italy, there is no legislation or normative instrument produced by executive authorities on predictive justice, nor is its introduction really under discussion, even in the context of the current efforts towards digitalisation.

However, from a general perspective, legal rules concerning right to privacy and data protection are relevant in this field, due to the fact that the systems at issue involve a massive treatment of personal information. For this reason, Legislative Decree n. 51 of 2018, implementing EU Directive 2016/680, on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, comes into consideration.[11]

As far as the legal principles established by these sources are concerned, it is possible to recall what has been observed with reference to predictive policing. Nevertheless, in this context, the ban on decisions based solely on automatic processing – already mentioned in the first section – is particularly important. As said above, Article 8 of Legislative Decree n. 51 of 2018, implementing Article 11 of the Directive, prohibits this kind of completely automated decisions, including profiling, which produces an adverse legal effect concerning the data subject, unless authorised by European Union or Member State law to which the controller is subject and which provides for appropriate safeguards for the rights and freedoms of the data subject. According to a strict way of interpreting the norm, it requires not only a human contribution to the decision, intended as an effective and significant control on the machine output, but also the imposition, to a human judge, of the evaluation of further evidence, different from the machine output, as a basis of any kind of decision.[12.]

The 'user under control' principle and the need for a critical approach towards the algorithm output, at the European level, is also expressed by the European Commission for the Efficiency of Justice (CEPEJ)'s European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment, as well as by Article 14 of the European Commission's proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on AI.

Moreover, Article 8 of the mentioned Legislative decree prohibits automated decisions based on the sensitive personal data listed in Article 9 of GDPR, unless specific measures

[11] Federica Resta, 'La direttiva sulla protezione dei dati personali in ambito giudiziario penale e di polizia e la tutela dei terzi' [2020], www.giustiziainsieme.it

[12] Mitja Gialuz, 'Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa', [2019] *Diritto penale contemporaneo* (29th May 2019), 16-18.

aimed at protecting the individual's rights and liberties are adopted. In any case, according to the same disposition, a discrimination of persons by the use of profiling based on this kind of information can never be admitted.

Finally, even if, to our knowledge, no specific soft law source concerning predictive justice has been published yet, it is possible to mention, even in this field, the already cited AgID's White Paper on Artificial Intelligence (Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino), of 2018.

Regarding the implementation of international sources, for the reasons explained above, it is important to recall the implementation of EU Directive n. 680/2016; moreover, it is necessary to mention Articles 7 CFREU and 8 ECHR, concerning the right to respect for private life, and Article 8 CFREU, on the right to data protection.

Moreover, as it will be said below, other supranational principles come into play regarding predictive justice, as specific profiles of the fundamental guarantee of the access to a human judge, implicit in the provisions of Article 6 ECHR and of Article 111 of Italian Constitution, on fair trial.

With specific reference to risk assessment tools in the field of personal liberty, Article 5 ECHR establishes two significant aspects of this right: firstly, the right to be conducted promptly before a human judge; secondly, the right to an effective control on the legitimacy of the restriction of personal liberty, within an adversarial procedure, which should involve the defence's access to the functioning of the algorithm.[13]

Moreover, as it will be explained below, either supranational and internal sources establish the principle of foreseeability of law, provided for by Articles 7 ECHR, 49 CFREU and 25 of the Italian Constitution (see the complete Italian report, e-RIDP 2023, Section I, 3.9.).

2.2 Case Law

Criminal tribunals or courts haven't been confronted with AI-based systems used for predictive justice.

To our knowledge, the same can be said about the civil and the constitutional courts, and about other independent authorities: none of them have issued decisions concerning this subject. However, the conclusion is partly different as regards the administrative courts. In fact, it is worth noting the case law of the Italian administrative supreme court, the State Council, which emphasises the need for AI-based systems to respect the fundamental principles of the Italian legal system (see the complete Italian report, e-RIDP 2023,

[13] Mitja Gialuz, 'Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa', [2019] *Diritto penale contemporaneo* (29th May 2019), 14-15.

Section I, 2.8. and 3.6). Despite the fact that these judgements specifically refer to administrative decisions, they seem relevant due to the general scope of their assertions.[14] According to the Italian judges, a procedure involving AI should not be stigmatised, but rather, in principle, encouraged: it has many advantages, such as, for example, the significant reduction in the time frame for purely repetitive and discretionary operations; the exclusion of interference due to negligence of the competent person and the consequent greater guarantee of impartiality of the automated decision.[15] However, the Council of State establishes that using artificial intelligence must be in line with certain fundamental principles: the principles of accessibility and transparency, the prohibition of decisions based solely on the automated processing of data and the prohibition of algorithmic discrimination. Thus, an important conclusion can be drawn from this position. If the Italian legislator decided to introduce risk assessment tools in the criminal area, these principles would prevent him from adopting tools similar, for example, to the well-known AI-based system COMPAS, which is clearly in contrast with the principle of transparency. Finally, the statements of the Italian administrative supreme court have been warmly welcomed by legal commentators, arguing for the extension of these dicta to criminal matters as well.[16]

2.3 Substantive Guarantees

In light of what we reported above, in Italy specific legislation about the reliability, impartiality, equality, and adaptability of AI-based systems used for predictive justice is absent. Anyway, it is important to once again draw attention to the case law of the Italian administrative supreme court about using AI in the decisions adopted by public administrations in administrative matters. As already observed, the guarantees elaborated by the Council of State may constitute a significant barrier to the use of AI-based systems used for predictive justice that contrast with the above-mentioned fundamental principles. This conclusion appears relevant even concerning criminal proceedings.

Given that there are no tools officially used for predictive justice in the Italian legal system at the moment, the answers to the other questions related to this sub-section are negative or not relevant.

3 General principles of law

At the time being, Italian society is not ravaged, like others, by issues of discrimination. Due to many social factors, the public's complaints about justice do not focus on discriminatory aspects, rather on the time-consuming process of it, inducing a strong feeling of ineffectiveness. However, if there is an aspect of predictive justice that has captured the

[14] Jacopo Della Torre, 'Le decisioni algoritmiche all'esame del Consiglio di Stato', [2021] *Rivista di diritto processuale*, 724 ff.

[15] State Council , sect. IV, 4 February 2020, n. 881.

[16] Jacopo Della Torre, 'Le decisioni algoritmiche all'esame del Consiglio di Stato', [2021] *Rivista di diritto processuale*, 724 ff; Mitja Gialuz, 'Intelligenza artificiale e diritti fondamentali in ambito probatorio', in *Giurisdizione penale, intelligenza artificiale ed etica del giudizio*, (2021, Giuffré), 66 ff.

interest of a small part of the literature, it is the risk of discriminatory practices being hidden behind the black box of predictive software.[17] Against the backdrop of several, serious problems in the functioning of our criminal justice system, what is certainly accepted and widely recognised is that the current code of criminal procedure offers wide room for discussion and confrontation between the parties, based on a strongly rooted adversarial scheme, especially with regard to the first instance decision. For these reasons, reading about the Loomis case, ruled by the Supreme Court of Wisconsin some years ago, focused a part of the literature on the unprecedented risk of importing the usage of implicitly discriminatory software and practices into criminal proceedings. The topic of risk assessment immediately echoed the period of Lombroso's theories, also based on discriminatory beliefs, attracting a strong reaction of mistrust. However, for the many reasons explained above, the feeling is that of observing a phenomenon that goes on abroad, with no serious repercussions for our legal system.

Burgeoning case-law of the ECtHR established that the fundamental guarantee of judicial independence is multifaceted. It implies different layers of action, both external and internal to the judiciary, objective and subjective. As Italian contemporary history also demonstrates, the most dangerous external incursion into judicial independence is represented by other public powers and authorities. Nevertheless, private interference in the function may also be detrimental and thus, contrary to the Convention. In fact, social constraint is a negative factor impinging on judicial independence, regardless of the form that it holds: large-scale access to 'prediction' can be a serious risk for constraint over the judge, who would feel 'encouraged' to follow the normative force of numbers… And this would have an impact also on the internal side of independence, even in legal systems which are not based on the stare decisis rule. The ECtHR has stressed the importance that the organisation of each judicial office allows each individual to perform their task without being influenced by other judges, from higher courts or from the same court: in particular, the Court endorsed the freedom of a judge from her peers' influence (ECtHR, *Findlay v. UK*, 25.2.1997). Pushing judges to use and follow the 'prediction' would mean constraining them into the respect of decisions taken by other judges, that – due to the public's expectations - they should feel forced to comply with. The intensity of the constraint could vary, according to the position taken by each institutional context: from a bare moral-suasion to a precise disciplinary duty to follow the prediction, clearly impinging on the independence of every single judge, in case a legal order should recognise in case-law consistency a superior interest in the administration of justice.

The right to access to a human judge has not been formalised yet because it can be considered an implicit premise of the whole range of guarantees of a fair trial. Going either from Art. 6§1 ECHR or Art. 111 §§ 1-3 of the Italian Constitution, the whole theory of the prerogatives of the fair judge is implicitly based on a human judge. All notions of 'tribunal', 'independence', 'impartiality' have been elaborated by the ECtHR (and by national constitutional courts) on the basis of a function performed by humans. Even the most

[17] Claudio Castelli, Giustizia predittiva [2022] Questione Giustizia (8th February 2022)

recent analysis of the Strasbourg case-law demonstrates that the court takes it for granted that such prerogatives play their crucial role in a context in which the judiciary is human. Probability, doubt, conviction, are all processed by human intuition, human comprehension of the facts, and the reasoning of the decision grants accessibility to that process. As a consequence, the center of the discussion should be whether a non-human decision is a judicial decision at all.[18] I personally doubt it. While it is important to study if and how automated decision systems may improve the real problems of criminal justice, it is crucial to recognise that, although assisted by a sufficient range of guarantees, automated decisions will not be judicial decisions, leaving room for an appeal to human courts.

Given the general ban of psycho-criminological expertise in the judgement upon the merits, digital risk assessment tools could be used in Italy only in the correctional phase, that is to say after the conviction became final and the presumption of innocence has been defeated for good.

As to the right to a fair trial, it is suitable to proceed from one specific feature of quantitative legal prediction, or predictive justice. In my book I tried to demonstrate that quantitative law predictions (QLP) work on the basis of correlations established between a pending case and previous decisions in similar cases. Setting aside the traditional discussion, in the common law, about the concept of similarity between two cases, I tried to list a very general set of variables that are impinging on a decision in criminal cases. The relevant variables can be divided into:[19] i) substantive and ii) procedural. The substantive variables can be distinguished into: a) material and b) legal. The material variables refer to: a1) factual (actus reus) and a2) subjective (mens rea) elements.

i) Substantive; a) material; a1) factual variables: the conduct, the subsequent natural even, causality, the circumstances of the case, aggravating or mitigating, in case of attempt (or other forms of inchoate offence), the stage of development in the action

i) Substantive; a) material; a2) subjective variables: mens rea (bearing in mind that different legal systems provide for different classifications of mens rea), liability, omplicity (in some jurisdiction, like the Italian e.g., complicity refers to the material aspect and not the subjective one), propensity towards crime

i) Substantive, b) legal variables: Nomen iuris, any legal condition impinging on the sentence, such as recidivism and other forms of habitualness in re-offending, statute of limitation (usually depending on nomen iuris: e.g. classifying the facts under a different nomen iuris, the statute of limitation may not occur), continued or concurrent offences.

ii) Procedural variables are much more complicated to list.[20] Although not the only relevant aspect, the evidence process is the factor that most affects the possibility to establish

[18] Antonio Punzi, Judge in the machine, in A. Carleo, *Decisione robotica* (2019, Il Mulino), 319 ff.
[19] Serena Quattrocolo, Artificial Intelligence, Computational Modelling and Criminal Proceedings (2020 , Springer), 201 ff.
[20] Ibid., 203.

relevant variables, to build up useful correlations. All the elements listed above must be appreciated on the basis of evidence provided by the parties (although many European legal orders recognise the judges ex officio power of introducing decisive evidence): based on such evidence, the judge must reach conviction beyond any reasonable doubt, on each of the listed points. However, the whole evidence process relies on discretional judicial evaluations. It is highly debatable that a quantitative computational model can grasp useful correlations in the realm of assessing reliability beyond any reasonable doubt, not only because of an intrinsic limitation of AI techniques. In fact, in this field, establishing similarities between different cases is particularly complicated, because the factors listed above are discretionally weighted by the judge, in each case. For instance, comparing two cases from the standpoint of the judicial evaluation of mitigating and aggravating circumstances may prove pointless: given the same circumstances, judge 1 may consider the aggravating circumstances predominant, while judge 2 may reach the opposite conclusion. And, although a machine learning system is trained to learn from previous errors, there are no right or wrong decisions, but only decisions giving different, discretional interpretations of the same factor.

As to the procedural aspects, the selection and admission of evidence is the first step of a crucial process that will bring the judge to 'establish the truth', in the adjudication of the case. Every legal order provides for different criteria to admit evidence, and it is difficult to generalise. However, it is possible to argue that all over the world, the courts' activities must orient the fact-find towards the truth, encouraged to pursue these basic goals through admitting and taking into consideration any evidence that appears to be factually relevant for the disposition of the case. 'Relevance' is a multi-fold concept that not only undergoes the typical process of a discretional evaluation but can vary significantly, depending on the stage of the procedure in which it is applied.

Moreover, criminal evidence evaluation is a factor that seriously impinges on the effectiveness of a computational model. In fact, evidence in criminal proceedings is still mostly oral. Documents are valuable and frequently used as evidence, of course; however, witnesses and expert-witnesses are key evidence in the majority of cases. The evaluation of the accuracy, reliability, and relevance of each piece of evidence is crucial in the decision of a case and cannot be standardised into a computational model. Based on very similar factual elements and evidence, two cases can be decided in different ways, because of a different evaluation of reliability and accuracy of a witness or expert witness.

For these reasons, the whole discussion about the risk of jeopardising the right to a fair trial must rather concentrate on the question of whether a QLP process is a trial at all. In fact, as demonstrated, a trial needs to take into account a long list of variables. A reliable automated decision-making process based on QLP can be conceived exclusively in relation to simple cases, in which the number of the variables at stake, both material and procedural, are extremely limited. Outside these boundaries, there cannot be the illusion of accomplishing the task of a trial, either fair, or unfair: the number of variables and the subjectivity in the evaluation of such variables exclude that QLP from performing the same task and functions of a trial. If the number of variables is extremely reduced and

the consistency of the case law is extremely high, QLP may possibly perform a coincident function, delivering a result which may satisfy the parties, but that, according, to 3.3., is not a judicial decision and should likely be submitted to a court for a human review.

Given the absence of currently used QLP systems, the effectiveness of the right to defence has not been discussed in practical terms, so far. However, the most basic idea of right to defence implies the defendant's right to give arguments upon each of the variables that have been listed in 3.5. The basic idea of equality of arms elaborated by the ECtHR (see *Martinie v. France*) implies that every part of the trial must be in the condition to convince the judge upon her reconstruction of the facts. It is arguable whether predictive systems may comply at all with this basic principle.[21]

The matter does not change in terms of appeals. Appeals are meant to be solutions against judicial errors and, insofar, they must allow the parties to convince the court of their reconstruction of facts, against a first instance decision that came to the wrong conclusions. In this sense, the right to appeal, as a refined aspect of the most general right to defence, would be totally illusory in a process in which a real second instance judgment would not be allowed, like in the case mentioned by the question, of using the same AI system.

Without lingering over aspects that will be analysed in Section III, it is crucial to reflect on whether and how it is possible to assess the reliability of data and the correctness of a calculation generated by a digital system. The Italian administrative supreme court (Consiglio di Stato) has recently recognised the right of those who suffer the effects of an algorithmic public decision to get a review on how the algorithm works and what the datasets used are (Consiglio di Stato: Sez. VI, 8.4.2019, n. 2270; Sez. VI, Sent., 13.12.2019, n. 8472; Sez. VI, 2.1.2020, n. 30; Sez. VI, 4.2.2020, n. 881). Although this position is not referred to judicial decisions, it appears to be a paradigm for algorithmic decisions taken, at any level, by the public administration.[22] With more specific regard to criminal proceedings, given that the minimum standard of the equality of arms is the chance to 'effectively influence the court's decision', what if the defendant claims that the impossibility to assess the reliability of an automatedly generated piece of evidence deprived her of the chance to 'effectively influence the court's decision'? In my opinion, it seems compliant with the principle of the equality of arms that the court discharges such automated calculations. Given that, in many cases, technology can provide sufficient validation of an automated process, when ex post validation is not available, the court should exclude the results of that process from the adjudication on the defendant's guilt, in order not to violate the basic expression of the fair trial, the equality of arms.

[21] Serena Quattrocolo, Artificial Intelligence, Computational Modelling and Criminal Proceedings (2020 , Springer), 90 ff.

[22] Jacopo Della Torre, 'Le decisioni algoritmiche all'esame del Consiglio di Stato', [2021] *Rivista di diritto processuale*, 724 ff.

As said above, predictive justice is arguably compatible with several constitutional principles and, in particular, independence and impartiality of judiciary, but also, in my opinion with the foreseeability of criminal law, which is provided for by art. 25 of the Italian constitution, art. 7 ECHR, and art. 49 CFREU. With specific regard to this aspect, the guarantee of foreseeability acknowledges that if the individual is not in a position to understand what the criminal law imposes or bans, her compliance with the law cannot be expected. Actually, inconsistency in the judicial interpretation of a criminal command can affect the foreseeability of what is legitimate and what is not. It was said that the foreseeability of the legal consequences of our actions is one of the dimensions of legal certainty. More precisely, certainty guarantees that individuals, before acting, can foresee these three aspects: whether their conduct will be considered legitimate; if illegitimate, whether it will amount to a crime; what punishment they may undergo. Insofar, predictive justice seems to improve and foster the certainty and foreseeability of criminal law, reducing the risk of inconsistency in the case-law. However, even from a semantic perspective, the distinction between fore-see (pre-vedere, in Italian) and fore-tell, (i.e. to pre-dict, based on etymology, pre-dire, in Italian) is based on the same difference as between 'foreseeability of the criminal law' and 'predictive justice'. The fundamental right of legal certainty, established by the main bills of human rights in the world, is a guarantee of accessibility, comprehensibility, awareness that people must have of the penal consequences of their behaviours.[23] The interpretation by the Strasbourg Court reiterates that it is a matter of cognitive comprehension[24], the complexity of which implies the suitable intervention of a counsel. On the contrary, 'predictive justice' is not aimed to clarify the meaning and the comprehension of legal precepts, but to predict the outcome of a potential litigation. It is a prediction of the success rate of an action and not an instrument to clarify the interpretation of the law. In this sense, such instruments are not supposed to foster the principle of legality and the accessibility of the criminal precept. Rather, they are about the personal expectation of the decision in each specific case, and this is much different from the core guarantee protected by the principle of legal certainty. Far from enhancing legal certainty, predictive justice appears to reduce or exclude the individualisation of the judicial response: individualisation is crucial not only in sentencing but throughout the whole of criminal proceedings. The result of a judgment must depend on the peculiar circumstances of each case, otherwise other fundamental rights would be violated, such as Art. 6 ECHR, as said above. Predictive justice can seriously jeopardise such a right.

Some have argued that applying a computational model based on precedent decisions of a court means to predict the decision in a new case. In fact, what QLP can do is to provide accurate calculations of how a court or a judge decided in previous cases on a

[23] Serena Quattrocolo, Artificial Intelligence, Computational Modelling and Criminal Proceedings (2020 , Springer), 219 ff.

[24] Alessandra Santangelo, 'Ai confini tra common law e civil law: la prevedibilità del divieto nella giurisprudenza di Strasburgo', [2019] *Rivista italiana di diritto e procedura penale*, 332 ff.

similar claim. Looking at the achievements in AI from the legal point of view, it has been said that the best that the modelling of judicial decisions can do, so far, is to give a model of the possible different solutions to a legal problem, as it has been observed. As a consequence, such programs can express probabilities, even high probabilities, that the court will stick to the precedent (especially in a common law jurisdiction), refusing to distinguish or to overrule, or that it will follow the mainstream interpretation, rejecting the more eccentric one (in a civil law jurisdiction). Thus, it is undisputed that the first step to address the matter of 'predictive justice' is to abandon the (dystopic) idea of a machine being able to predict the outcome of a future decision, that is to say, being able to decide: such a process delivers probabilities, based on what has occurred in the past and no serious scientific approach can use past events to build previsions of the future. As said, it is still the case of a law machine 'to inform', rather than a law machine 'to decide'. Given the undisputed value of statistics in hard science, it has been said that the application of probabilities to human activity is non-sensical, as it is governed by uncertain rules and factors, constantly changing over time. These remarks seriously impinge on how the purported 'accurate numbers' of QLP can be used. As an example, in many legal orders prosecutors have discretional power to drop cases they are not interested in. Within such a framework, the judicial statistics completely overlook the cases dropped by the prosecutor, either because of an 'immunity bargain', based on a deal between the suspect and the prosecutor, or other discretional evaluations. Such cases escape a complete review of precedents, as they do not even reach the trial stage. For these cases, there is no judicial decision and thus they are completely overlooked by the software. This has a great impact on the reliability of the results of 'predictive justice tools'. Moreover, the case outcome depends, in reality, on how the parties express their arguments, on the evidence they bring, on the application of procedural rules based on incidental conditions. Although modern techniques of natural language processing have reduced the gap between human and digital processing in the realm of semantics, the distinction between syntactic and semantic elaboration is still crucial. Traditionally, data is syntactically processed, while information is semantically processed: digital agents outperform human agents in syntactic analysis; humans excel in semantics, while digital technologies are not able to process data with a semantic function. Given that natural language processing has been at the centre of AI and law research, since the very initial stages and the achievements in such a field have been great, it is still really hard to establish similarities between different cases, due to the hindrance of such a semantic gap, provided that, in law, there is no 'application', rather 'interpretation'. For these reasons, the epistemological discussion has still to reach its higher level: the most recent literature opened the floor for such a discussion, which will be enlarged and enriched, also in Italy, in the next decade.[25]

[25] Antonio Punzi, Judge in the machine, in A. Carleo, Decisione robotica (2019, Il Mulino), 319 ff.; Massimo Durante, Potere computazionale. L'impatto delle ICT su diritto, società e sapere, (2019, Meltemi), passim.

Against the backdrop of the current situation, in Italy there is no discussion about the risk of privatisation of criminal justice. Given the large space devoted by the Italian constitution to Justice (from the point of view of judiciary recruitment and organisation but also the basic guarantees of the jurisdiction), I do not see real risks in this sense.

For several reasons, in many countries, including Italy, income creates a divide between an effective and ineffective defence. Predictive justice seems to be able to rephrase this axiom into expensive and inexpensive justice, forcing the not wealthy to accept the consequences of automated and possibly unfair justice, while the wealthy may have the chance to access more expensive human justice. The risks have been listed above and it is certainly not desirable to leave the foundation for such a scenario. This remark reinforces the wish that: 1) predictive justice is narrowed down to a few offences, in which: a) the number of variables to be examined is extremely limited; b) cases are repetitive and the judge's individual intuition does not play a relevant role in the decision; 2) predictive justice is used only with the purpose to tackle the backlog of cases: in this sense, given the defendant's right to appeal to a human judge, the automated solution of the case is acceptable as an alternative to the relinquishment of the file, that would consist of a patent denial of justice.

Selected literature

Bourcier D (2011), 'L'acte de juger est-il modélisable ? De la logique à la justice', in *Cahiers de Philosophie du Droit*, L'E-justice, Dalloz

Brigan J M (2018), 'Les risques accentués d'une justice pénale predictive', in *Archives de Philosophie du Droit*, La justice predictive, Dalloz, Paris

Castelli C, *Giustizia predittiva, Questione Giustizia* (2022) (8th February 2022)

Cevolani G, Crupi V, *Come ragionano i giudici: razionalità, euristiche e illusioni cognitive, Criminalia*, (2017), 181 ff

Della Torre J, 'Le decisioni algoritmiche all'esame del Consiglio di Stato', in *Rivista di diritto processuale*, (2021), 724 ff

Deumier P, 'La justice prédicitive et les sources du droit : la jurisprudence du fond', in *Archives de Philosophie du Droit*, La justice predictive, (2018) Dalloz, Paris

Durante M, *Potere computazionale. L'impatto delle ICT su diritto, società e sapere*, (2019), Meltemi, Milano

Engel C, Gigerenzer G, 'Law and Heuristics. An Interdisciplinary Venture', in G. Gigerenzer, C. Engel (eds) *Heuristics and the Law*, (2006), MIT Press, Cambridge (MA)

Galgani B, *Forme e garanzie nel prisma dell'innovazione tecnologica*, (2022), Wolters Kluwer, Milano

Gialuz M, 'Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa', in *Diritto penale contemporaneo* (2019), (29th May 2019), p. 1 ff

Gialuz M, 'Intelligenza artificiale e diritti fondamentali in ambito probatorio', in *AA.VV. Giurisdizione penale, intelligenza artificiale ed etica del giudizio*, (2021), Giuffré Milano, 66 ff

Katz D M., 'Quantitative legal prediction or how I learned to stop worrying and start preparing for the data-driven future of the legal services industry', in *Emory Law Journal*, (2013) p. 928 ff

Nieva Fenoll J, *Inteligencia artificial y proceso judicial*, (2018) Marcial Pons, Madrid

Ost F, '(Re)Learning to Think About Law from Cases', in S. Glanert (ed.), *Comparative legal reasoning. Essays in Honour of Geoffrey Samuel*s, (2018), Wildy, Simmond&Hill Publishing, London, 67 ff

Posner R A., *How Judges Think*, (2008), Harvard University Press, Cambridge (MA)

Punzi A, 'Judge in the machine', in A. Carleo, *Decisione robotica*, (2019), Il Mulino, Bologna, 319-330

Quattrocolo S, *Artificial Intelligence, Computational Modelling and Criminal Proceedings*, (2020), Springer, Chaim

Resta F, 'La direttiva sulla protezione dei dati personali in ambito giudiziario penale e di polizia e la tutela dei terzi' (2020), www.giustiziainsieme.it

Rottier E, 'La justice prédictive et l'acte de juger: quelle prévisibilité pour la justice?', in *Archives de Philosophie du Droit*, La justice predictive, (2018), Dalloz, Paris

Santangelo A, 'Ai confini tra common law e civil law: la prevedibilità del divieto nella giurisprudenza di Strasburgo', *Rivista italiana di diritto e procedura penal* 3 (2019), 332-357

Summers R S., 'Precedent in the US', in D.N. MacCormick, R.S. Summers, *Interpreting precedents*, (1997), Ashgate-Dartmouth, Aldershot

Taruffo M., 'Institutional Factors Influencing Precedents', in D.N. MacCormick, R.S. Summers, *Interpreting Precedents*, (1997), Ashgate-Dartmouth, Aldershot, p. 451 ff

Zara G, *Valutare il rischio in ambito criminologico. Procedure e strumenti per l'assessment psicologico*, (2016), Il Mulino, Bologna

Zara G, Farrington D P., *Criminal Recidivism: explanation, prediction and prevention*, (2016), Routledge, Oxon

PREDICTIVE JUSTICE IN THE UNITED STATES OF AMERICA

*By Emily Silverman**

Abstract

Rapid growth in the use of increasingly sophisticated risk assessment tools in criminal justice systems across the United States is due in part to reform efforts undertaken to reduce the country's extremely high incarceration rates. Other potential advantages of harnessing these tools, some of which already employ AI-based technology, include decreasing the disparities caused by the cash-bail system and providing outcomes at various stages of the criminal process that are fairer and less punitive than those produced by unfettered human decision-makers. Existing studies have not yet shown conclusively that these goals have – or have not – been achieved. In addition, use of AI-based tools implicates fundamental tenets of criminal procedure. As these tools become more prevalent, it remains to be seen how and whether courts and legislators will step up to protect these hard-won principles.

1 National practices

1.1 Definition of 'predictive justice'

There is no single official legal definition of 'predictive justice' in the United States; nevertheless, the term has been in use for decades. One working definition, articulated in early 2022 in an editorial on the use of artificial intelligence (AI) in the administration of justice, viewed it as a process involving the use of machine learning algorithms 'that perform a probabilistic analysis of any given particular legal dispute using case law precedents'.[1] Another aspect of predictive justice, discussed in a piece published in 2018, involves machine learning systems that employ risk-assessment algorithms to estimate the likelihood of recidivism.[2]

Writing in 2008, a prolific law professor referred to lectures he delivered at the University of Cincinnati in 1973 as 'an occasion to lay out a general theory of predictive justice ... to articulate a theory of preventive actions based on predictive decisions'.[3] The focus of the

* Senior Researcher, Max Planck Institute for the Study of Crime, Security and Law.

[1] Raffaele Giarda, 'International: Artificial Intelligence in the Administration of Justice' (*LegalBytes*, January 2022) <https://bakerxchange.com/rv/ff008a110bc355ed8ab8ecd14f4f8822ba8d30ae/p=0> accessed 27 March 2023.

[2] See, eg, Slava Polonski, 'Mitigating Algorithmic Bias in Predictive Justice: 4 Design Principles for AI Fairness' (24 November 2018) Towards Data Science <https://towardsdatascience.com/mitigating-algorithmic-bias-in-predictive-justice-ux-design-principles-for-ai-fairness-machine-learning-d2227ce28099> accessed 28 March 2023.

[3] Alan M Dershowitz, 'Visibility, Accountability and Discourse as Essential to Democracy: The Underlying Theme of Alan Dershowitz's Writing and Teaching' (2008) 71 Alb L Rev 731.

professor's theory was preventive confinement.[4] And, already decades earlier, there was so much literature on the topic of preventive justice in 1958 that a detailed description and illustration of 'the predictive devices developed for sentencing to various types of imprisonment, for placement on probation, for release on parole, and for predicting the postparole conduct of former prisoners over a considerable span of time' would require 'too extensive a discussion' for a single article.[5]

1.2 Selected AI-based systems used for predictive justice

According to the Partnership Report on Artificial Intelligence published in 2019,[6] criminal justice risk assessment tools are basic forms of AI, even though they are usually much simpler than the deep neural networks used in many modern AI systems. While some of them use heuristic frameworks to produce their scores, 'most use simple machine learning methods to train predictive models from input datasets.'[7] Arguably, there are a number of AI-based systems being used for predictive justice in the various jurisdictions of the United States. Three such systems will be introduced here: the commercially available Correctional Offender Management Profiling for Alternative Sanctions (COMPAS); the federal Prisoner Assessment Tool Targeting Estimated Risk and Needs (PATTERN); and the bespoke tool developed by the Pennsylvania Board of Probation and Parole.

COMPAS, a well-known commercially available instrument that seems to date to at least the late 1990s,[8] is widely used in the United States.[9] In Wisconsin, for example, it was implemented in 2012.[10] A proprietary algorithm sold by the private company currently

---

[4] Dershowitz (n 3) 745 ('although preventive confinement has always been and will always be practiced, no jurisprudence of preventive intervention has ever emerged. … . No philosopher, legal writer, or political theorist has ever, to this writer's knowledge, attempted to construct a systematic theory of when it is appropriate for the state to confine preventively.').

[5] Sheldon Glueck, 'Predictive Devices and the Individualization of Justice' (1958) 23 Law & Contemp Probs 461, 471.

[6] Partnership on AI is a 'non-profit partnership of academic, civil society, industry, and media organizations creating solutions so that AI advances positive outcomes for people and society'. Partnership on AI, 'About Us' <https://partnershiponai.org/about/> accessed 29 March 2023.

[7] Partnership on AI, 'Report on Algorithmic Risk Assessment Tools in the US Criminal Justice System' 7 (*Partnership on AI*, 23 April 2019) <https://partnershiponai.org/paper/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system> accessed 28 March 2023.

[8] JC Oleson, 'Risk in Sentencing: Constitutionally Suspect Variables and Evidence-Based Sentencing' (2011) 64 SMU L Rev 1329, 1349 fn125 (2011). See Mapping Pretrial Injustice, 'Common Pretrial Risk Assessments' <https://pretrialrisk.com/the-basics/common-prai> accessed 28 March 2023; Alexandra 'Mac' Taylor, 'AI Prediction Tools Claim to Alleviate and Overcrowded American Justice System … But Should They Be Used?' Stanford Politics (13 September 2020) <https://stanfordpolitics.org/2020/09/13/ai-prediction-tools-claim-to-alleviate-an-overcrowded-american-justice-system-but-should-they-be-used> accessed 28 March 2023; Tim Brennan, William Dieterich, and Beate Ehret, 'Evaluating the Predictive Validity of the COMPAS Risk and Needs Assessment System' (2009) 36 Crim Just & Behavior 21.

[9] EPIC, 'AI in the Criminal Justice System' <https://epic.org/issues/ai/ai-in-the-criminal-justice-system> accessed 28 March 2023.

[10] Andrew Lee Park, 'Injustice Ex Machina: Predictive Algorithms in Criminal Sentencing' (UCLA Law Review, 19 February 2019) <www.uclalawreview.org/injustice-ex-machina-predictive-algorithms-in-criminal-sentencing> accessed 28 March 2023.

known as equivant,[11] it is referred to as a fourth generation tool.[12] Fourth generation tools 'use machine learning in their modeling' and, in contrast to third generation tools, 'can output an explicit forecast, rather than a score'; when such a forecast is generated, 'it can be difficult to understand precisely what led to the system's determination'.[13] In the jurisdictions where COMPAS has been applied or adapted, judges may draw on the algorithm's output when making sentencing decisions.[14]

The Prisoner Assessment Tool Targeting Estimated Risk and Needs (PATTERN) was developed and implemented by the Federal Bureau of Prisons in accordance with legislation known as the First Step Act of 2018.[15] PATTERN, which was initially released in July 2019,[16] takes an AI-like approach.[17] It should be noted, however, that the question of whether machine learning was used to develop PATTERN – a question that was raised in Congressional testimony – was not immediately answered by the Department of Justice.[18] Staff of the Federal Bureau of Prisons use PATTERN to score inmates in their custody.[19]

---

[11] Three corporations, Northpointe, CourtView Justice Solutions, and Constellation Justice Systems, consolidated into a single branded entity called equivant in January 2017. Anne L. Washington, 'How to Argue with an Algorithm: Lessons from the COMPAS-ProPublica Debate' (2018) 17 Colo Tech LJ 131, 133 fn5.

[12] See, eg, Susan Turner and others, 'Development of the California Static Risk Assessment (CSRA): Recidivism Risk Prediction in the California Department of Corrections and Rehabilitation' 2 (September 2013) UC Irvine Center for Evidence-Based Correction Working Paper <https://bpb-us-e2.wpmucdn.com/sites.uci.edu/dist/0/1149/files/2013/12/Development-of-the-CSRA-Recidivism-Risk-Prediction-in-the-CDCR.pdf> accessed 28 March 2023.

[13] Michael E Donohue, 'A Replacement for Justitia's Scales: Machine Learning's Role in Sentencing' (2019) 32 Harv JL & Tech 657, 661. See also Orna Rabinovich-Einy and Ethan Katsh, 'Artificial Intelligence and the Future of Dispute Resolution – The Age of AI-DR' in Mohamed Abdel Wahab, Daniel Rainey, and Ethan Katsh (eds), *Online Dispute Resolution: Theory and Practice* (2nd edn, Eleven International Publishing 2021) pp. 471-488 (referring to COMPAS as an AI-based predictive algorithm).

[14] Ellora Thadaney Israni, 'When an Algorithm Helps Send You to Prison' The NY Times (New York, 26 October 2017) <https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html> accessed 27 March 2023.

[15] Public Law No 115-391, 132 Stat 5195 (21 December 2018). See Michael Santos, 'PATTERN Risk and Needs Assessment Under First Step Act' (Prison Professors) <https://prisonprofessors.com/pattern-first-step-act> accessed 28 March 2023. See also National Institute of Justice, *2021 Review and Revalidation of the First Step Act Risk Assessment Tool* (No 303859, December 2021) <www.ojp.gov/pdffiles1/nij/303859.pdf> accessed 28 March 2023.

[16] DOJ, 'Department of Justice Announces Enhancements to the Risk Assessment System and Updates on First Step Act Implementation' (15 January 2020 <www.justice.gov/opa/pr/department-justice-announces-enhancements-risk-assessment-system-and-updates-first-step-act> accessed 28 March 2023.

[17] Harold J Krent and Robert Rucker, 'The First Step Act - Constitutionalizing Prison Release Policies' (2022) 74 Rutgers UL Rev 631, 643.

[18] Amy B Cyphert, 'Reprogramming Recidivism: The First Step Act and Algorithmic Prediction of Risk' (2020) 51 Seton Hall L Rev 331, 360. '"The DOJ Report provides so few details on weighting, it is unclear what type(s) of models were used (such as regressions) and/or whether any type of machine learning (supervised or unsupervised) was employed."' Id. at 360 fn176.

[19] National Institute of Justice (n 15).

In 2013, members of the Pennsylvania Board of Probation and Parole began using machine learning forecasts to help inform discrete parole release decisions. Funding to develop the Board's state-of-the-art risk assessment tools was provided by the National Institute of Justice.[20]

At this point it should be emphasized that the question of which tools in use in the various criminal justice systems of the United States in fact rely on machine-learning algorithms does not lead to uniform, straight-forward answers. According to an article published in 2022, 'a number of states now rely on algorithmic and Artificial Intelligence ("AI") systems to fine tune the assessment of future dangerousness.'[21] In contrast, the following was claimed in a 2021 article:

> Algorithmic tools have taken root in some court systems at least as aids to human decision-making in criminal cases with respect to questions of bail, sentencing, and parole. But so far, virtually none of these tools appear to rely on machine-learning algorithms. … As best we can determine, only one jurisdiction (Pennsylvania) has implemented any risk assessment tool in criminal justice that is based on machine learning. … . Despite somewhat frequent claims to the contrary in the popular media, all other algorithmic tools used by courts appear to be based on standard indices or conventional logistic regression models – not machine-learning algorithms.[22]

This article refers specifically to COMPAS as a non-learning algorithmic tool.[23]

1.3 Description and role in the decision-making process of AI-based systems in use in the United States

According to some authors, AI-based systems have been in use in the criminal justice systems of the United States since at least the early years of the 21st century. Of these, risk assessment tools – some of which may incorporate machine learning – are used in a variety of contexts, including pretrial risk assessment (pretrial detainment/bail), sentencing, parole and probation, and prison rehabilitation programs.[24] According to a 2019 law

[20] Richard Berk, 'An Impact Assessment of Machine Learning Risk Forecasts on Parole Board Decisions and Recidivism' (2017) 13 J Experimental Criminology 193, 195. See also Aziz Z Huq, 'Racial Equity in Algorithmic Criminal Justice' (2019) 68 Duke LJ 1043, 1076.

[21] Krent and Rucker (n 17) 633 (footnotes omitted).

[22] Cary Coglianese and Lavi M Ben Dor, 'AI in Adjudication and Administration' (2021) 86 Brook L Rev 791, 801-803 (footnotes omitted).

[23] Coglianese and Ben Dor (n 22) 803.

[24] See Danielle Kehl, Priscilla Guo, and Samuel Kessler, 'Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing' (2017): <https://dash.harvard.edu/handle/1/33746041> accessed 28 March 2023. See also Mirko Bagaric and others, 'The Solution to the Pervasive Bias and Discrimination in the Criminal Justice System: Transparent and Fair Artificial Intelligence' (2022) 59 Am Crim L Rev 95, 130 (discussing impact of AI in the use of bail).

review article, AI, 'void of all human interaction, has been used to inform probation, sentencing, and parole decisions on the state level, and probation on the federal level'.[25]

In the context of sentencing, states tend to make the use of risk assessment tools advisory, rather than presumptive or mandatory.[26] In the Loomis decision of 2016, for example, the Wisconsin Supreme Court held that a sentencing court may consider a COMPAS risk assessment at sentencing but that COMPAS scores are but one of many factors that may be considered and weighed: risk scores alone may not be used to determine whether an offender is incarcerated or to determine the severity of an offender's sentence, and they may not be the determinative factor in deciding whether an offender can be supervised safely and effectively in the community.[27] As a result of the limitations placed on the use of COMPAS, the discretion of the decision-maker continues to play an important role in the sentencing process; furthermore, there is very little information available about how judges actually use risk assessments in practice.[28]

In the federal prison system, in contrast, eligibility for early release is determined by PATTERN alone. No discretion on the part of prison authorities is involved:[29] 'Unlike COMPAS, PATTERN is not just one factor that is weighed in deciding who is eligible for benefits like early release, it is THE factor.'[30]

Outside the field of risk assessment tools, AI does not yet seem to have advanced to the point where it is relied upon to 'produce judicial decisions', but it has been used in other ways, such as predicting a Supreme Court ruling on a particular issue.[31] Also, by analyzing massive amounts of data, software developed in recent years can assist in the exercise of legal judgment, work that was traditionally thought to be immune to automation.[32] For instance, thanks to its 'machine-learning, artificial intelligence and natural language

[25] Rachel DiBenedetto, 'Reducing Recidivism or Misclassifying Offenders: How Implementing Risk and Needs Assessment in the Federal Prison System Will Perpetuate Racial Bias' (2019) 27 JL & Pol'y 414, 417 (footnotes omitted).
[26] Brandon Garrett and John Monahan, 'Assessing Risk: The Use of Risk Assessment in Sentencing' (2019) 103 Judicature 42, 43.
[27] *State v Loomis*, 881 NW2d 749, 769 (Wis 2016).
[28] Garrett and Monahan (n 26), 43.
[29] Krent and Rucker (n 17) 634.
[30] Cyphert (n 18) 342 (emphasis in original).
[31] See, eg, Matthew Hutson, 'Artificial Intelligence Prevails at Predicting Supreme Court Decisions' (Science, 2 May 2017): <www.sciencemag.org/news/2017/05/artificial-intelligence-prevails-predicting-supreme-court-decisions> accessed 28 March 2023. See Taylor B. Schaefer, 'The Ethical Implications of Artificial Intelligence in the Law' 55 Gonz L Rev 221, 225.
[32] Dana Remus and Frank Levy, 'Can Robots Be Lawyers: Computers, Lawyers, and the Practice of Law' (2017) 30 Geo J Legal Ethics 501, 524 (2017).

processing technologies', Ravel Law, acquired by LexisNexis in 2017,[33] 'provides strategic insight into an array of factors that affect a judge's decision-making'.[34]

In conclusion, in 2019, 'the more fantastic ideas such as using AI to objectively decide cases by analyzing facts and applying law' were still 'figments of creative imaginations'.[35] And as recently as 2021, authors who knew of 'no machine-learning tool that has been adopted in any court in the United States to make an ultimate, fully automated determination on a legal or factual question',[36] made the following statement:

> Although it is still early in courts' assessment of judicial use of algorithmic tools, it seems noteworthy that, in all the cases decided to date that have actually wrestled with the issues, courts appear to have taken pains to emphasize that such tools only serve as one of multiple factors that a judge takes into account in reaching a decision.[37]

On the other hand, when John Roberts, Chief Justice of the US Supreme Court, was asked in April 2017 whether 'smart machines, driven with artificial intelligences, will assist with courtroom fact finding or, more controversially even, judicial decision making', he replied, 'It's a day that's here, and it's putting a significant strain on how the judiciary goes about doing things.'[38]

1.4 How AI-based systems used for predictive justice in the United States work

According to some scholars, COMPAS uses machine learning.[39] But because COMPAS is proprietary software, it is difficult to say much about how it functions. Indeed, 'there is almost no transparency about its inner workings'.[40] The COMPAS tool 'is organized around an algorithm that uses the answers to some 137 questions about a criminal suspect to rank them on a scale of 1 to 10 … with higher scores indicating a greater risk of

---

[33] LexisNexis, 'LexisNexis Announces Acquisition of Ravel Law' (8 June 2017): <www.lexisnexis.com/community/pressroom/b/news/posts/lexisnexis-announces-acquisition-of-ravel-law> accessed 28 March 2023.

[34] PRWeb, 'Ravel Law Announces Unprecedented Judge Analytics Offering' (16 April 2015) <www.prweb.com/releases/2015/04/prweb12656883.htm> accessed 28 March 2023.

[35] See also Richard C Kraus, 'Artificial Intelligence Invades Appellate Practice: The Here, The Near, and The Oh My Dear' (2019 Winter Edition) Appellate Issues: <www.americanbar.org/groups/judicial/publications/appellate_issues/2019/winter/artificial-intelligence-invades-appellate-practice-the-here-the-near-and-the-oh-my-dear> accessed 28 March 2023.

[36] Coglianese and Ben Dor (n 22) 798 (footnote omitted).

[37] Coglianese and Ben Dor (n 22) 811.

[38] Adam Liptak, 'Sent to Prison by a Software Program's Secret Algorithms' The NY Times (New York, 1 May 2017), A22.

[39] Donohue (n 13) 661. But Coglianese and Ben Dor (n 22) 803 (COMPAS is a 'non-learning algorithmic tool adopted by several state court systems for pretrial decisions'); Jeff Ward, 'Black Box Artificial Intelligence and the Rule of Law' 84 Law & Contemp Prob i, ii (2021)(referring to COMPAS as a simple, statistically-based algorithm).

[40] Kehl, Guo, and Kessler (n 24) 9, 11.

recidivism'.[41] It considers variables from five main areas (criminal involvement, relationships and lifestyles, personality and attitudes, family, and social exclusion) and uses a combination of static and dynamic factors to assess the risk of recidivism. The algorithm is 'largely considered to be a black box: though its basic input information is available, the weighting of these inputs within the algorithm are proprietary, and thus not available to the public'.[42]

PATTERN takes an AI-like approach,[43] where AI is defined as 'the ability of a machine to perceive and respond to its environment independently and perform tasks that would typically require human intelligence and decision-making processes, but without direct human intervention'.[44] Although PATTERN does not utilize a fully autonomous AI or machine-learning algorithm, 'its algorithm nonetheless provides the foundation for greater application as a more AI-like tool, including for example, automatic updating independent of human intervention.'[45] PATTERN (as updated following publication of the July 2019 Risk and Needs Assessment Report) incorporates fifteen factors: eleven dynamic and four static.[46]

As far as the bespoke Pennsylvania Board of Probation and Parole risk assessment tools are concerned, training data were provided by the state's Department of Corrections, and several machine learning procedures were applied. Random forests were determined to be the most effective.[47] The Pennsylvania tool was trained using data provided by the Department of Corrections. The data included information concerning the inmate's capacity for violence, sex offender status, conduct in prison, arrest and conviction history, gender, age, and intelligence as well as information from the inmate's Level of Service Inventory-Revised interview.[48]

---

[41] Huq, 'Racial Equity' (n 20) 1047.

[42] Taylor (n 8). See also Kehl, Guo, and Kessler (n 24) 11.

[43] Krent and Rucker (n 17) 643.

[44] Christopher Rigano, 'Using Artificial Intelligence to Address Criminal Justice Needs' (January 2019) NIJ Journal 280: <https://www.nij.gov/journals/280/Pages/using-artificial-intelligence-to-address-criminal-justice-needs.aspx> accessed 28 March 2023

[45] Krent and Rucker (n 17) 643 fn85.

[46] Dynamic factors: 1. Conviction(s) for any type of infraction during current incarceration period; 2. Conviction(s) for serious and violent infractions during current incarceration period; 3. Infraction-free (any type) during current incarceration period; 4. Infraction-free (serious and violent) during current incarceration period; 5. Number of programs completed (any); 6. Work programming; 7. Drug treatment while incarcerated; 8. Non-compliance with financial responsibility; 9. History of violence; 10. History of escapes; 11. Education score. Static factors: 1. Age at time of assessment; 2. Violent offense of conviction; 3. Sex offender status; 4. Criminal history score. See DOJ, 'The First Step Act of 2018: Risk and Needs Assessment System – UPDATE' 10–11 (footnotes omitted) (January 2020): <www.bop.gov/inmates/fsa/docs/the-first-step-act-of-2018-risk-and-needs-assessment-system-updated.pdf> accessed 28 March 2023.

[47] Berk (n 20) 195.

[48] See Berk (n 20) 195, 213-214.

1.5 Use of AI-based systems at various stages of the criminal process

Judicial authorities in numerous jurisdictions throughout the United States are required to use risk assessment tools at one or more stages of the criminal process. While the nature of the tools in question is not always clear, the Partnership Report states generally that 'criminal justice risk assessment tools are basic forms of AI', even if they are 'usually much simpler than the deep neural networks used in many modern artificial intelligence systems'.[49]

As far as the pretrial stage is concerned, information on the use throughout the country of risk assessment tools is provided by a website run since 2020[50] by the organizations 'Media Alliance Project' and 'MediaJustice'. According to the website ('Mapping Pretrial Injustice'), risk assessment tools are required by court order in Jefferson County, Alabama; they are required by legislation in Connecticut, Delaware, Hawaii, Kentucky, New Jersey, Rhode Island, Vermont, Virginia, and West Virginia; they are required by the state supreme court in Indiana and Nevada; and they are required by judicial council in Minnesota.[51]

Furthermore, the website reports that COMPAS is in use in at least 11 counties.[52] In addition to common national pretrial tools (such as COMPAS),[53] 11 states have developed their own tools and 49 counties across 22 states use locally-developed or otherwise county-specific tools.[54]

As far as sentencing is concerned, by 2017, 'numerous states', including Kentucky and Oklahoma, required sentencing judges to consider the results of 'evidence-based tools'.[55] In Kentucky, judges are required to consider the results of a defendant's risk and needs assessment;[56] in Oklahoma, the judge is required to review the defendant's risk and needs assessment if the defendant is a felony offender being considered for a community

[49] Partnership on AI, 'Report' (n 7) 7 ('Some [of the tools] use heuristic frameworks to produce their scores, though most use simple machine learning methods to train predictive models from input datasets.').
[50] Ethan Corey, 'New Data Suggests Risk Assessment Tools have Little Impact on Pretrial Incarceration' (*The Appeal*, 7 February 2020): <https://theappeal.org/new-data-suggests-risk-assessment-tools-have-little-impact-on-pretrial-incarceration> accessed 28 March 2023.
[51] Mapping Pretrial Injustice, 'State Laws on Rats' <https://pretrialrisk.com/national-landscape/state-laws-on-rats> accessed 28 March 2023.
[52] Mapping Pretrial Injustice, 'How Many Jurisdictions Use Each Tool?' <https://pretrialrisk.com/national-landscape/how-many-jurisdictions-use-each-tool> accessed 28 March 2023. Note that COMPAS has separate tools for use at sentencing and during the pretrial stage, Public Safety Risk Assessment Clearinghouse, 'Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)': <https://bja.ojp.gov/sites/g/files/xyckuh186/files/media/document/compas.pdf> accessed 28 March 2023.
[53] Mapping Pretrial Injustice, 'Common Pretrial Risk Assessments' <https://pretrialrisk.com/the-basics/common-prai> accessed 28 March 2023.
[54] 'How Many Jurisdictions' (n 52).
[55] John Lightbourne, 'Damned Lies & Criminal Sentencing Using Evidence-Based Tools' (2017) 15 Duke L & Tech Rev 327, 332.
[56] Ky Rev Stat § 532.007(3) (2022).

punishment pursuant to the Oklahoma Community Sentencing Act.[57] In other states, the use of risk assessment tools in the context of sentencing is advisory, rather than presumptive or mandatory.[58] In any case, given the lack of available information, it is difficult to determine what it means, in practice, for a sentencing courts to 'use' such a tool. Finally, the role of discretion in a decision-maker's determination should not be underestimated.[59]

1.6 Incentives for using AI-based systems

Rapid growth in the use of increasingly sophisticated risk assessment tools in criminal justice systems across the United States is due in part to reform efforts undertaken in order to reduce the country's extremely high incarceration rates.[60] Other potential advantages of harnessing these tools include decreasing the disparities caused by the cash-bail system[61] and providing outcomes that are fairer and less punitive than those produced by human decision-makers when they act with complete discretion.[62] One of the aims of the federal First Step Act of 2018, which led to the development of PATTERN, was to lower federal prison numbers by providing for the early release of non-violent offenders.[63]

1.7 Alternative dispute resolution based on AI calculations

Online dispute resolution, which is in the process of taking the place of alternative dispute resolution, has 'gained significant traction in the United States'.[64] While AI has begun showing up in this context, mostly in the form of AI-based predictions, its use has been limited, at least where decision-making is involved.[65]

1.8 Public, media, and scholarly responses to AI-based systems for predictive justice

AI-based systems for predictive justice have received a great deal of negative press in recent years. The media, NGOs, and legal scholars tend to emphasize the negative aspects of the technology, particularly the risks of bias that accrue to the detriment of poorer communities and communities of color, groups already suffering from structural racism and human-emanating bias in the criminal justice context. The public perception

---

[57] 22 Okl St § 988.18 (2022). The Oklahoma Community Sentencing Act is codified at 22 Okl St §§ 988.1–988.24.
[58] Garrett and Monahan (n 26) 43.
[59] Garrett and Monahan (n 26) 43.
[60] Partnership on AI, 'Report' (n 7) 7.
[61] Jessica Brand and Jessica Pishko, 'Bail Reform: Explained' (*The Appeal*, 14 June 2018) <https://theappeal.org/bail-reform-explained-4abb73dd2e8a> accessed 28 March 2023.
[62] Partnership on AI, 'Report' (n 7) 7.
[63] Bagaric and others (n 24) 123.
[64] Amy J Schmitz and Janet Martinez, 'ODR and Innovation in the United States' in Mohamed Abdel Wahab, Daniel Rainey, and Ethan Katsh (eds), *Online Dispute Resolution: Theory and Practice* (2nd edn, Eleven International Publishing 2021) pp. 611-638.
[65] Rabinovich-Einy and Katsh (n 13).

reflects this. Fewer scholars, it seems, focus on the advantages that AI-based systems have to offer.

1.9 Reliability and impartiality of AI-based systems for predictive justice in use in the United States

A 2013 study of 19 criminal risk and need assessment tools in use in the United States found that validity, in most cases, had been examined in 'one or two studies' and that these investigations were frequently completed by the very people who had developed the instrument.[66] Another study, conducted between September 2019 and July 2020 by the Electronic Privacy Information Center (EPIC),[67] consisted of a survey of state usage (pre-trial as well as other contexts) of risk assessment tools. A table summarizing the results of the survey indicates which of the numerous tools in use had been subject to a validation study. The table does not, however, indicate who carried out the study, nor does it identify tools that are AI-based.[68] As far as the performance of algorithmic risk tools with minorities is concerned, the few studies to date show 'some evidence that minorities are more likely to be ranked at higher risk, though this result is not consistent across studies and for all tools'.[69]

COMPAS has been evaluated by numerous entities, both independent and internal.[70] It has been the subject of research that questions accuracy, utility, and fairness.[71] In 2021, available validation studies of COMPAS were 'typically performed by employees, consultants, or research funding recipients' of the tool's owners.[72] In a summary published

[66] Sarah L Desmarais and Jay P Singh, Risk Assessment Instruments Validated and Implemented in Correctional Settings in the United States 2 (*Council of State Governments Justice Center*, 27 March 2013): <https://csgjusticecenter.org/wp-content/uploads/2020/02/Risk-Assessment-Instruments-Validated-and-Implemented-in-Correctional-Settings-in-the-United-States.pdf> accessed 28 March 2023.

[67] EPIC is an NGO that was established in 1994 to 'protect privacy, freedom of expression, and democratic values in the information age'. EPIC, 'About Us': < https://epic.org/about> accessed 29 March 2023.

[68] EPIC, 'Liberty at Risk: Pre-Trial Risk Assessment Tools in the US' (September 2020) <https://epic.org/wp-content/uploads/2022/02/Liberty-At-Risk-Report-FALL-2020-UPDATE.pdf> accessed 28 March 2023.

[69] Melissa Hamilton, 'Algorithmic Risk Assessment: A Progressive Policy in Pre-trial Release' (2021) 57 Idaho L Rev 615, 630 (citing Whitney Threadcraft-Walker and others, 'Gender, Race/Ethnicity and Prediction: Risk in Behavioral Assessment' (2018) 54 J Crim Just 12 (note that the studies do not indicate whether the tools were AI-based).

[70] See, eg, Northpointe Research and Development Department, 'COMPAS Scales and Risk Models – Validity and Reliability: A Summary of Results from Internal and Independent Studies' (*Elec Priv Info Ctr*, 20 July 2010): <https://epic.org/algorithmic-transparency/crim-justice/EPIC-16-06-23-WI-FOIA-201600805-COMPASSummaryResults.pdf> accessed 28 March 2023.

[71] Mirko Bagaric, Dan Hunter, and Nigel Stobbs, 'Erasing the Bias against Using Artificial Intelligence to Predict Future Criminality: Algorithms Are Color Blind and Never Tire' (2020) 88 U Cin L Rev 1037, 1044.

[72] Melissa Hamilton, 'Algorithmic Risk Assessment: A Progressive Policy in Pretrial Release' (2021) 57 Idaho L Rev 615, 628 (2021).

in 2010 of research findings from multiple studies,[73] the Northpointe Research and Development Department came to the overall conclusion that COMPAS was reliable and had both good predictive and construct validity.[74] The authors acknowledged that:

> much of the evidence for the reliability and validity of the COMPAS is found in the results of research studies conducted by Northpointe. We know that critics may discount this research. However, most of our in-house research is conducted for state agencies, and that [sic] competent research divisions within those agencies closely scrutinize our methods and results. Such state-sponsored studies are, thus, often subjected to a far more thorough vetting than that provided by the editors of peer-reviewed journals often resulting from the fact that such agencies have direct access to the same data, can scrutinize such data and often can replicate and test our findings.[75]

By law, PATTERN is subject to annual review and validation by the Attorney General.[76] In August 2020, the National Institute of Justice contracted with two investigators to serve as consultants and to conduct the annual review and revalidation of PATTERN.[77]

The impact of machine learning forecasts used to help the Pennsylvania Board of Probation and Parole make parole release decisions was evaluated in a paper published in 2017.[78]

---

[73] Northpointe Research and Development Department, 'COMPAS Scales and Risk Models – Validity and Reliability: A Summary of Results from Internal and Independent Studies' (*Elec Priv Info Ctr*, 20 July 2010) <https://epic.org/algorithmic-transparency/crim-justice/EPIC-16-06-23-WI-FOIA-201600805-COMPAS-SummaryResults.pdf> accessed 28 March 2023.

[74] 'Construct validity' is relevant with regard to COMPAS' 'needs' scales; 'predictive validity' is relevant with regard to its 'risk scales'. Northpointe Research and Development Department, 'COMPAS Scales and Risk Models – Validity and Reliability: A Summary of Results from Internal and Independent Studies' 2-4, 7 (*Elec Priv Info Ctr*, 20 July 2010): <https://epic.org/algorithmic-transparency/crim-justice/EPIC-16-06-23-WI-FOIA-201600805-COMPASSummaryResults.pdf> accessed 28 March 2023.

[75] Northpointe Research and Development Department, 'COMPAS Scales and Risk Models – Validity and Reliability: A Summary of Results from Internal and Independent Studies' 2 (*Elec Priv Info Ctr*, 20 July 2010): <https://epic.org/algorithmic-transparency/crim-justice/EPIC-16-06-23-WI-FOIA-201600805-COMPASSummaryResults.pdf> accessed 28 March 2023. A 2021 review of the literature on the two popular risk tools that were the focus of prior empirical studies, including COMPAS, revealed that 'the available validation studies were typically performed by employees, consultants, or research funding recipients of the tools' owners.' Melissa Hamilton, 'Evaluating Algorithmic Risk Assessment' (2021) 24 New Crim L Rev 156, 181.

[76] 18 USC § 3631(b)(4). See National Institute of Justice (n 15).

[77] National Institute of Justice (n 15).

[78] Berk (n 20).

1.10 Findings of studies mentioned in question 1.9

1.10.1 COMPAS

One of the evaluations of COMPAS, carried out by ProPublica[79] and published in 2016,[80] garnered an enormous amount of attention in both the popular media and the scholarly literature.[81] According to the ProPublica study, the risk scores calculated by COMPAS were 'remarkably unreliable in forecasting violent crime: Only 20 percent of the people predicted to commit violent crimes actually went on to do so'; furthermore, when a full range of crimes, including misdemeanors, was taken into account, 'the algorithm was somewhat more accurate than a coin flip.'[82] Also, as far as false positives were concerned, ProPublica found that 'the formula was particularly likely to falsely flag black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants' and that, with regard to false negatives, 'white defendants were mislabeled as low risk more often than black defendants.'[83]

The ProPublica study was not, however, without its detractors. For example, in September 2016, just a few months after the ProPublica study was published, an article strongly critical of its findings appeared in Federal Probation. Its authors stated:

> We think ProPublica's report was based on faulty statistics and data analysis, and that the report failed to show that the COMPAS itself is racially biased, let alone that other risk instruments are biased. Not only do ProPublica's results contradict

[79] ProPublica describes itself as 'an independent, nonprofit newsroom that produces investigative journalism with moral force'. With a team of more than 100 journalists, ProPublica 'covers a range of topics including government and politics, business, criminal justice, the environment, education, health care, immigration, and technology'. See ProPublica, 'About Us': <www.propublica.org/about> accessed 28 March 2023.

[80] Julia Angwin and others, 'Machine Bias' (*ProPublica*, 23 May 2016) <www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> accessed 28 March 2023; Jeff Larson and others, 'How We Analyzed the COMPAS Recidivism Algorithm' (*ProPublica*, 23 May 2016): <www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm> accessed 28 March 2023.

[81] For scholarly literature, see, eg, Mirko Bagaric and Gabrielle Wolf, 'Sentencing by Computer: Enhancing Sentencing Transparency and Predictability and (Possibly) Bridging the Gap between Sentencing Knowledge and Practice' (2018) 25 Geo Mason L Rev 653; Huq, 'Racial Equity' (n 20); Sandra G Mayson, 'Dangerous Defendants' (2017) 127 Yale LJ 490. For popular media, see, eg, Julia Angwin, 'Make Algorithms Accountable' The NY Times (New York, 1 August 2016) A17; Sam Corbett-Davies and others, 'A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased against Blacks. It's Actually Not That Clear' The Washington Post (Washington, D.C., 17 October 2016): <www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas> accessed 27 March 2023; Thadaney Israni (n 14).

[82] Angwin and others (n 80).

[83] Angwin and others (n 80).

> several comprehensive existing studies concluding that actuarial risk can be predicted free of racial and/or gender bias, a correct analysis of the underlying data (which we provide below) sharply undermines ProPublica's approach.[84]

And in October 2017, in response to the ProPublica-Northpointe contretemps, a group of computer science researchers wrote the following in the Washington Post:

> Algorithms have the potential to dramatically improve the efficiency and equity of consequential decisions, but their use also prompts complex ethical and scientific questions. … The problems we discuss apply equally to human decision-makers, and humans are additionally biased in ways that machines are not. We must continue to investigate and debate these issues as algorithms play an increasingly prominent role in the criminal justice system.[85]

1.10.2 PATTERN

As reported in the 2021 publication 'Review and Revalidation of the First Step Act Risk Assessment Tool' (Report), discrepancies with some of the measures used to create PATTERN version 1.2 were identified. The staff of the Bureau of Prison's Office of Research and Evaluation together with the National Institute of Justice's review and revalidation expert consultants collaborated to correct the discrepancies.[86] Updated data were used to created PATTERN version 1.3. Among other things, the Report reviewed and analyzed the predictive validity and the racial and ethnic neutrality of PATTERN version 1.3: According to the Report, the results suggested that PATTERN 1.3 displayed a high level of predictive accuracy. And, with respect to racial and ethnic neutrality, the review of the risk and needs assessment system (as mandated by the First Step Act) must include 'an evaluation of the rates of recidivism among similarly classified prisoners to identify any unwarranted disparities, including disparities among similarly classified prisoners of different demographic groups, in such rates'.[87] The Report contains results of evaluations of PATTERN using a number of approaches that 'reflect the current scientific standards for assessing instrument neutrality'.[88] Racial and ethnic neutrality was examined in a number of different ways, including through differential prediction analyses, which assess a key question: 'Do racial and ethnic subgroups have different probabilities of recidivism controlling for PATTERN score?' According to the Report, PATTERN shows relatively high predictive accuracy across the five racial/ethnic (White, Black, Hispanic, Native American, Asian) groups.

---

[84] Anthony W Flores, Kristin Bechtel, and Christopher T Lowenkamp, 'False Positives, False Negatives, and False Analyses: A Rejoinder to Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks' (2016) 80 Fed Probation 38.

[85] Corbett-Davies and others (n 81).

[86] National Institute of Justice (n 15).

[87] 18 USC § 3631(b)(4)(E).

[88] National Institute of Justice (n 15).

> The predictive value … and differential prediction results …, however, are mixed and complex. The differential prediction analyses reveal statistically significant results in 28 of 48 tests (analyses of main effects). These include the overprediction of Black, Hispanic, and Asian males and females on some of the general recidivism tools and the underprediction of Black males and females and Native American males, relative to white individuals, on some of the violent recidivism tools. The magnitudes of differential prediction include:
>
> • 6 to 7 percent relative overprediction for Black females on the general recidivism tool
>
> • 12 to 15 percent relative underprediction of Native American males and females on the general recidivism tools
>
> • 5 to 8 percent relative overprediction of Asian males on the general and violent recidivism tools
>
> … . Statistically significant results do not necessarily invalidate a tool, particularly with large sample sizes. However, due to the importance of the FSA mandate to examine the risk and needs assessment system for racial and ethnic neutrality, these results will be a central focus of subsequent review and revalidation efforts.
>
> The NIJ consultants will also continue to investigate potential solutions for the differential prediction issues identified during this review, including testing emerging debiasing techniques and engaging with stakeholders to explore the most promising and supportable approaches.[89]

According to an evaluation of PATTERN carried out by a legal scholar and published in 2020, the tool 'will have a disproportionate impact on Black inmates'[90] although it includes 'certain best practices in recidivism prediction, and its developers have made a good faith effort to engage advocates and scholars about the tool's development'.[91]

1.10.3 Pennsylvania Board of Probation and Parole's machine-learning protocol

The performance evaluation of the Pennsylvania Board of Probation and Parole's machine-learning protocol, published in 2017, showed that the machine learning forecasts 'apparently had no effect on the overall parole release rate but did appear to alter the mix of inmates released'.[92] The forecasts appeared to lead to reductions in rearrests for both nonviolent and violent crime.[93]

---

[89] National Institute of Justice (n 15).
[90] Cyphert (n 18) 331.
[91] Cyphert (n 18) 381.
[92] Berk (n 20).
[93] Huq, 'Racial Equity' (n 20) 1076. See Berk (n 20) 212-213.

1.11 Neutrality compared: AI-based systems used for predictive justice versus humans

The question of whether AI-based systems for predictive justice provide more neutrality in the criminal justice system than humans has been and continues to be hotly debated. One side of the debate emphasizes the advantages of AI-based systems. This approach points out that the current process for making sentencing decisions, a process dominated by judges, has been shown to be heavily biased against disadvantaged groups, and it refers to research findings showing that under this process 'groups such as African Americans and unattractive people receive disproportionately heavier sentences than other people.' Scholars in this camp emphasize the fact that algorithms, unlike humans, 'have no subconscious thinking paths' and 'do exactly what they are programmed to do'.[94]

On the other side of the debate are exponents of Melvin Kranzberg's first law of technology.[95] The following quote, made in the context of AI risk assessments, stems from an AI sceptic: '[T]hese algorithms are neither good nor bad, but they are certainly not neutral. To accept AI in our courts without a plan is to defer to machines in a way that should make any advocate of judicial or prosecutorial discretion uncomfortable.'[96]

1.12 Consistency compared: AI-based systems versus humans

Opinions regarding the consistency of AI-based systems compared to that of humans appear to be mixed and findings limited. In one law review article, published in 2021, the authors wrote in support of the consistency of machine-learning tools: 'If machine-learning tools are used as substitutes for – or even just as complements to – human decision-making, they could potentially reduce inconsistencies and other foibles that permeate human judgment.'[97] In contrast, the authors of a 2017 law review article were more critical of risk assessment tools and software, including those that incorporate machine learning. In their eyes, while such tools 'have the potential to improve sentencing accuracy in the criminal justice system and reduce the risk of human error and bias, they also have the potential to reinforce or exacerbate existing biases and to undermine certain basic tenets of fairness that are central to our justice system'.[98]

---

[94] Mirko Bagaric, Dan Hunter, and Nigel Stobbs, 'Erasing the Bias against Using Artificial Intelligence to Predict Future Criminality: Algorithms Are Color Blind and Never Tire' (2020) 88 U Cin L Rev 1037, 1039, 1065, available at: https://scholarship.law.uc.edu/uclr/vol88/iss4/3.

[95] 'Technology is neither good nor bad; nor is it neutral.' Melvin Kranzberg, 'Technology and History: "Kranzberg's Laws"' (1986) 27 Technology and Culture 544, 545.

[96] Jason Tashea, 'Courts Are Using AI to Sentence Criminals. That Must Stop Now' 17 April 2017, Wired: <www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now> accessed 28 March 2023.

[97] Coglianese and Ben Dor (n 22) 828 (footnote omitted).

[98] Kehl, Guo, and Kessler (n 24) 36.

A qualitative study published in 2020 examined attorney attitudes – specifically, prosecutors and defense attorneys – towards risk assessment in sentencing and plea bargaining.[99] The findings of the study can be summarized as follows:

> Prosecutors, for example, favored the use of risk assessment tools for sentencing, arguing they "were likely a more consistent and fair way than relying on intuition or personal experience." On the other hand, defense attorneys "were consistently opposed to using future recidivism risk as a factor in sentencing," as tools measuring future recidivism were based on "group means" rather than individual ones.[100]

1.13 Effect of AI-based systems on responses to crime

There is not a great deal of information available to show whether AI-based systems lead to harsher or more lenient responses to crimes or other violations of the law. In a 2017 assessment of the machine learning risk forecasts used by the Pennsylvania Board of Probation, it was found that the forecasts did not seem to have an effect on the overall parole release rate but did seem to alter the mix of inmates released.[101] The conclusion was that 'risk assessments based on machine learning forecasts can improve parole release decisions, especially when distinctions are made between re-arrests for violent and nonviolent crime.'[102]

1.14 Future of AI-based systems for predictive justice purposes

Despite controversy surrounding the use of (AI-based) systems for predictive justice, numerous jurisdictions in the United States continue to use COMPAS, and PATTERN remains in use on the federal level. Examples of public authorities that have terminated their use of AI-based systems for predictive justice purposes are not readily apparent.

2 Normative framework

2.1 National legal rules governing the use of AI-based systems for predictive justice

As of 2023, there were no national legal rules specifically governing the use of AI-based systems for predictive justice in the United States.

> Given the federalist structure of the United States, the development and implementation of AI technology in the public sector … is not determined by any central institution. … Decisions about digital technologies used by courts throughout the

[99] Anne Metz and others, 'Valid or Voodoo: A Qualitative Study of Attorney Attitudes Towards Risk Assessment in Sentencing and Plea Bargaining' 8 (10 March 2020) Virginia Public Law and Legal Theory Research Paper No 2020-25, Duke Law School Public Law & Legal Theory Series No 2020-15: <https://papers.ssrn.com/sol3/papers.cfm?abstractid=3552018> accessed 28 March 2023.
[100] Bagaric and others (n 24), 122 (footnotes omitted).
[101] Berk (n 20) 193.
[102] 'An Impact Assessment of Machine Learning Risk Forecasts on Parole Board Decisions and Recidivism' (2017) 13 J Experimental Criminology 193, 193.

> United States are … made by a plethora of institutions and actors. … Any one of these numerous … entities could in principle have its own policy with respect to … the use of algorithms to support decision-making.[103]

In April 2021, however, a bill, the 'Justice in Forensic Algorithms Act of 2021,'[104] was introduced in the US House of Representatives. Had it been enacted into law before the end of the 117th Congress (2021-2022), the Act would have established a federal framework to govern the use of computational forensic software. The bill defined computational forensic software as 'software that relies on an automated or semiautomated computational process, including one derived from machine learning, statistics, or other data processing or artificial intelligence techniques, to process, analyze, or interpret evidence.' The framework would have contained various elements, including the following:

> – requirements for the establishment of testing standards and a testing program for computational forensic software,
>
> – requirements for the use of computational forensic software by federal law enforcement agencies and related entities (e.g., crime labs),
>
> – a ban on the use of trade secret evidentiary privilege to prevent federal criminal defendants from accessing evidence collected using computational forensic software or information about the software (e.g., source code), and
>
> – limits on the admissibility of evidence collected using computational forensic software.[105]

Arguments for adoption of this federal legislation included the advantages of interpretable, not black box, technologies: if interpretable information is accessible to judges, prosecution, and defense counsel, they can understand the results produced by the technologies and can, in turn, explain them to jurors and other stakeholders in the criminal justice system.[106] Most importantly, defendants and defense counsel who are able to understand how forensic technologies reach their conclusions could have contested them in a meaningful way if such findings were used as evidence against them.[107] As of March 2023, the bill had not been reintroduced in the 118th Congress (2023-2024).

---

[103] Coglianese and Ben Dor (n 22) 793.

[104] HR 2438 (117th Congress, 2021-2022).

[105] Quoted from the Bill Summary authored by CRS (Congressional Research Service). Both the Bill Summary and the text of the bill are available at <www.congress.gov/bill/117th-congress/house-bill/2438?r=3&s=6> accessed 31 March 2023.

[106] See Duke Government Relations, 'The Need for Transparency and Interpretability at the Intersection of AI and Criminal Justice' (22 November 2021): <https://governmentrelations.duke.edu/2021/11/22/the-need-for-transparency-and-interpretability-at-the-intersection-of-ai-and-criminal-justice> accessed 28 March 2023.

[107] Julie Pattison-Gordon, 'Justice-Focused Algorithms Need to Show Their Work, Experts Say' (12 May 2022) government technology <www.govtech.com/computing/justice-focused-algorithms-need-to-show-their-work-experts-say> accessed 28 March 2023. See also Brookings Institution, 'Forensic Algorithms:

Legislative activity in this area also takes place at the state and local levels. Idaho, for example, enacted legislation in 2019 that specifically addresses the transparency, accountability, and explainability of pretrial risk assessment tools. The law requires all information used to build or validate such tools to be open to public inspection; entitles parties to criminal cases in which the court has considered or an expert witness has relied upon such a tool to review all calculations and data used to calculate the defendant's risk score; and prohibits builders and users of pretrial risk assessment tools from asserting trade secret or other intellectual property protections to quash discovery of relevant information in criminal and civil cases.[108]

2.2 Normative instruments produced by the executive authorities of your country deal with AI-based systems for predictive justice

In February 2019, President Donald Trump promulgated Executive Order 13859, entitled 'Maintaining American Leadership in Artificial Intelligence.' The order[109] required the Office of Management and Budget to issue a memorandum to agencies urging them to 'consider ways to reduce barriers to the use of AI technologies in order to promote their innovative application while protecting civil liberties, privacy, American values, and United States economic and national security'.[110] In November 2019, in response to the order, the memorandum 'Guidance for Regulation of Artificial Intelligence Applications' was issued for the heads of executive departments and agencies.[111] While not specifically focused on AI-based systems for predictive justice, the memorandum encouraged agencies to coordinate with each other 'to ensure consistency and predictability of AI-related policies that advance American innovation and adoption of AI' and reminded them of the need appropriately to protect 'privacy, civil liberties, national security, and American values" and to allow 'sector- and application-specific approaches'.[112]

The Future of Technology in the US Legal System' (*Brookings*, 12 May 2022): <www.brookings.edu/events/forensic-algorithms-the-future-of-technology-in-the-us-legal-system> accessed 28 March 2023; Doug Austin, 'The Justice in Forensic Algorithms Act: Legal Technology Trends' (*eDiscovery Today*, 26 May 2022): https://ediscoverytoday.com/2022/05/26/the-justice-in-forensic-algorithms-act-legal-technology-trends> accessed 28 March 2023.

[108] Idaho Code § 19-1910 (2022). For additional examples of proposed and enacted legislation at the state and local levels (validation study requirements, transparency in how prosecutors use risk assessments, AI task forces and commissions, etc.), see 'Liberty at Risk' (n 68).

[109] An executive order is a declaration by the president that has the force of law. Executive orders do not require any action by Congress to take effect, and they cannot be overturned by Congress. See Legal Information Institute, 'Executive Order' (*Legal Information Institute*): <www.law.cornell.edu/wex/executive_order> accessed 28 March 2023.

[110] Exec Order No 13859 of 11 February 2019, Maintaining American Leadership in Artificial Intelligence, 84 Fed Reg 3967 (14 February 2019).

[111] Russell T Vought, Memorandum for the Heads of Executive Departments and Agencies, *Guidance for Regulation of Artificial Intelligence Applications* (17 November 2020): <https://trumpwhitehouse.archives.gov/wp-content/uploads/2020/11/M-21-06.pdf> accessed 28 March 2023.

[112] Vought (n 111).

2.3 Soft law sources concerning predictive justice

In October 2022, the White House Office of Science and Technology Policy published a document entitled 'Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People'. The document, a white paper intended to 'support the development of policies and practices that protect civil rights and promote democratic values in the building, deployment, and governance of automated systems', is non-binding and does not constitute US government policy.[113] As stated in the title, it is a blueprint rather than an actual AI bill of rights. Its five principles state that people should be protected from automated systems that are unsafe or ineffective; they should be protected from algorithmic discrimination; they should enjoy data privacy; they should be notified when an automated system is being used, and explanations of outcomes should be provided; and they should, where appropriate, be able to opt out from automated systems in favor of a human alternative. While not specifically targeting predictive justice, the blueprint calls for 'enhanced protections and restrictions for data and interferences related to sensitive domains', including criminal justice; furthermore, automated systems intended for use within sensitive domains such as criminal justice should be 'tailored to the purpose, provide meaningful access for oversight, include training for any people interacting with the system, and incorporate human consideration for adverse or high-risk decisions'.[114]

Another soft-law source, the Model Penal Code,[115] prominently endorsed the consideration of risk in the sentencing process in its 2017 revision (MPC-S).[116] Once risks and needs processes developed by the sentencing commission[117] – including, presumably, those based on AI – prove to be sufficiently reliable, they may be incorporated into the sentencing guidelines:

> MPC-S § 6B.09. Evidence-Based Sentencing; Offender Treatment Needs and Risk of Reoffending.

[113] The White House, Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People, 2 (October 2022): <www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf> accessed 29 March 2023.

[114] White House (n 113) 6-7.

[115] The Model Penal Code, first promulgated in 1962, is a model code assembled by the American Legal Institute. Am Law Inst, '*Model Penal Code: Official Draft and Explanatory Notes*' (The American Law Institute, 1985). The part on sentencing has been revised in light of the changes in sentencing philosophy and practice that have taken place in the last half century. The project was approved in 2017; publication of the official text is expected in 2023. Am Law Inst, 'Model Penal Code: Sentencing' <www.ali.org/projects/show/sentencing> accessed 29 March 2023.

[116] Am Law Inst, Model Penal Code: Sentencing, Proposed Final Draft § 6B.09(3) (Proposed Final Draft, 2017): <https://robinainstitute.umn.edu/publications/model-penal-code-sentencing-proposed-final-draft-approved-may-2017> accessed 28 March 2023.

[117] The MPC-S recommends that all American jurisdictions establish a permanent sentencing commission as an essential agency of the criminal justice system. See MPC-S § 6A.

(1) The sentencing commission shall develop instruments or processes to assess the needs of offenders for rehabilitative treatment, and to assist the courts in judging the amenability of individual offenders to specific rehabilitative programs. When these instruments or processes prove sufficiently reliable, the commission may incorporate them into the sentencing guidelines.

(2) The commission shall develop actuarial instruments or processes, supported by current and ongoing recidivism research, that will estimate the relative risks that individual offenders pose to public safety through their future criminal conduct. When these instruments or processes prove sufficiently reliable, the commission may incorporate them into the sentencing guidelines.

(3) The commission shall develop actuarial instruments or processes to identify offenders who present an unusually low risk to public safety, but who are subject to a 32 presumptive or mandatory sentence of imprisonment under the laws or guidelines of the state. When accurate identifications of this kind are reasonably feasible, for cases in which the offender is projected to be an unusually low-risk offender, the sentencing court shall have discretion to impose a community sanction rather than a prison term, or a shorter prison term than indicated in statute or guidelines. The sentencing guidelines shall provide that such decisions are not departures from the sentencing guidelines.

2.4 Case law that addresses AI-based systems used for predictive justice: Criminal courts

The courts have only just begun to grapple with the legal implications of the use of algorithmic risk assessment tools in sentencing. Several such cases have been litigated in recent years in the United States. While the focus has not been on the use of AI, the holdings would seem to be applicable in the context of AI-based systems as well. The most prominent of these cases is *State v Loomis*. After a short introduction to the *Loomis* case, three additional cases (*Malenchik v State, State v Rogers*, and *State v Walls*) will be introduced. The section will end with a summary of the discussion.

2.4.1 State v Loomis

Defendant Loomis pleaded guilty in Wisconsin state court to charges relating to his involvement in a drive-by shooting.[118] He challenged the state's use of the risk assessment portion of the COMPAS report at sentencing. In determining his sentence, the court relied in part on the fact that Loomis had been 'identified, through the COMPAS assessment, as an individual who is at high risk to the community'. Loomis argued that the use of the COMPAS risk assessment violated his right to due process for three reasons: first,

[118] *State v Loomis*, 881 NW2d 749 (Wis 2016), cert denied, *Loomis v Wisconsin*, 137 S Ct 2290 (June 26, 2017). For a detailed summary of the *Loomis* case, see Recent Cases, 'Criminal Law - Sentencing Guidelines - Wisconsin Supreme Court Requires Warning before Use of Algorithmic Risk Assessments in Sentencing - *State v Loomis*, 881 NW2d 749 (Wis 2016)' (2017) 130 Harv L Rev 1530.

it violated his right to be sentenced on the basis of accurate information; second, it violated his right to an individualized sentence; and third, it improperly used gendered assessments in sentencing.

In its 2016 holding, the Wisconsin Supreme Court rejected all of Loomis's due process challenges: First, the variables used by the COMPAS algorithms were publicly available and the outcome of the risk assessment was based entirely on Loomis's answers to the questions or on publicly available information. Due process was satisfied because Loomis had 'the opportunity to verify that the questions and answers listed on the COMPAS report were accurate'. Second, while the COMPAS assessment did involve group data, the assessment was only one of multiple factors considered by the sentencing court so that Loomis received an individualized sentence.[119] Third, COMPAS's use of gender in calculating risk scores did not violate any due process rights since its use simply accounted for differences in recidivism rates between men and women; also, there was no proof that the court actually relied on gender as a factor in sentencing.

Despite denying Loomis's claims, the court expressly recognized the finding that risk assessment tools may not perform as well for non-whites as for whites. It also pointed out that the accuracy of such tools, without constant re-norming, is short-lived. As a result, it established the requirement that all Presentence Investigation Reports containing a COMPAS risk assessment inform the sentencing court of the following cautions regarding the risk assessment's accuracy:

> - The proprietary nature of COMPAS has been invoked to prevent disclosure of information relating to how factors are weighed or how risk scores are determined.
>
> - Because COMPAS risk assessment scores are based on group data, they are able to identify groups of high-risk offenders –not a particular high-risk individual.
>
> - Some studies of COMPAS risk assessment scores have raised questions about whether they disproportionately classify minority offenders as having a higher risk of recidivism
>
> - A COMPAS risk assessment compares defendants to a national sample, but no cross-validation study for a Wisconsin population has yet been completed. Risk assessment tools must be constantly monitored and re-normed for accuracy due to changing populations and subpopulations.
>
> - COMPAS was not developed for use at sentencing, but was intended for use by the Department of Corrections in making determinations regarding treatment, supervision, and parole.[120]

[119] *State v Loomis*, 881 NW2d 749, 765 (Wis 2016).
[120] *State v Loomis*, 881 NW2d 749, 769-770 (Wis 2016).

In sum, the *Loomis* opinion 'essentially implemented a mandatory disclaimer on the practice of using a COMPAS risk assessment at sentencing'.[121] It also stressed that risk scores may not be used as the sole determinative factor in sentencing. Emphasis of this point was expressed in a concurring opinion: 'consideration of COMPAS is permissible; reliance on COMPAS for the sentence imposed is not permissible.'[122] Loomis appealed the decision to the United States Supreme Court, but the Court declined to hear the case.[123]

2.4.2 Malenchik v State

Defendant Malenchik challenged the trial court's use of the results of two risk assessment tests at sentencing, both of which indicated that Malenchik was at high risk of recidivism.[124] In its 2010 holding that a trial court can properly 'supplement and enhance' its evaluation of the evidence before it at sentencing by considering assessment tool scores, the Indiana Supreme Court stressed that the sentence imposed by the trial court was not based solely on the risk assessments but that other factors had also been considered (eg, the defendant's prior criminal history, unwillingness to change his behavior, and refusal to accept responsibility for his actions); furthermore, it pointed out that the trial court had not relied on either test as an independent aggravating factor.

2.4.3 State v Rogers

This case addressed the question of whether a defendant was entitled to reconsideration of a sentence if the sentence was imposed without the use of a risk assessment instrument.[125] While West Virginia's highest court in 2015 denied the motion on procedural grounds, a concurring opinion sought to clarify the role of risk and needs assessments in relation to sentencing: accordingly, a risk and needs assessment 'is merely a tool that may be used by [trial court] judges during sentencing' and '[trial court] judges are not required to consider or use any of the information contained in [such an] assessment.'[126]

2.4.4 State v Walls

Defendant Walls' sentence reflected the risk assessment report that deemed him a high-risk, high-needs probation candidate, but the sentencing court refused to make the report available to defense counsel.[127] Walls challenged the sentence, arguing that he had a statutory and constitutional right to review and verify the question, answers, and scoring

---

[121] Chris Miller, 'The Prospects of Constitutional Challenges to COMPAS Risk Assessment (26 April 2021)': <www.americanbar.org/groups/litigation/committees/privacy-data-security/articles/2021/constitutional-challenges-compas-risk-assessment> accessed 28 March 2023.
[122] *State v Loomis*, 881 NW2d at 774 (Roggensack, CJ, concurring).
[123] *Loomis v Wisconsin*, 137 S Ct 2290 (26 June 2017).
[124] *Malenchik v State*, 928 NE2d 564 (Ind 2010).
[125] *State v Rogers*, No 14-0373, 2015 W Va LEXIS 3, 2015 WL 869323 (W Va, 9 January, 2015) (memorandum decision).
[126] ibid.
[127] *State v Walls*, 2017 Kan App Unpub LEXIS 487; 396 P3d 1261 (Kan App 2017).

decisions contained in the report. In 2017, the Kansas Court of Appeals found in his favor: Depriving him of the report 'necessarily denied him the opportunity to challenge the accuracy of the information upon which the court was required to rely in determining the conditions of his probation'. Since a defendant has a right to an 'effective opportunity to rebut the allegations likely to affect the sentence', the sentencing court's decision to deny him access to the output of the risk assessment tool on which it had relied in setting his sentence violated Wall's right to procedural due process.

2.4.5 Summary

Many legal scholars point out with approval the fact that courts 'appear to have taken pains to emphasize that [algorithmic assessment] tools only serve as one of multiple factors that a judge takes into account in reaching a decision'.[128] One law professor, for example, has stated that 'it would be a dark future if computer algorithms ever replaced a judge's sentencing decision' and that she 'can't imagine that a risk tool alone could produce just verdicts'. In her opinion, 'the judicial function can't be outsourced to a math problem.'[129] In contrast, another law professor sees this differently. In his view, the results of well-constructed risk assessment instruments are superior to lay judgments and should be given presumptive effect. 'Unfortunately', he writes, 'that rarely occurs'; instead, judges see the results of risk assessment instruments as mere tools and themselves as the definitive answer. This scholar views critically the holdings of judicial decisions (such as *Loomis* and *Malenchik*) according to which 'the results of a [risk assessment instrument] are but one factor to consider and should not be dispositive.'[130] He argues that:

> Judges and parole boards are clearly the ultimate decision-makers about offender risk. But they should be aware that evaluator, judicial, and parole board adjustments to [a risk assessment instrument] usually do not improve on the actuarial assessment. In fact, consistent with the studies comparing actuarial and clinical judgment, several studies find that professional "overrides" of [a risk assessment instrument's] risk estimate, whether by judges, probation officers, or other correctional professionals, decrease accuracy in predicting offending.[131]

While underscoring the superiority of risk assessment systems, the author also advocates for transparency, specifically, for risk algorithms to be made available for evaluation so as to enable defendants to engage in meaningful challenges to the results of risk assessments. Transparency is still suboptimal, both with regard to COMPAS – '[T]he company that produces the COMPAS refuses to reveal its algorithm or the weights assigned to risk

---

[128] Coglianese and Ben Dor (n 22) 811. See also Bagaric and others (n 24) 134.

[129] Joe Forward, The Loomis Case: The Use of Proprietary Algorithms at Sentencing, The InsideTrack (July 19, 2017): www.wisbar.org/newspublications/insidetrack/pages/article.aspx?Volume=9&ArticleID=25730 (quoting Prof. Cecelia Klingele).

[130] Christopher Slobogin, 'Preventive Justice: How Algorithms, Parole Boards, and Limiting Retributivism Could End Mass Incarceration' (2021) 56 Wake Forest L Rev 97, 138.

[131] Slobogin (n 130) 138.

factors, claiming trade secret protection.' – and the purportedly publicly developed PATTERN – 'Congress required that the PATTERN be made public, but did not require that the validation procedure that led to development of the instrument nor the data underlying it be disclosed.'[132] And he points out that 'the integration of sophisticated machine learning into [risk assessment instrument] construction could make matters worse.'[133]

2.5 Case law that addresses AI-based systems used for predictive justice: Civil courts

Various aspects of risk assessment tools used for predictive justice have claimed the attention of civil courts. Two such cases will be mentioned here. The first, *Henderson v Stensberg*, raised equal protection claims involving the use of COMPAS in a parole decision. The second, *Rodgers v Christie*, raised products liability claims regarding a risk assessment tool. Neither claim was successful.

2.5.1 Henderson v Stensberg

Plaintiff Henderson, incarcerated in Wisconsin, was denied parole in 2015.[134] He argued that prison officials discriminated against him and other Black prisoners by using COMPAS to assess their suitability for parole. Among other things, he brought Fourteenth Amendment equal protection claims in federal court against those prison officials as well as against the company that developed COMPAS. The judge granted summary judgment to the defendants. He expressly stated that Henderson's equal protection claims were not foreclosed and acknowledged that 'there is growing concern that risk-assessment algorithms unfairly disadvantage Black offenders.'[135] But he granted summary judgment because 'Henderson's recidivism score was the lowest possible'; he could not 'show that his COMPAS recidivism score was the reason he was denied parole'; and he thus 'failed to adduce admissible evidence that he was harmed by his COMPAS assessment or that he was denied parole for a discriminatory reason'.[136]

2.5.2 Rodgers v Christie

In 2017, Christian Rodgers[137] was murdered, allegedly by a man who, days before, had been granted pretrial release by a state court due to a decision informed in part by the court's use of a risk estimation tool (the Public Safety Assessment, PSA).[138] The victim's

---

[132] Slobogin (n 130) 164.

[133] Slobogin (n 130) 164.

[134] *Henderson v Stensberg*, 2021 US Dist LEXIS 58010 (WD Wis, 26 March 2021).

[135] *Henderson v Stensberg*, 2021 US Dist LEXIS 58010 at *2.

[136] *Henderson v Stensberg*, 2021 US Dist LEXIS 58010 at *2, *18.

[137] *Rodgers v Christie*, 795 F App'x 878 (3d Cir 2020) (note that the disposition of this case is not an opinion of the full Court and does not constitute biding precedent).

[138] While Coglianese and Ben Dor refer to the Public Safety Assessment as a non-learning algorithmic tool (see Coglianese and Ben Dor (n 22) 803), the outcome of the case is nevertheless of interest in the context of this study.

mother brought products liability claims in federal court against the foundation responsible for developing the PSA, alleging that the tool was designed in a defective manner. The Third Circuit, ruling in 2020, affirmed dismissal of the case, however, holding that the PSA is not a 'product' pursuant to the New Jersey Products Liability Act.[139]

2.5.3 Summary

As of yet, there is not much relevant legal scholarship assessing these rulings: commentators have yet to weigh in on the *Henderson* opinion of March 2021. As for the *Rodgers* decision, it was described by a law professor in 2021 as the only case to date involving accusations that the PSA tool harmed a third party not involved in a criminal matter. The scholar argued that both the *Loomis* and the *Henderson* decisions show that the lack of transparency of algorithms restricts the ability of plaintiffs to be heard and to prepare plausible causes of action.[140]

2.6 Laws governing reliability, impartiality, equality, and adaptability of AI-based predictive justice

In the United States, reliability, impartiality, equality, and adaptability are not specifically addressed by federal legislation for the context of AI-based predictive justice (risk assessment). These or related issues have, however, been addressed indirectly in a recent executive order, in a bill introduced in a state legislature (Massachusetts), and in model legislation (EPIC).

In May 2022, President Joe Biden issued an executive order entitled 'Advancing Effective Accountable Policing and Criminal Justice Practices to Enhance Public Trust and Public Safety'.[141] The order directs the National Academy of Sciences to conduct and publish a study of – among other things – predictive algorithms, with a particular focus on the use of such algorithms by federal law enforcement agencies. The study must assess concerns in the areas of privacy, civil rights, civil liberties, accuracy, or disparate impact that arise in association with the use of such algorithms. Subsequently, the study will be used to make any necessary changes to Federal law enforcement practices.[142]

In February 2022, an 'Act Establishing a Commission on Automated Decision-Making by Government in the Commonwealth' was introduced in the Massachusetts legislature.[143]

---

[139] *Rodgers v Christie*, 795 F App'x 878 (3d Cir 2020) (note that the disposition of this case is not an opinion of the full Court and does not constitute biding precedent).

[140] Sonia M Gipson Rankin, 'Technological Tethereds: Potential Impact of Untrustworthy Artificial Intelligence in Criminal Justice Risk Assessment Instruments' (2021) 78 Wash & Lee L Rev 647, 705–706.

[141] Exec Order No 14074 of 25 May 2022, Advancing Effective, Accountable Policing and Criminal Justice Practices to Enhance Public Trust and Public Safety, 87 Fed Reg 32945 (31 May 2022).

[142] See The White House, 'Fact Sheet' (*The White House*, 25 May 2022) <www.whitehouse.gov/briefing-room/statements-releases/2022/05/25/fact-sheet-president-biden-to-sign-historic-executive-order-to-advance-effective-accountable-policing-and-strengthen-public-safety> accessed 29 March 2023. See also Exec Order No 14074 of 25 May 2022, s 13 (d).

[143] Bill S.2688: <https://malegislature.gov/Bills/192/S2688> accessed 29 March 2023.

Virtually the same bill, Bill S.33, was introduced in early 2023.[144] The act would establish a commission to study and make recommendations related to the use in Massachusetts of automated decision systems that may affect human welfare, including the legal rights and privileges of individuals. It describes the responsibilities and composition of the commission and lists the reporting requirements with which it would have to comply.

In particular, the commission would undertake a survey of

> (b) the training specific Massachusetts offices provide to individuals using automated decision systems, and the procedures for enforcing the principles, policies, and guidelines regarding their use;
>
> (c) the manner by which Massachusetts offices validate and test the automated decision systems they use, and the manner by which they evaluate those systems on an ongoing basis …;
>
> (d) matters related to the transparency, explicability, auditability, and accountability of automated decision systems in use in Massachusetts offices …;
>
> (e) the manner and extent to which Massachusetts offices make the automated decision systems they use available to external review …; and
>
> (f) procedures and policies in place to protect the due process rights of individuals directly affected by Massachusetts offices' use of automated decision systems, including but not limited to public disclosure and transparency procedures.[145]

The commission would also consult with experts in the fields of machine learning, algorithmic bias, algorithmic auditing, and civil and human rights[146] and would examine research related to the use of automated decision systems that directly or indirectly result in disparate outcomes for individuals or communities based on an identified group characteristic.[147]

Sometime before 2020, EPIC developed a model law for state AI commissions.[148] In its 'findings' section, the model law proposes that the enacting legislature find that the state has begun to deploy AI and other automated decision systems in numerous areas, including in the area of criminal law; that there is an inherent risk of bias and inaccuracy in the use of these technologies; that there is limited public knowledge about the systems; and that existing regulation of automated decision systems is insufficient.[149] The model law calls for the creation of an 18-person commission to carry out a two-phased study.[150]

---

144 Bill S.33: <https://malegislature.gov/Bills/193/S33> accessed 29 March 2023.
145 Bill S.33 Sec. 11. (b)(i)(b)-(f).
146 Bill S.33 Sec. 11. (b)(ii).
147 Bill S.33 Sec. 11. (b)(iii).
148 EPIC, 'Model State Artificial Intelligence Commission Law': <https://archive.epic.org/EPIC-Model-State-AI-Commission-Bill.pdf> accessed 28 March 2023.
149 Model State (n 148) s 2.
150 Model State (n 148) s 4(b).

Phase one would involve the reviewing and cataloguing of how algorithms or other automated decision systems are being used by the state, including:

> the identity of the developer and pertinent contract terms between the state and the developer; any state bodies or subdivisions using automated decision systems; the inputs used; the source of the inputs used; the purposes for which such systems are used; the validation policies, the logic of the automated decision system; the data maintenance and deletion policies; and the potential harms that could arise from the use of the system and how those risks are currently addressed.[151]

In a second phase, the commission would propose recommendations regarding, among other things, minimum technological standards for all automated decision systems; uniform data security provisions; procedures by which individuals affected by a decision made by an automated decision system used by the state could seek information concerning that decision; procedures by which individuals could seek human review of automated decisions made about them; procedures to ensure that automated decision system do not reflect unfair bias or make impermissible discriminatory decisions; procedures to ensure that such systems are adequately evaluated; procedures to ensure the accuracy, reliability, and validity of decisions made by such systems; and procedures to establish data provenance.[152]

2.7 Restrictions on marketing AI-based systems for predictive justice

There is no federal law governing the marketing of AI-based systems for predictive justice nor is there a federal law that imposes technological requirements on producers of AI-based systems for predictive justice. There is no federal law that requires producers of AI-based systems for predictive justice to consult criminal justice professionals regarding the design of the software, and there is no federal law that requires producers of AI-based systems for predictive justice to regularly monitor and update the software. Finally, there is no federal law governing the certification or labelling of AI-based systems for predictive justice.

In *State v Loomis*, the Wisconsin Supreme Court recognized in 2016 that the accuracy of risk assessment tools, without constant re-norming, is short-lived. In its holding, the court required all presentence investigation reports submitted to sentencing judges that contain a COMPAS risk assessment to include a 'written advisement'. The purpose of the advisement was to inform the sentencing court (among other things) that risk assessment tools 'must be constantly monitored and re-normed for accuracy due to changing populations and subpopulations'.[153] It should be noted that the holding is binding only in Wisconsin and that it does not require the monitoring and re-norming of risk assessment tools; it requires only that the advisement regarding the accuracy of risk assessment

[151] Model State (n 148) s 4(g)(1).
[152] Model State (n 148) s 4(g)(2).
[153] *State v Loomis*, 881 NW2d 749, 769 (Wis 2016).

tools be included in presentence investigation reports that contain a COMPAS risk assessment.

2.8 Training of professionals who rely on AI-based systems

Of the many jurisdictions that use risk assessment instruments (some of which are – or may soon be – based on AI), very few train judges, lawyers, and correctional officials in their use.[154] As far as the use of risk assessment instruments by sentencing judges is concerned, their discretion continues to play an important role, and very little information is available about how judges actually use these risk assessments in practice.[155]

2.9 Transparency and the technological functioning of AI-based systems

There is no federal law guaranteeing the transparency of the technological functioning of AI-based systems for predictive justice. Generally speaking, companies are allowed to claim their technology is a trade secret and can refuse to be transparent about how their product works. In *State v Loomis*, for example, the defendant requested access to information concerning the inner workings of the COMPAS tool, but the Wisconsin Supreme Court denied the request: the court permitted the proprietary nature of the COMPAS tool – as asserted by its developer, Northpointe, Inc. – to prevent disclosure of information about how factors are weighed or how risk scores are determined.[156]

There is, however, a statute in the state of Idaho,[157] enacted in 2019, that addresses the transparency, accountability, and explainability of pretrial risk assessment tools. Pursuant to the statute, all pretrial risk assessment tools must be transparent; information used to build or validate such tools must be open to public inspection; parties to criminal cases in which such a tool has been relied upon are entitled to review calculations and data used to calculate the defendant's risk score; and builders and users of such tools may not assert trade secret or other intellectual property protections in order to quash discovery of information used in the development or validation of such tools.[158]

---

[154] Slobogin (n 130) 168.

[155] Garrett and Monahan (n 26) 43. See discussion at 1.3. above.

[156] *State v Loomis*, 881 NW2d at 761, 769 (Wis 2016).

[157] Idaho Code § 19-1910 (2022).

[158] Idaho Code § 19-1910 (2022). Pretrial Risk Assessment Tools.
'(1) All pretrial risk assessment tools shall be transparent, and:
(a) All documents, data, records, and information used by the builder to build or validate the pretrial risk assessment tool and ongoing documents, data, records, and written policies outlining the usage and validation of the pretrial risk assessment tool shall be open to public inspection, auditing, and testing;
(b) A party to a criminal case wherein a court has considered, or an expert witness has relied upon, a pretrial risk assessment tool shall be entitled to review all calculations and data used to calculate the defendant's own risk score; and
(c) No builder or user of a pretrial risk assessment tool may assert trade secret or other intellectual property protections in order to quash discovery of the materials described in paragraph (a) of this subsection in a criminal or civil case.

In contrast to the transparency required by statute in Idaho at the pretrial stage, Massachusetts does not require such transparency at the parole stage. According to an amicus brief filed by EPIC[159] in the case of *Jose Rodriguez v Massachusetts Parole Board*,[160] parole applicants in Massachusetts are given only a redacted version of the report provided by the predictive analytical tool in use in that state (LS/CMI[161]). They are not given information about the sources of data that went into their assessment nor are they given information about the logic of the tool or about the role the report played in their parole decisions. Furthermore, they and the public are unable to access even blank scoresheets, scoring guides, training manuals or validation studies. EPIC argued that Massachusetts' lack of transparency concerning use of its predictive analytical tool prevents the public from fully understanding the tool's accuracy and potential for bias and prevents inmates from understanding how the tool decided their recidivism risk and whether those decisions were accurate. While EPIC stated that the predictive analytical tool in use in Massachusetts 'seems' to fall into the category of 'checklist-type tools' rather than the more advanced category of tools that use machine learning, the argumentation in favor of transparency is equally if not more applicable to AI-based tools.

2.10 Transparency and the use of AI-based systems for predictive justice

There are no federal rules specifically governing the right of affected individuals to be informed about the use of AI-based systems for predictive justice. Other general rules concerning the right to be informed may, however, apply. In New York State, for example, inmates whose application for parole are denied have a statutory right to be informed in writing of the factors and reasons for such denial, and 'such reasons shall be given in detail and not in conclusory terms.'[162] This right will not always suffice, however, to enable inmates to challenge all the factors that have contributed to their parole denials. Take, for example, inmate Glenn Rodriguez, who was granted parole in 2017 but whose previous parole application was denied on the basis of a COMPAS 'high risk' ranking:

---

(2) For purposes of this section, "pretrial risk assessment tool" means a pretrial process that creates or scores particular factors in order to estimate a person's level of risk to fail to appear in court, risk to commit a new crime, or risk posed to the community in order to make recommendations as to bail or conditions of release based on such risk, whether made on an individualized basis or based on a grid or schedule.'

[159] Benjamin Winters, 'Brief of Amicus Curiae,' *Rodriguez v Massachusetts Parole Board*, SJC-13197 (*Elec Priv Info Ctr*, 14 February 2022): <https://epic.org/documents/rodriguez-v-massachusetts-parole-board> accessed 28 March 2023.

[160] SJC-13197, available at <www.ma-appellatecourts.org/docket/SJC-13197> accessed 30 March 2023. In September 2022, the Massachusetts Supreme Judicial Court affirmed the judgment of the superior court in favor of the parole board, holding that the superior court was correct in affirming the board's decision to deny Rodriguez release on parole. Slip opinion available at <https://cases.justia.com/massachusetts/supreme-court/2022-sjc-13197.pdf?ts=1662552216> accessed 30 march 2023.

[161] Level of Service/Case Management Inventory.

[162] NY Exec Law § 259-i(2)(a) (McKinney 2018).

> When inmate Glenn Rodriguez was denied parole, he had a statutory right to be informed in writing of the "factors and reasons" for the denial." Rodriguez filed a grievance showing that there was an error in one of the inputs used to generate his risk assessment score. The tool relies on manual inputs from surveys filled out by a human evaluator. In Rodriguez's case, the evaluator had checked "yes" where he should have checked "no" in one survey response. Rodriguez knew that when another inmate had received a reassessment to correct the same error, that person's final risk score dropped significantly. But Rodriguez could not prove that the error had any significant effect in his own case because the weights of the input variables are alleged trade secrets. Ultimately, he was unable to convince anyone to correct the mistake and had to return to the parole board six months later with the same erroneous score.[163]

While the problems Glenn Rodriguez encountered involved interactions between human error and trade secrets, the combination of AI and trade secrets, it would seem, would pose inmates with even more challenging situations. (See also discussion of Jose Rodriguez at 2.9. above.)

3 General principles of law

3.1. Right to equality (right to non-discrimination) with regard to AI-based systems used for predictive justice

In the context of AI-based systems used for predictive justice, there is lively discussion in the United States about equality and non-discrimination – issues of interest from many perspectives, including that of constitutional law. This discussion is taking place both in the popular media[164] and in the academic community.[165] Authors (primarily law faculty members) of a law review article published in 2022, for example, who point to the wide acceptance of the principle of equality before the law as a fundamental tenet of justice,[166] recognize that the use of algorithms in the criminal justice system has both positive and negative effects on the realization of this principle. Citing an opinion piece written by

---

[163] Rebecca Wexler, 'Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System' (2018) 70 Stan L Rev 1343, 1371 (footnotes omitted). See also Rebecca Wexler, 'Code of Silence: How Private Companies Hide Flaws in the Software That Governments Use to Decide Who Goes to Prison and Who Gets Out' (June/July/August 2017) Washington Monthly.

[164] See, eg, Thadaney Israni (n 14); Liptak (n 38); Jens Ludwig and Cass R. Sunstein, 'Discrimination in the Age of Algorithms' The Boston Globe (Boston, 24 September 2019) A8.

[165] See, eg, Itay Ravid and Amit Haim, 'Progressive Algorithms' (2022) 12 UC Irvine L Rev 527; Note, 'Beyond Intent: Establishing Discriminatory Purpose in Algorithmic Risk Assessment' (2021) 134 Harv L Rev 1760; Vincent M. Southerland, 'The Intersection of Race and Algorithmic Tools in the Criminal Legal System' (2021) 80 Md L Rev 487; Aziz Z. Huq, 'Constitutional Rights in the Machine-Learning State' (2020) 105 Cornell L Rev 1875; Huq, 'Racial Equity' (n 20); Michael Simon and others, 'Lola v Skadden and the Automation of the Legal Profession' (2018) 20 Yale JL & Tech 234.

[166] Bagaric and others (n 24) 100; Michael Simon and others, 'Lola v Skadden and the Automation of the Legal Profession' (2018) 20 Yale JL & Tech 234.

two professors for the Boston Globe in 2019,[167] these authors opine that 'eliminating algorithms – or reducing their use – would cause "more, not less, discrimination.' And, citing an article published in a nonprofit news organization in 2019, they state that 'the bias present in algorithms presents less of an obstacle than human bias because it can more easily "be observed, studied, and corrected in ways that human bias cannot."'[168] On the other hand, they recognize that the use of algorithms trained with '"past biased data" are likely to recreate the same biases in their decision-making processes, further exacerbating discrimination and unfairness'.[169] In the reform section of their article, they propose that predictive systems 'be developed carefully with a focus on preventing the operation of factors that lead to indirect discrimination' in order to 'minimize the potential for race and other immutable factors to influence the outcomes of risk assessment algorithms'.[170]

In a 2021 law review article, the author, a law professor, defended the use of risk assessment instruments against wide-ranging attacks on accuracy and fairness grounds. His conclusion against claims of egalitarian injustice was that 'with a few caveats, such instruments are not violative of equal protection if they provide relevant and probative results.'[171]

Finally, specifically in the machine-learning context, another law professor reexamined questions of intent and classification – issues at the heart of the constitutional jurisprudence of the Equal Protection Clause and federal antidiscrimination statutes. In a 2020 law review article, he suggested that 'the equality concerns commonly raised by algorithmic systems in practice are better conceptualized in terms of their impact on pernicious social stratification.'[172] And in another 2020 journal article, the co-authors, both law professors, saw the use of machine learning and other forms of AI in the adjudication of criminal proceedings as a 'context in which questions of equity and fairness receive heightened attention'.[173]

3.2. AI-based systems and judicial independence

Discussion in the United States on the effects of AI-based systems on judicial independence is not widespread. There are no means or methods designed specifically to guarantee judicial independence in the context of AI usage.

---

[167] Bagaric and others (n 24) 98, citing Ludwig and Sunstein (n 164).
[168] Bagaric and others (n 24) 98 (citing Matt Henry, 'Risk Assessment: Explained' (*The Appeal*, 25 March 2019): <https://theappeal.org/risk-assessment-explained> accessed 28 March 2023.
[169] Bagaric and others (n 24) 133.
[170] Bagaric and others (n 24) 135.
[171] Slobogin (n 130) 143.
[172] Huq, 'Constitutional Rights' (n 165) 1917.
[173] Carla L Reyes and Jeff Ward, 'Digging into Algorithms: Legal Ethics and Legal Access' (2020) 21 Nev LJ 325, 333.

3.3. Right of access to a human judge

There is discussion in the United States about whether and under what circumstances there should be a right of access to a human judge.[174] For example, a 2020 law review article, the author, a law professor, pointed to the holding of the Wisconsin Supreme Court in Loomis[175] when he wrote that 'American law is … making tentative moves toward a … right to a human decision.'[176] (Loomis held that a risk score generated by an algorithm cannot, as a matter of due process, 'be considered as the determinative factor in deciding whether the offender can be supervised safely and effectively in the community'.[177]) According to the author, '[The Loomis] decision precludes full automation of bail determinations. There must be a human judge in the loop.'[178] He also argued that there is no reason why it should not be possible to invoke the Sixth Amendment's right to a jury trial[179] to preclude the use of at least some forms of algorithmically generated inputs in criminal sentencing: 'Indeed, it would seem to follow a fortiori that a right precluding a jury's substitution with a judge would also block its displacement by a mere machine.'[180]

The author's own position in this discussion is to favor 'a right to a well-calibrated machine decision'[181] rather than a right to a decision taken by a human judge, in part because 'machines have the capacity to classify and predict with fewer errors than humans.'[182]

> Algorithmic technologies used by machine decisions are still in their infancy. Now, they can be flawed in many ways. It seems too early, however, to assume that human decisions will be globally superior to machine decisions such that a right to the former is warranted. Sometimes the opposite might be true. We should, therefore, at least consider the possibility that under certain circumstances a right to a well-calibrated machine decision might be the better option.[183]

---

[174] See, eg, Michael Simon and others, 'Lola v Skadden and the Automation of the Legal Profession' (2018) 20 Yale JL & Tech 234 (citing Joseph Weizenbaum, Computer Power and Human Reason: From Judgment to Calculation (WH Freeman and Co, 1976), for the argument that computers should not be permitted to make judicial decisions).

[175] *State v Loomis*, 881 NW2d 749 (Wis 2016). See discussion at 2.4. above.

[176] Aziz Z Huq, 'A Right to a Human Decision' (2020) 106 Va L Rev 611, 617.

[177] *State v Loomis*, 881 NW2d 749, 760 (Wis 2016).

[178] Huq, 'A Right to a Human Decision' (n 176) 617.

[179] US Const amend VI ('In all criminal prosecutions, the accused shall enjoy the right to a speedy and public trial, by an impartial jury of the state and district wherein the crime shall have been committed ... .').

[180] Huq, 'A Right to a Human Decision' (n 176) 617.

[181] Huq, 'A Right to a Human Decision' (n 176) 619, 686.

[182] Huq, 'A Right to a Human Decision' (n 176) 650.

[183] Huq, 'A Right to a Human Decision' (n 176) 687-688.

3.4. The presumption of innocence and the use of AI-based systems to establish the probability that a person is dangerous or is likely to reoffend

There is discussion in the United States about protecting the presumption of innocence when AI-based risk assessment tools are used to determine whether a person is dangerous or is likely to recidivate. In a 2020 law review article, for example, the author explains the role of the constitutional presumption of innocence in the pretrial phase, when risk assessment instruments are used to help decide whether a detainee should be incarcerated or released. She examines the implications of using machine learning to develop the instruments used in this phase of the criminal justice system and asks whether AI-based tools represent a threat to the system. After comparing seven such instruments, two of which (COMPAS and the Kleinberg et al. tool) have a machine-learning component, she concludes that there are 'more similarities than differences' between tools using machine learning and those using regression analysis and that 'adding a machine learning aspect to risk assessment tools will not worsen the outcome, and in many cases may improve it.'[184]

In contrast, in a 2021 journal article on the cascading effect of algorithmic bias in risk assessments, the author came to the following conclusion: 'To the extent the U.S. justice system is predicated on the presumption of innocence, the use of algorithmic tools to predict the probability of future crime in deciding the length of one's sentence is a contradiction.'[185]

3.5. Fair trial rights and AI-based systems used for predictive justice

There is widespread discussion in the United States about due process/fair trial issues in the context of AI-based systems used for predictive justice. A 2021 note in the Harvard Law Review, for example, cited case law rejecting a defendant's due process claim against the use of the COMPAS risk assessment systems[186] and argued that with algorithmic systems 'concerns of bias … can infringe upon the individual liberty interest in a fair trial.'[187]

As far as the right to contest decisions made by AI is concerned, the authors of a comprehensive law review article published in 2021 showed that the issue of what rights to an appeal, if any, people should have when they are subjected to decision-making by artificial intelligence is unclear. This lack of clarity exists even though 'the right to challenge decisions with significant effects is a core principle of the rule of law.'[188] Their article

[184] Doaa Abu Elyounes, 'Bail Or Jail? Judicial versus Algorithmic Decision-Making in the Pretrial System' (2020) 21 Colum Sci & Tech L Rev 376, 445.
[185] Tim O'Brien, 'Compounding Injustice: The Cascading Effect of Algorithmic Bias in Risk Assessments' (2021) 13 Geo JL & Mod Critical Race Persp 39, 80-81.
[186] Note, 'Beyond Intent' (n 165) 1762 (citing *People v Younglove*, No 341901, 2019 WL 846117, at *3 (Mich Ct App, 21 February 2019).
[187] Note, 'Beyond Intent' (n 165) 1771.
[188] Margot E Kaminski and Jennifer M Urban, 'The Right to Contest AI' (2021) 121 Colum L Rev 1957, 1959-1960.

reviews suggestions made by numerous legal experts over the past several years and concludes that 'while earlier scholars called for some kind of due process, the recent trend has been to favor systemic governance over the companies or government entities that build and use Al over establishing individual rights such as a right to contest.'[189]

3.6. Right to defense against algorithmic calculations

Robot lawyers are making inroads in the legal profession. Such intelligent machines represent a challenge to the existing liability regime. According to the author of a 2019 law review article, a law professor, human lawyers who fail to deliver competent legal services to their clients are subject to both ethical discipline and malpractice suits.[190] Indeed, 'their responsibility can extend to the actions of third parties' involved in the provision of legal services.[191] When lawyer robots make mistakes, however, the question of who should be held responsible and who should compensate injured clients is still an open one. The author points out that some clients, particularly sophisticated corporate clients, 'are likely to negotiate warranties and other protections into their engagements to shield themselves from any errors resulting from the use of artificial intelligence'.[192] In contrast, to these clients, ordinary individuals, he argues 'are not in a position to negotiate these protections or to assess the quality of the legal services they receive'.[193]

Regarding the hurdles defendants face if they seek meaningfully to challenge algorithmic assessments used to determine their sentences, we turn again to the Loomis case. Here, defendant Loomis challenged the system (COMPAS) that labeled him a high risk for recidivism. The Wisconsin Supreme Court denied his challenge on a number of grounds and, recognizing its protection by trade secret law, did not grant Loomis full access to the COMPAS algorithm. Co-authors of a journal article published in 2020, both law professors, stated the following:

> Despite the [Wisconsin] Supreme Court's decision, concern remains that denying access to the process by which sentencing and other impactful determinations are made represents a due process problem in and of itself. Those engaged in debates at the intersection of law and algorithms employ Loomis' experience and the court's response as a rallying cry for technologies' potential to inject additional inequity into the criminal justice system, rather than less. These concerns amplify as state and federal government institutions adopt technological tools in an increasing number of government-citizen interactions.[194]

---

[189] Kaminski and Urban (n 188) 1984.
[190] Milan Markovic, 'Rise of the Robot Lawyers' (2019) 61 Ariz L Rev 325, 343.
[191] Markovic (n 190) 343.
[192] Markovic (n 190) 344.
[193] Markovic (n 190) 344.
[194] Reyes and Ward (n 173) 328. See also Justin Snyder, 'RoboCourt: How Artificial Intelligence Can Help Pro Se Litigants and Create a "Fairer" Judiciary' (2022) 10 Ind JL & Soc Equal 200.

3.7. Principles of constitutional law and the use of AI-based systems for predictive justice

There is wide-ranging discussion in the United States about principles of constitutional law affected by the use of AI-based systems for predictive justice.[195] (A number of these principles – due process, equal protection, right to jury trial, etc. – are discussed elsewhere in this report.) In 2020, for example, a group of experts from Harvard (mathematics, economics, and law) published a comprehensive law review article analyzing the constitutional issues presented by the use of risk-assessment technologies – including those based on AI – in the criminal justice system. The issues addressed include the relationship between due process and certain proprietary algorithmic models and the challenges to existing equal protection jurisprudence posed by the discriminatory nature of risk assessment instruments. The article discusses possible ways to challenge the constitutionality of risk assessment technologies in state courts and concludes with suggestions for how to improve the technology and satisfy constitutional standards simultaneously.[196]

3.8. Privatization of criminal justice and equality of litigants

There is discussion about the privatization of aspects of criminal justice in the United States.[197] Privatization is particularly problematic in the context of sentencing: 'private developers play a significant part in sentencing determinations without being subject to traditional constitutional accountability mechanisms.'[198]

The question of the equality of litigants in the criminal justice system is also a topic of concern in the United States. The question has been asked whether increased reliance on AI will lead

> to one or more inequitable two-tiered systems. Some fear an eventual system with expensive – but superior – human lawyers and inexpensive – but inferior – AI driven legal assistance. Others fear almost the reverse problem: that AI will be superior to human lawyers but will be expensive and available only to large law firms and their wealthy clients. Still others fear that AI's impact will not overcome the status quo where some can afford legal services while others cannot.[199]

---

[195] See, eg, Krent and Rucker (n 17) (due process, ex post facto issues); Huq, 'Constitutional Rights' (n 165) (due process, privacy, equality); Andrea Nishi, 'Privatizing Sentencing: A Delegation Framework for Recidivism Risk Assessment' (2019) 119 Colum L Rev 1671 (due process, equal protection).

[196] Michael Brenner and others, 'Constitutional Dimensions of Predictive Algorithms in Criminal Justice' (2020) 55 Harv CR-CL L Rev 267.

[197] See, eg, Farhang Heydari, 'The Private Role in Public Safety' (2022) 90 Geo Wash L Rev 696.

[198] Nishi (n 195) 1688.

[199] Drew Simshaw, 'Access to AI Justice: Avoiding an Inequitable Two-Tiered System of Legal Services' (2022) 24 Yale JL & Tech 150, 156 (footnotes omitted).

And asymmetries may arise if law enforcement can access data possessed by private companies but investigators for the defense cannot.[200] In such cases, 'law enforcement but not defendants will benefit from deploying new algorithmic artificial intelligence and machine learning tools to search and analyze that data.'[201]

Selected literature

Abu Elyounes D, 'Bail Or Jail? Judicial versus Algorithmic Decision-Making in the Pretrial System' (2020) 21 Colum Sci & Tech L Rev 376

Am Law Inst, *'Model Penal Code: Official Draft and Explanatory Notes'*, The American Law Institute, 1985

– – 'Model Penal Code: Sentencing'

– – Model Penal Code: Sentencing, Proposed Final Draft, 2017

Angwin J, 'Make Algorithms Accountable' The NY Times (New York, 1 August 2016) A17

Angwin J and others, 'Machine Bias', *ProPublica,* 23 May 2016

Austin D, 'The Justice in Forensic Algorithms Act: Legal Technology Trends', *eDiscovery Today,* 26 May 2022

Bagaric M and others, 'The Solution to the Pervasive Bias and Discrimination in the Criminal Justice System: Transparent and Fair Artificial Intelligence' (2022) 59 Am Crim L Rev 95

Bagaric M, Hunter D, and Stobbs N, 'Erasing the Bias against Using Artificial Intelligence to Predict Future Criminality: Algorithms Are Color Blind and Never Tire' (2020) 88 U Cin L Rev 1037

Bagaric M and Wolf G, 'Sentencing by Computer: Enhancing Sentencing Transparency and Predictability and (Possibly) Bridging the Gap between Sentencing Knowledge and Practice' (2018) 25 Geo Mason L Rev 653

Berk R, 'An Impact Assessment of Machine Learning Risk Forecasts on Parole Board Decisions and Recidivism' (2017) 13 J Experimental Criminology 193

Brand J and Pishko J, 'Bail Reform: Explained', *The Appeal,* 14 June 2018

[200] 'Privacy Asymmetries: Access to Data in Criminal Defense Investigations' (2021) 68 UCLA L Rev 212, 248

[201] Wexler, 'Privacy Asymmetries' (n 200) 248.

Brennan T, Dieterich W, and Ehret B, 'Evaluating the Predictive Validity of the COMPAS Risk and Needs Assessment System' (2009) 36 Crim Just & Behavior 21

Brenner M and others, 'Constitutional Dimensions of Predictive Algorithms in Criminal Justice' (2020) 55 Harv CR-CL L Rev 267

Brookings Institution, 'Forensic Algorithms: The Future of Technology in the US Legal System', *Brookings,* 12 May 2022

Coglianese C and Ben Dor LM, 'AI in Adjudication and Administration' (2021) 86 Brook L Rev 791

Corbett-Davies S and others, 'A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased against Blacks. It's Actually Not That Clear', *The Washington Post,* Washington, D.C., 17 October 2016

Corey E, 'New Data Suggests Risk Assessment Tools have Little Impact on Pretrial Incarceration', *The Appeal,* 7 February 2020

Cyphert AB, 'Reprogramming Recidivism: The First Step Act and Algorithmic Prediction of Risk' (2020) 51 Seton Hall L Rev 331

Dershowitz AM, 'Visibility, Accountability and Discourse as Essential to Democracy: The Underlying Theme of Alan Dershowitz's Writing and Teaching' (2008) 71 Alb L Rev 731

Desmarais SL and Singh JP, Risk Assessment Instruments Validated and Implemented in Correctional Settings in the United States 2, *Council of State Governments Justice Center,* 27 March 2013

DiBenedetto R, 'Reducing Recidivism or Misclassifying Offenders: How Implementing Risk and Needs Assessment in the Federal Prison System Will Perpetuate Racial Bias' (2019) 27 JL & Pol'y 414

DOJ, 'Department of Justice Announces Enhancements to the Risk Assessment System and Updates on First Step Act Implementation', 15 January 2020

– – 'The First Step Act of 2018: Risk and Needs Assessment System – UPDATE', *Department of Justice,* January 2020

Donohue ME, 'A Replacement for Justitia's Scales: Machine Learning's Role in Sentencing' (2019) 32 Harv JL & Tech 657

Duke Government Relations, 'The Need for Transparency and Interpretability at the Intersection of AI and Criminal Justice', 22 November 2021

EPIC, 'About Us' < https://epic.org/about> accessed 29 March 2023

– – ‘AI in the Criminal Justice System’ <https://epic.org/issues/ai/ai-in-the-criminal-justice-system> accessed 28 March 2023

– – ‘Liberty at Risk: Pre-Trial Risk Assessment Tools in the US’ (September 2020) <https://epic.org/wp-content/uploads/2022/02/Liberty-At-Risk-Report-FALL-2020-UP-DATE.pdf> accessed 28 March 2023

– – ‘Model State Artificial Intelligence Commission Law’ <https://archive.epic.org/EPIC-Model-State-AI-Commission-Bill.pdf> accessed 28 March 2023

Flores AW, Bechtel K, and Lowenkamp CT, ‘False Positives, False Negatives, and False Analyses: A Rejoinder to Machine Bias: There’s Software Used Across the Country to Predict Future Criminals. And It’s Biased Against Blacks’ (2016) 80 Fed Probation 38

Forward J, ‘The Loomis Case: The Use of Proprietary Algorithms at Sentencing’ InsideTrack, 19 July 19 2017

Garrett B and Monahan J, ‘Assessing Risk: The Use of Risk Assessment in Sentencing’ (2019) 103 Judicature 42

Giarda R, ‘International: Artificial Intelligence in the Administration of Justice’, *LegalBytes*, January 2022

Gipson Rankin SM, ‘Technological Tethereds: Potential Impact of Untrustworthy Artificial Intelligence in Criminal Justice Risk Assessment Instruments’ (2021) 78 Wash & Lee L Rev 647

Glueck S, ‘Predictive Devices and the Individualization of Justice’ (1958) 23 Law & Contemp Probs 461

Hamilton M, ‘Algorithmic Risk Assessment: A Progressive Policy in Pretrial Release’ (2021) 57 Idaho L Rev 615

– – ‘Evaluating Algorithmic Risk Assessment’ (2021) 24 New Crim L Rev 156

Henry M, ‘Risk Assessment: Explained’, *The Appeal*, 25 March 2019

Heydari F, ‘The Private Role in Public Safety’ (2022) 90 Geo Wash L Rev 696

Huq AZ, ‘Racial Equity in Algorithmic Criminal Justice’ (2019) 68 Duke LJ 1043

– – ‘Constitutional Rights in the Machine-Learning State’ (2020) 105 Cornell L Rev 1875

– – ‘A Right to a Human Decision’ (2020) 106 Va L Rev 611

Hutson M, ‘Artificial Intelligence Prevails at Predicting Supreme Court Decisions’, *Science*, 2 May 2017

Kaminski ME and Urban JM, ‘The Right to Contest AI’ (2021) 121 Colum L Rev 1957

Kehl D, Guo P, and Kessler S, 'Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing' (2017)

Kranzberg M, 'Technology and History: "Kranzberg's Laws"' (1986) 27 Technology and Culture 544

Kraus RC, 'Artificial Intelligence Invades Appellate Practice: The Here, The Near, and The Oh My Dear' (2019 Winter Edition) Appellate Issues

Krent HJ and Rucker R, 'The First Step Act - Constitutionalizing Prison Release Policies' (2022) 74 Rutgers UL Rev 631

Larson J and others, 'How We Analyzed the COMPAS Recidivism Algorithm', ProPublica, 23 May 2016

Lee Park A, 'Injustice Ex Machina: Predictive Algorithms in Criminal Sentencing', UCLA Law Review, 19 February 2019

LexisNexis, 'LexisNexis Announces Acquisition of Ravel Law' (8 June 2017)

Lightbourne J, 'Damned Lies & Criminal Sentencing Using Evidence-Based Tools' (2017) 15 Duke L & Tech Rev 327

Legal Information Institute, 'Executive Order' (*Legal Information Institute*) <www.law.cornell.edu/wex/executive_order> accessed 28 March 2023

Liptak A, 'Sent to Prison by a Software Program's Secret Algorithms' *The NY Times* (New York, 1 May 2017), A22

Ludwig J and Sunstein CR, 'Discrimination in the Age of Algorithms' *The Boston Globe* (Boston, 24 September 2019) A8

Mapping Pretrial Injustice, 'Common Pretrial Risk Assessments' <https://pretrialrisk.com/the-basics/common-prai> accessed 28 March 2023

– – 'How Many Jurisdictions Use Each Tool?' <https://pretrialrisk.com/national-landscape/how-many-jurisdictions-use-each-tool> accessed 28 March 2023

– – 'State Laws on Rats' <https://pretrialrisk.com/national-landscape/state-laws-on-rats> accessed 28 March 2023

Markovic M, 'Rise of the Robot Lawyers' (2019) 61 Ariz L Rev 325

Mayson SG, 'Dangerous Defendants' (2017) 127 Yale LJ 490

Metz A and others, 'Valid or Voodoo: A Qualitative Study of Attorney Attitudes Towards Risk Assessment in Sentencing and Plea Bargaining' 8 (10 March 2020) Virginia Public Law and Legal Theory Research Paper No 2020-25, Duke Law School Public Law & Legal Theory Series No 2020-15

Miller C, 'The Prospects of Constitutional Challenges to COMPAS Risk Assessment (26 April 2021)', ABA Privacy and Data Security

National Institute of Justice, *2021 Review and Revalidation of the First Step Act Risk Assessment Tool* (No 303859, December 2021)

Nishi A, 'Privatizing Sentencing: A Delegation Framework for Recidivism Risk Assessment' (2019) 119 Colum L Rev 1671

Note, 'Beyond Intent: Establishing Discriminatory Purpose in Algorithmic Risk Assessment' (2021) 134 Harv L Rev 1760

Northpointe Research and Development Department, 'COMPAS Scales and Risk Models – Validity and Reliability: A Summary of Results from Internal and Independent Studies', *Elec Priv Info Ctr*, 20 July 2010

O'Brien T, 'Compounding Injustice: The Cascading Effect of Algorithmic Bias in Risk Assessments' (2021) 13 Geo JL & Mod Critical Race Persp 39

Oleson JC, 'Risk in Sentencing: Constitutionally Suspect Variables and Evidence-Based Sentencing' (2011) 64 SMU L Rev 1329

Partnership on AI, 'Report on Algorithmic Risk Assessment Tools in the US Criminal Justice System', 23 April 2019

– – 'About Us' <https://partnershiponai.org/about/> accessed 29 March 2023

Pattison-Gordon J, 'Justice-Focused Algorithms Need to Show Their Work, Experts Say' (12 May 2022) government technology

Polonski S, 'Mitigating Algorithmic Bias in Predictive Justice: 4 Design Principles for AI Fairness' (24 November 2018) Towards Data Science

ProPublica, 'About Us' <www.propublica.org/about> accessed 28 March 2023

PRWeb, 'Ravel Law Announces Unprecedented Judge Analytics Offering' (16 April 2015) <www.prweb.com/releases/2015/04/prweb12656883.htm> accessed 28 March 2023

Public Safety Risk Assessment Clearinghouse, 'Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)'

Rabinovich-Einy O and Katsh E, 'Artificial Intelligence and the Future of Dispute Resolution – The Age of AI-DR' in MA Wahab, D Rainey, and E Katsh (eds*), Online Dispute Resolution: Theory and Practice* (2nd edn, Eleven International Publishing 2021) pp. 471-488

Ravid I and Haim A, 'Progressive Algorithms' (2022) 12 UC Irvine L Rev 527

Recent Cases, 'Criminal Law - Sentencing Guidelines - Wisconsin Supreme Court Requires Warning before Use of Algorithmic Risk Assessments in Sentencing - *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016)' (2017) 130 Harv L Rev 1530

Remus D and Levy F, 'Can Robots Be Lawyers: Computers, Lawyers, and the Practice of Law' (2017) 30 Geo J Legal Ethics 501

Reyes CL and Ward J, 'Digging into Algorithms: Legal Ethics and Legal Access' (2020) 21 Nev LJ 325

Rigano C, 'Using Artificial Intelligence to Address Criminal Justice Needs' (January 2019) NIJ Journal 280

Santos M, 'PATTERN Risk and Needs Assessment Under First Step Act' (*Prison Professors*)

Schaefer TB, 'The Ethical Implications of Artificial Intelligence in the Law' 55 Gonz L Rev 221

Schmitz AJ and Martinez J, 'ODR and Innovation in the United States' in MA Wahab, D Rainey, and Ethan Katsh (eds), *Online Dispute Resolution: Theory and Practice* (2nd edn, Eleven International Publishing 2021) pp. 611-638

Simon M and others, 'Lola v Skadden and the Automation of the Legal Profession' (2018) 20 Yale JL & Tech 234

Simshaw D, 'Access to AI Justice: Avoiding an Inequitable Two-Tiered System of Legal Services' (2022) 24 Yale JL & Tech 150

Slobogin C, 'Preventive Justice: How Algorithms, Parole Boards, and Limiting Retributivism Could End Mass Incarceration' (2021) 56 Wake Forest L Rev 97

Snyder J, 'RoboCourt: How Artificial Intelligence Can Help Pro Se Litigants and Create a "Fairer" Judiciary' (2022) 10 Ind JL & Soc Equal 200

Southerland VM, 'The Intersection of Race and Algorithmic Tools in the Criminal Legal System' (2021) 80 Md L Rev 487

Tashea J, 'Courts Are Using AI to Sentence Criminals. That Must Stop Now' 17 April 2017, Wired

Taylor AM, 'AI Prediction Tools Claim to Alleviate and Overcrowded American Justice System … But Should They Be Used?' Stanford Politics (13 September 2020)

Thadaney Israni E, 'When an Algorithm Helps Send You to Prison' *The NY Times* (New York, 26 October 2017)

Threadcraft-Walker W and others, 'Gender, Race/Ethnicity and Prediction: Risk in Behavioral Assessment' (2018) 54 J Crim Just 12

Turner S and others, 'Development of the California Static Risk Assessment (CSRA): Recidivism Risk Prediction in the California Department of Corrections and Rehabilitation' (September 2013) UC Irvine Center for Evidence-Based Correction Working Paper

Vought RT, Memorandum for the Heads of Executive Departments and Agencies, *Guidance for Regulation of Artificial Intelligence Applications* (17 November 2020)

Ward J, 'Black Box Artificial Intelligence and the Rule of Law' (2021) 84 Law & Contemp Prob

Washington AL, 'How to Argue with an Algorithm: Lessons from the COMPAS-ProPublica Debate' (2018) 17 Colo Tech LJ 131

Weizenbaum J, *Computer Power and Human Reason: From Judgment to Calculation* (WH Freeman and Co, 1976)

Wexler R, 'Code of Silence: How Private Companies Hide Flaws in the Software That Governments Use to Decide Who Goes to Prison and Who Gets Out' (June/July/August 2017) Washington Monthly

– – 'Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System' (2018) 70 Stan L Rev 1343

– – 'Privacy Asymmetries: Access to Data in Criminal Defense Investigations' (2021) 68 UCLA L Rev 212

The White House, 'Fact Sheet' (*The White House*, 25 May 2022)

The White House, Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People (October 2022) <www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf> accessed 29 March 2023

Winters B, 'Brief of Amicus Curiae,' *Rodriguez v. Massachusetts Parole Board*, SJC-13197 (*Elec Priv Info Ctr*, 14 February 2022)

NATIONAL REPORTS ON EVIDENCE THROUGH ARTIFICIAL INTELLIGENCE

AI AND ADMINISTRATION OF JUSTICE IN CHINA

By Haiyan Wang[*]

Abstract

AI technology is playing an increasingly important role in criminal justice. China is also deeply integrating AI with technology justice, not only releasing a series of guiding policies, laws and regulations, but also applying AI technology in the whole litigation stage of examination and prosecution and court trial. In addition to this, AI-driven evidence is also one of the important applications. However, AI technology also gives rise to urgent issues and challenges, such as algorithmic discrimination and privacy violation. These issues may infringe on the fundamental rights of citizens (e.g., equality, privacy, communications freedom and confidentiality). In order to achieve better application of AI technologies under the premise of risk control, the following solutions are currently proposed by Chinese academics: (1) when discrimination arises, algorithmic explanation is first conducted, and class action lawsuits can be filed if the algorithm user refuses to explain; (2) equality between prosecution and defense is achieved through information disclosure and information disclosure; (3) due process restricts mandatory measures to protect citizens' personal information rights; (4) judicial review system should be established to protect privacy, etc.

1 Overview of Intelligent Justice Construction and AI

The world is now stepping into the fast lane of digital and intelligent development. Driven by the troika of algorithms, computing power and data, while supplemented with big data, extreme algorithms, cognitive science and artificial neural network, artificial intelligence (AI) technology is now deeply affecting and reshaping various aspects of social development based on deep learning, driven by logic calculus and by means of command output. Based on this, in 2017, the State Council issued The Development Plan for a New-Generation Artificial Intelligence (新一代人工智能发展规划), marking the rise of AI technology to a national strategic level, and providing directional guidance for the in-depth R&D and wide penetration of China's intelligent technology. The Guidelines for The Construction of National New-Generation Artificial Intelligence Standard System (国家新一代人工智能标准体系建设指南) jointly issued in 2020 by the Standardization Administration of China and other four departments points out that by 2023, an AI standard system will be preliminarily established, focusing on the development of key urgently needed standards such as data, algorithms, systems and services; the AI standard system will be firstly applied in key industries and fields such as manufacturing,

[*] Haiyan Wang (汪海燕), Professor in Criminal Procedure Law and Evidence Law, China University of Political Science and Law, zhlwhy@sina.com.

transportation, finance, security, home, elderly care, environmental protection, education, health care and judicial justice, so as to build AI standard test and verification platforms, and to empower such platforms to provide public services. Under the new technological revolution, AI is now empowering traditional policing, public prosecution and court trials to move towards the intelligent justice stage characterized by digitization, networking and intelligence, which is an inevitable requirement to adapt to the development of the times.

The legal implication of AI integrating with intelligent justice is 'digital justice'.[1] Fairness and justice are unremitting pursuits on the way toward judicial modernization; in the AI era, fairness and justice have been transformed into a 'digital justice' driven by science and technology. 'Digital justice' not only represents justice but also measures justice efficiency by digits. It requires minimizing the waste of judicial resources and using limited judicial resources to maximize the justice effect, so as to optimize the allocation of judicial resources. Of course, during integrating justice artificial intelligence with intelligent justice construction, we should, on the one hand, be vigilant against ignoring or even sacrificing justice due to the pursuit of justice efficiency, and on the other hand, strike a balance between justice and justice efficiency.

At present, AI technology is applied in case investigation, examination, prosecution and court trials. The public security and judicial organs across the country all put forward the goals of empowering the police and the procuratorial organ by technology, building smart courts, as well as improving the intelligent level of public security and judicial organs in office, case handling, service provision, decision-making and supervision, based on information technologies such as big data, cloud computing, Internet and AI. It can be predicted that AI will play an increasingly important role in China's intelligent justice construction, reform and practice; it is also of great significance to accelerate the construction of 'Digital China' and 'Safe China' and continuously promote the modernization of national governance system and governance capacity.

Absorbing cutting-edge technologies such as AI, broadening the scalability of integrating technology with justice, and building the largest concentric circle are irresistible trends in the digital age. At the same time, we must clearly understand that with the rapid development of AI technology, there come many ethical challenges and emerging legal issues. However, at present, the construction of intelligent criminal justice mostly focuses on the development and utilization of a new-generation AI technology, and the transformation, upgrading and effective utilization of AI technology itself, but lacks the standard construction and regulation path of the application of AI technology in criminal justice.

[1] Bin Wei, 'Difficulties and Paths of Integrating Judicial Artificial Intelligence into Judicial Reform' (2021) 43 MLS 4.

2 Policy Planning, Laws and Regulations on integrating AI within criminal justice

The development and innovation of AI technology not only promote the operation of criminal justice and the construction of intelligent justice but also give rise to urgent issues and challenges such as algorithmic bias and privacy infringement. We should, on the basis of fully mastering the development and prospects of AI technology, speed up the formulation of justice protection schemes for the development of AI technology, safeguard the deep integration of AI technology with intelligent criminal justice, and strive to make the people feel fairness and justice in each judicial case. Therefore, around the overall strategic planning and specific application design of AI and criminal justice construction, China has issued a series of guiding policies, laws and regulations to facilitate AI to inject new momentum into criminal justice.

2.1. Policy planning

China pays more and more attention to the huge potential and application possibility of AI in the construction of criminal justice, which is embodied in the transformation from the iterative updating of AI technology itself to the upgrading of AI technology in the field of intelligent justice construction, so as to enhance the technological support for the innovation of fair justice and justice for the people, promote the social fairness and justice, and maintain social harmony and stability. Based on this transformation, China has successively issued pertinent policies and plans regarding the examination and prosecution and court trial, as well as promoted and guided the integration of AI with intelligent justice construction step by step.

2.1.1. Examination and prosecution

In 2016, the Supreme People's Procuratorate issued The Outline of Procuratorial Work Development Plan During the 13th Five-Year Plan ("十三五"时期检察工作发展规划纲要), which established the overall goal of intelligent procuratorial work application system and the task of procuratorial big data construction of procuratorial organs at all levels, marking that the construction of intelligent procuratorial work has entered the intelligent development stage. Thereafter, the Supreme People's Procuratorate officially issued The Opinions of the Supreme People's Procuratorate on Deepening Intelligent Procuratorial Work Construction (最高人民检察院关于深化智慧检务建设的意见) on January 3, 2018, outlining the grand blueprint of intelligent procuratorial work construction in the future. In January 2021, the Supreme People's Procuratorate issued The Provisions of the People's Procuratorate on Handling Cybercrime Cases (人民检察院办理网络犯罪案件规定), which once again emphasized the active exploration of using big data, cloud computing, AI and other information technologies to assist in case handling, so as to improve the professional level of handling cybercrime cases. It can be seen that AI in the construction of intelligent procuratorial work has moved from the overall blueprint planning to the specific application design.

2.1.2. Court trial

In January 2016, the Supreme People's Court proposed for the first time to build smart courts. It refers to a people's court organization, construction, operation and management form, which relies on AI supporting judicial adjudication, litigation services and judicial management in a highly information-based way. They intend to process full business online and provide a full range of intelligent services, based on the state of the technology.

In 2017, the Supreme People's Court issued The Opinions of the Supreme People's Court on Accelerating the Construction of Smart Courts (**最高人民法院关于加快建**设智慧法院的意见), proposing to explore the establishment of a knowledge map for court business such as case filing, court trial, judgment rendering and enforcement, construct AI perception interactive system and knowledge-based AI aided decision-making system for various users, use big data and AI technology to provide targeted and intelligent services on demand, and promote the 'similar judgments for similar cases' and the standardization of sentencing. In 2018, the Artificial Intelligence Standardization White Paper (2018 Edition) (**人工智能**标准化白皮书(2018**版**)) specifies that the construction and application of smart courts need to rely on several AI technologies such as intelligent big data analysis, speech recognition, image and video analysis, so as to realize the functions such as case element analysis, automatic transcription of court speech recognition, analysis of court trial video, forwarding and scheduling of court video streaming media, etc. In 2022, the Supreme People's Court issued The Opinions on Regulating and Strengthening the Applications of Artificial Intelligence (**最高人民法院关于**规范和加强人工智能司法应用的意见) in the Judicial Fields, aiming at constructing an improved functional system for the application of AI in the judicial field by the year 2025.

China not only pays attention to the application of technology, but also to exploring practical samples for the construction of intelligent justice, thus producing advanced experience that can be duplicated and popularized, and giving full play to the leading and exemplary role of smart courts in concept innovation, technological innovation and institutional innovation, so as to effectively promote judicial reform and innovation.

2.2. Laws and regulations

At present, although China has not promulgated special legal provisions on the integration of AI with criminal justice, the relevant provisions present the possibility of AI being applied in criminal justice. For example, Article 53 of The National Security Law of China (**国家安全法**) stipulates that 'in carrying out intelligence information work, we should make full use of modern technologies to strengthen the identification, screening, synthesis, research, judgment and analysis of intelligence information.' Intelligence information work is an important part of criminal justice, especially in the investigation stage, which provides guidance for the integration of modern technologies represented by AI technol-

ogy with the analysis/judgment of intelligence information. Act 21,28,44,52 of The Network Security Law of China (网络安全法) also stipulates the application of relevant technical measures and other necessary measures.

Compared with the foregoing two laws, The Data Security Law of China (数据安全法) and The Personal Information Protection Law of China (个人信息保护法) newly issued in 2021 further clarify the application scenarios of AI technology, and preliminarily regulate the application of AI technology in principle. For example, the provisions of The Data Security Law of China on data development and utilization technology and the construction of standard systems provide a reference for integrating AI with the construction of intelligent criminal justice. The Personal Information Protection Law of China regulates automated decision-making for the first time. According to Article 24 of The Personal Information Protection Law, 'When using personal information for automated decision-making, personal information processors shall ensure the transparency of decision-making and the fairness and impartiality of the results.' Article 55 further stipulates that 'a personal information processor who uses personal information for automated decision-making shall conduct a personal information protection impact assessment in advance and record the processing.' It means that transparency, impact assessment and fairness of the results are conditions of automated decision-making. Besides, The Personal Information Protection Law of China also pays further attention to new technologies and applications such as facial recognition and AI, as well as stipulates special personal information protection rules and standards.

It can be seen that most of the relevant laws and regulations on AI in China are still of a guiding and fundamental nature, without much operability. At the same time, there are also no specific legal provisions for the integration of AI with intelligent criminal justice. The Standing Committee of the National People's Congress clearly mentioned the need to strengthen the relevant legislative work concerning the new applications and technologies such as digital economy, Internet finance, AI, big data and cloud computing in 2021, so as to create a law-based environment for healthy development. This provides a significant guiding value for the integration of AI with the legal regulation of criminal justice.

2.3. Relevant industry rules or standards

In order to promote the healthy development of AI in the new era, in June 2019, China's National Professional Committee for the Governance of New-Generation Artificial Intelligence issued The Governance Principles of New-Generation Artificial Intelligence - Developing Responsible Artificial Intelligence (新一代人工智能治理原则——发展负责任的人工智能) in Beijing, explicitly proposing eight principles, i.e., harmonious and kind, fair and justice, inclusive and sharing, respect for privacy, safe and controllable, shared responsibility, open and cooperation, and agile governance. The foregoing principles provide an important reference for the application of AI in criminal justice. In September 2021, China's National Professional Committee for the Governance of New-Generation

Artificial Intelligence issued The Code of Ethics for New-Generation Artificial Intelligence (新一代人工智能伦理规范), which aims to integrate ethics with the whole life cycle of AI and provide ethical guidance for natural persons, legal persons and other relevant institutions engaged in AI-related activities. The Code of Ethics puts forward six fundamental ethical requirements, i.e., enhancing human well-being, promoting fairness and justice, protecting privacy, ensuring controllability and credibility, strengthening the sense of responsibility and improving ethical literacy. Besides, The Code of Ethics puts forward 18 specific ethical requirements for specific activities such as AI management, R&D, supply and use. This also provides fundamental guidance for the integration of AI with intelligent criminal justice.

3 Practical Applications of Artificial Intelligence in Criminal Justice

3.1.1. Application of artificial intelligence in examination and prosecution

At present, the application of AI in the process of examination and prosecution in China is mainly reflected in the guidance of evidence standards, evidence verification, evidence chain examination, procurator performance assessment (supervision), accuracy of sentencing suggestions, examination and arrest, etc. The specific application examples are as follows.

3.1.2. Application of data intelligence

The core of intelligent procuratorial work lies in the construction of intelligent systems. In recent years, local procuratorial organs have earnestly implemented the requirements of The Action Guide for Procuratorial Big Data (2018-2020) (检察大数据行动指南(2018—2020年)), strengthened the construction of intelligent systems of procuratorial organs through independent innovation and external forces, and made a breakthrough in the application of data intelligence.

The case intelligent research and judgment system is a typical application. This application is to use intelligent analysis in the system to make a pre-research and judgment on the nature of the case and the standard of evidence. For example, the procuratorial organs of Guizhou Province have developed an intelligent case research and judgment system, which can use the crime constitution theory in China's criminal law theory and the crime constitution system of different crimes in the specific provisions of criminal law to produce the knowledge map of different crime constitution elements. At the same time, it compares the weights of various statutory sentencing circumstances and discretionary sentencing circumstances in criminal law to produce a standardized map of conviction and sentencing, as well as systematically analyzes and weights the criminal evidence standards involved in The Criminal Procedure Law (刑事诉讼法) and the probative force of the evidence chain.

3.1.3. Application of perceptual intelligence

The application of perceptual intelligence is an important aspect of enriching the construction of intelligent procuratorial work through the continuous application of high-end perception technologies such as image recognition, character recognition, speech recognition and biometric recognition in the construction of intelligent procuratorial work systems.

First, the application of video recognition technology. For example, the perceptual intelligent application system of Shanghai procuratorates mainly uses video recognition technology to improve the business application level: by installing video recognition technology system in the penalty execution organ, it can automatically identify the behavior and status of the personnel under supervision, and can carry out intelligent analysis and behavior early warning in the system; it can standardize the behavior of procurators responsible for reception through video recognition technology to provide better procuratorial services; it can supervise the behavior of lawyers, parties and relevant personnel in related businesses, and prevent the occurrence of unnecessary trouble and unreasonable request through video recognition technology.

Second, the intelligent speech recognition system. For example, the 'intelligent speech recognition system' developed by iFLYTEK has been adopted by many judicial organs. The system empowers the information equipment and system to 'listen and remember' through speech recognition and speech synthesis technology, so as to realize the man-machine speech interaction. By automatically converting voice into text, the system has outstanding performance in document preparation, file reading and excerption. The case handling personnel only need to make dictation, and the system will automatically convert the oral content into written text and generate documents immediately, which greatly improves the case handling efficiency. The court trial speech recognition system developed by the Suzhou Intermediate People's Court under entrustment by the Supreme People's Court can automatically transcribe speech into text, automatically distinguish the speakers and contents of the court hearing, and the judges, parties and other participants can see the transcribed text in real time.[2]

The procuratorial organs of Anhui Province have achieved good results in terms of taking the initiative to embrace modern technology, making use of the application of intelligent voice technology, and developing a new intelligent procuratorial work model with Anhui characteristics. They have also implemented 'Three Applications': the application of intelligent voice input method (it is widely used in office/case handling scenarios such as document drafting, case information input, making file-reading notes, legal instrument drafting, etc.), the application of intelligent voice conference system (it is widely used in Procuratorial Committee, Party group meeting, office meeting and other occasions. Through human-computer interaction, it can realize the functions of role separation, text segmentation, key mark, audio delayed play, recording playback, rapid gener-

[2] Guofeng Ding, 'The Construction of "Smart Courts" in Jiangsu Injects New Impetus into the Modernization of Judicial Capacity' *Legal Daily* (Beijing, 20 March 2017) 1.

ation of meeting minutes, and the recognition accuracy is more than 90%.), and the application of intelligent voice inquiry system (it is mainly used in the inquiry process of investigation supervision, public prosecution and other departments. Through speech recognition technology, the voice of both sides of the inquiry will be transcribed into text in real-time according to the inquiry record format, to produce a standardized inquiry record in time; 'Two Integrations': to promote the application of intelligent case handling aided system in cases applicable to summary procedures and the application of voice file-reading and evidence presentation system in cases applicable to ordinary procedures; 'One Center': to explore and establish the first procuratorial intelligent voice cloud center and intelligent voice cloud computer room of procuratorial organs in China.[3]

3.1.4. Application of cognitive intelligence

In the application of robot intelligence, cognitive intelligence is that machines have the ability of active thinking and understanding similar to human beings. The application of cognitive intelligence in intelligent procuratorial work is to develop intelligent systems step by step and tap the cognitive system and understanding ability of machines in intelligent development during the construction of intelligent procuratorial work, so that robots can learn the 'general expression' closest to human brain cognition and obtain the perception ability, understanding and analysis ability similar to human brain, so as to push the AI-enabled intelligent procuratorial work construction to a new level and further promote the intelligent construction level of intelligent construction.[4]

First, the case management robot. For example, the procuratorial organs of Jiangsu Province have developed a 'case management robot'. The 'case management robot' can compare and analyze the case card filling and various legal documents of the procuratorial organs. Through the comparative analysis, it can check the obtained data, and further remind, warn and evaluate the possible qualitative or evidential problems of the case itself. It can correct errors through robots, analyze data or clerical errors, and timely find out mistakes and defects in case handling documents.

Second, the procuratorial work robot. For example, the Xiqing District People's Procuratorate of Tianjin Municipality has developed a procuratorial work robot in the 12309 Procuratorial Service Center. The procuratorial work robot can not only provide a touch operation menu for handling related businesses but also make available the function of man-machine interaction and communication. Through the facial recognition function of the procuratorial work robot, the new visitor's face will be registered and automatically remembered. In addition to serving as the 'guide' of the procuratorial service hall, the basic responsibilities of the procuratorial work robot can also handle preliminary busi-

[3] Mian Zhang, 'Embracing the New Technology of Intelligent Voice and Creating a New Engine of Intelligent Procuratorial Work' (2017) 753 PPS 28.

[4] Shu Xie, 'How Can Artificial Intelligence "Unbiasedly" Help Criminal Justice -- From "Evidence Guidance" to "Proof Assistance"' (2020) 38 JNUPSL 109,117.

nesses such as case management, prosecution/appeal reception and business consultation according to the different needs of the public, freeing human personnel from many basic operation services.

3.1.5. Others

As early as 2005, the Minhang District People's Procuratorate of Shanghai Municipality developed the feasibility evaluation system of non-custodial measures for minors, and then Beijing Municipality, Taiyuan City and other places developed a variety of quantitative evaluation tools.[5]

The public prosecution in court.[6] The procuratorates of Tianjin Municipality have conducted evidence presentation by multimedia through all links of court trial, forming a new mode of multimedia-driven cross examination of evidence in special case handling; the Ziyang Procuratorate of Sichuan Province has developed the 'integrated platform for court appearance' (a 'integrated platform for court appearance' system containing pretrial preparation, charges during court trial and background support) evidence presentation system based on electronic files, which endeavors to solve the contradiction between the diversity of evidence types and the lag of evidence presentation methods by multimedia-driven evidence presentation, and to improve the public prosecution in court; the procuratorates of Beijing Municipality have developed a court appearance management system, which integrates various functions such as court appearance information collection and release, court appearance observation and appointment, online comments on court appearance, court appearance problems and experience summary, and court appearance experience value ranking, so as to strengthen court appearance management.

As regards the application of AI in the process of examination and prosecution, the academic circle has also made relevant responses. In the application of AI in evidence judgment, some scholars believe that in terms of evidence validity, AI cannot conduct substantive examination, but can conduct formal examination, such as whether the interrogation meets the procedural requirements; in terms of probative force, AI cannot function independently, and may play an auxiliary and reference role in examining the authenticity of evidence; in terms of standard of proof, the role of AI is not to judge the standard of proof regarding evidence specification and analysis, but is only an auxiliary means for judges to judge the standard of proof.[7] Some scholars believe that evidence standards and proof standards occur in different stages; the evidence standard mainly

[5] Zhenhui Wang,' Principle and Construction of Quantitative Model for Review of Social Risk Assessment of Arrest' (2016) 34 PLF 73,74.
[6] Qian Sun,' Promoting the Deep Integration of Procuratorial Work and New Technology, Effectively Improving the Quality and Efficiency of Case Handling and Judicial Credibility' (2017) 752 PPS 7.
[7] Bo Zong, 'Analysis on the Application of Artificial Intelligence in Criminal Evidence Judgment' (2018) 37 JNUPSL 61.

appears in the pre-trial stage, such as case filing, arrest, investigation conclusion and public prosecution; the proof requirements of evidence on the facts of a case can be referred to as the evidence standard. However, what generally appears in the court trial is the proof standard; therefore, the participants of evidence standard and proof standard are different. The standard of evidence in the pre-trial stage is the result of unilateral investigation of the facts of the case, which is monopolized by the public power; the standard of proof is the degree to which the three parties (i.e., the prosecution, the defense and the judge) jointly procure the evidence to prove the facts of the case through cross-examination, debate, and investigation.[8] Some scholars have pointed out that the limitations (the subjectiveness in perception, the uncertainty in practice, the unity of the criminal procedure, the idealization of value) of the standard of proof have become the direct cause of the establishment of basic evidence standard guidelines in judicial practice. We can achieve a revolutionary leap in criminal examination by developing an intelligent case-handling-aided system, turning the evidence standards into rigid requirements, transforming them into standardized data models, and embedding them into the intelligent case-handling-aided system to give full play to the advantages of big data such as objectivity, accuracy, and resistance to external factors, if organically combined with the subjective initiative of law enforcement personnel, and together with the transformation from manual examination only to the combination of manual and artificial intelligent examination.[9] Some scholars have put forward the concept 'unified standard of evidence' in response to the standard of evidence. They believe that the standard of evidence is a derivative concept from China's judicial practice, and is sometimes interchangeable with the standard of proof; sometimes it is used to distinguish the standard of proof (evidence standard) in the pre-trial stage from the standard of proof in the trial stage. The AI in criminal proof can take a 'unified standard of evidence' as the core and develops around the guidance of evidence standard, the guidance of evidence rules, the verification of single evidence, the examination and judgment of the evidence chain and the whole case evidence, the guidance of factor-based interrogation and the exclusion of illegal verbal evidence.[10] In this regard, some scholars have put forward the 'digital evidence standard', that is, the digital evidence standard uses AI technology to machine learn and deeply mine the typical criminal cases, judicial information resources, case handling experience accumulated in judicial practice, as well as the evidence standards, evidence rules and evidence models formulated by local judicial organs, as well as to enumerate the types of evidence and procedures that should be available before trial for certain types of cases from the long-term accumulated judicial experience and form a list of guidelines, and to verify the consistency of each evidence to be verified, the logical consistency between different evidence and the controversy between evidence, which is

[8] Kun Dong, 'Evidence Standard: Connotation Reinterpretation and Path Prospect' (2020) 19 CLR 109.
[9] Guosheng Cai, 'Origin, Development and Function of Criminal Evidence Standard Guidance' (2021) 306 SSS 187.
[10] Qiuhong Xiong,'Application of Artificial Intelligence in Criminal Proof' (2020) 34 CLR 75.

a standard to regulate the 'quantity' and 'quality' in a lesser significance of conclusive evidence.[11]

In the application of AI technology to assist procuratorial organs in the accuracy and standardization of sentencing suggestions, most studies are carried out in the context of plea for leniency. Focusing on 'AI-assisted accurate prediction and sentencing', a scholar proposes that theoretical prediction and data prediction form a 'dual core' collaboration, the two links 'match' and verify each other, and the necessary manual intervention mechanism is configured to ensure the output of accurate sentencing suggestions analyzed and determined jointly by theoretical basis, data support, prediction verification and manual intervention. It can be seen that the scholar believes that the supporting status and reference function are the 'double drive' fulcrum for the implementation of AI-assisted accurate prediction and sentencing.[12] Some scholars have proposed that procuratorial organs at all levels can adopt mandatory regulations to require case handlers to make full use of technological means such as big data and AI to assist in accurate sentencing. The procuratorial organ shall provide material guarantee for prosecutors to use big data for sentencing.[13] The deviation degree early warning mechanism based on legal reasoning, intelligent prediction and deviation degree analysis function proposed by some scholars can not only ensure the correct exercise of judges' jurisdiction, but also effectively ensure the accuracy and standardization of procurators' sentencing suggestions under the current background of plea for leniency.[14] In addition, some scholars believe that in the field of criminal procedure, the intelligent judgment aided system, including aided sentencing and similar judgments for similar cases, has been applied to judicial practice, providing a strong 'external brain' support for judicial decision-making. After the plea for leniency system was written into the law, local procuratorial organs have gradually routinized their case handling relying on the sentencing suggestion aided system. This intelligent judgment aided system can not only effectively help procurators put forward sentencing suggestions, but also shorten the time for procurators to handle cases of plea for leniency, which has become a critical link in deepening the construction of 'intelligent procuratorial work'.[15]

In addition, some scholars have conducted a quantitative assessment of the social risk of arrest from the perspective of "arrest", and put forward the quantitative assessment model of the social risk of arrest, which is to use the social learning theory in criminology theory to predict the social risk by analyzing the factors affecting people's social learning progress; the index system of the model is generally developed around the "core eight

[11] Tao Yang, 'Rationality and Limit Analysis of Digital Evidence Standard -- Focusing on Shanghai "206" Intelligent System' (2020) 47 JSNU 45.
[12] Daocui Sun, 'Artificial Intelligence Assisted Accurate Prediction of Sentencing in China -- Taking Plea for Leniency Cases as the Applicable Field' (2020) 42 JJU 76,77.
[13] Yong Yang, 'Problems and Optimization in the Practice of Sentencing Suggestions in Plea for Leniency Cases' (2020) 312 AE 93.
[14] Ran Wang, 'Research on Judicial Supervision Mechanism of Big Data' (2021) 24 HUST (SSE) 136.
[15] Siyuan Wu, 'The Dilemma and Transformation of China's Plea Bargain Mode -- from "Confirmation and Approval Mode" to "Negotiation and Review Mode"' (2020) 1 CS 154.

indicators", including criminal history or litigation evasion history, antisocial personality, criminal attitude, criminal connection, educational background and occupation, family members and military service, drug abuse, entertainment and rest habits.[16] In addition to the deviation early warning mechanism, some scholars, from the perspective of judicial supervision, also propose that big data provides a new path for the supervision of judicial power, which is reflected in the real-time supervision mechanism based on data collection, the performance evaluation mechanism based on data portrait and the evidence examination mechanism based on knowledge map.[17] Some scholars, from the perspective of preventing criminal wrongful conviction, put forward three stages for AI to intervene in the prevention of criminal wrongful conviction, namely, the data coding stage, text generation-data link stage and standardized judicial product output stage.[18] Some scholars have put forward the application of AI in the fields of judicial case handling, management and service based on the construction of electronic procuratorial projects; specifically, in the field of judicial case handling, it mainly includes intelligent speech recognition, criminal sentencing suggestions and automatic generation of legal documents; in the field of judicial management, it mainly includes the dynamic circulation of procuratorial office and the team management data portrait; in the field of judicial services, it mainly includes procuratorial work publicity and intelligent services.[19]

Of course, we should be vigilant about the application of AI technology in the judicial field. In terms of AI technology in the criminal law application, we should not consider or excessively consider the limitations of criminal law but should prevent AI technology from stepping into the legal forbidden zone such as case-based rule[20] and informal institutions[21]. At the same time, we should explore the scientization and standardization of AI-driven criminal justice in practice.[22] Some procuratorial personnel fail to properly update their ideas and actively make full use of technology to serve case handling; generally speaking, the application of intelligent prosecution still requires further improvement and cannot fully meet the needs of case handling; the working mechanism innovation cannot keep up with the technological innovation; problems such as the unbalanced development of intelligent case handling among different regions cannot be ignored. Some scholars have also put forward three principles to be followed by AI-enabled evidence judgment, including : (1) the auxiliary principle, that is, AI can only play an auxiliary

[16]Tong Gao, 'Research on Quantitative Assessment of Social Risk of Arrest -- From the Perspective of Automated Decision-making and Algorithmic Regulation' (2021) 15 NLS 135.
[17]Ran Wang, 'Research on Judicial Supervision Mechanism of Big Data' (2021) 24 JHUST (SSE) 132.
[18] Xiumei Wang and Ling Tang, 'Application and System Design of Artificial Intelligence in Preventing Wrongful Conviction' (2021) 42 LM 100.
[19] Xia Cui, 'Towards Intelligentization: The Practical Path of Artificial Intelligence Embedded in Procuratorial Work Reform' (2021) 290 SS 132.
[20] Case-based rule, that is, Cases in the Criminal Trial Reference complied by the business department of the Supreme People's Court to guide law enforcement and handing cases. Local people's courts at all levels compile and publish 'Case Reference' 'Model cases' 'Typical Cases' to summarize judicial experience and guide judicial work.
[21] Informal institution refers to criminal policy, reform experiment and local regulation.
[22] Jingping Huang, 'Negative List of Criminal Justice Artificial Intelligence' (2017) 10 EFV 85,94.

role in evidence judgment, but cannot replace the judge's examination and judgment of evidence ; (2) limitation principle, that is, when AI is used for evidence judgment, it can only be limited to specific aspects, and not all evidence judgment can be made by AI ; (3) rebuttable principle, that is, when AI is used in one aspect of evidence judgment, it must be clear that the calculation results of AI in evidence judgment are not 'absolutely accurate', but refutable and revocable. Not only can judicial personnel directly abandon the calculation results of AI with justified reasons, the party concerned may also raise an objection to the AI calculation results and ask the judicial organ not to consider unreasonable calculation results. Some scholars have pointed out the problems existing in the sentencing proposal in the event of plea for leniency: the interval sentencing proposal accounts for the vast majority and the range of sentencing proposal is too wide, the proposal for the application of fine and probation is relatively arbitrary, the expression of sentencing circumstances is relatively messy, the laws and regulations referred to for sentencing are not unified, the application of non-prosecution is pretty rare, and the production of bill of prosecution is not standardized. It also puts forward that procuratorial organs at all levels can mandatorily require case handlers to make full use of technological means such as big data and AI to assist in accurate sentencing. The procuratorial organ shall provide material support for procurators to use big data-driven sentencing.23 Some scholars have put forward handling suggestions for the weakening of rational factors in evidence judgment due to the combination of AI and evidence standard, and the hidden worries of case handling personnel suffering from case handling inertia and path dependence. For example, some case handling personnel think that the cases handled meet the system requirements is the end of story. However, this is not only an escape from the responsibility of handling cases, but also may lead to mechanical justice. In order to solve foregoing problems, first of all, it should be made clear from the concept that the integration of evidence standard and AI should be moderate rather than absolute, and the legal problems must not be completely trusted to the algorithm, which will lead to the weakening or even elimination of factors such as human rationality and goodness in judicial case handling. Secondly, we should clarify the functional boundaries of two different fields: online intelligent operation and offline independent case handling. Finally, regarding the path dependence of case handling personnel, the usual practice is to link the case handling accountability system of case handling personnel with the case handling quality. Through the evaluation of case handling quality, case handling personnel are forced to actively improve their competence and get rid of the bad working habit of path dependence. However, at a deeper level, the real purpose of eliminating path dependence lies in the mutual restriction among case handling organs. Especially under the background of trial-centered litigation system reform, we should further substantiate the court trial, and give priority to the role of the prosecution, the defense and the judge in examining evidence in court trial, so as to solve a series of problems such as mechanized justice caused by path dependence in the pre-trial stage.[24] Some scholars

[23] Yong Yang, 'Problems and Optimization in the Practice of Sentencing Suggestions in Plea for Leniency Cases '(2020)312 AE 93.

[24] Kun Dong, 'Evidence Standard: Connotation Reinterpretation and Path Prospect' (2020) 19 CLR 118.

have put forward that the application of AI in the construction of intelligent procuratorial work is affected by the people's feelings of fairness and justice, the internal business needs of procuratorial organs and the driving force of AI integrating into judicial reform. Although AI has been widely used in procuratorial work, it is also restricted in many aspects. The lack of data samples, the defects of data quality and the shackles of data sharing are still unavoidable difficulties in terms of judicial data; there are also problems such as legal reasoning and knowledge labeling in the representation of legal knowledge with judicial logic; problems such as algorithm discrimination and algorithm black box associated with the operation of AI algorithms have not been solved. However, generally speaking, the application of AI in the construction of intelligent procuratorial work has become the mainstream trend. We should not only pay attention to the resource integration of judicial big data from vertical dimension, horizontal dimension and practical dimension, but also strengthen the in-depth integration of AI with procuratorial work.[25]

3.2. Application of artificial intelligence in court trial

In the judicial field, the article Some Speculation about Artificial Intelligence and Legal Reasoning by Buchanan and Headrick published in 1970 ushers in the research on AI in the field of judicial adjudication.

At present, AI in China's court trial is mainly used in evidence judgment (examination), aided sentencing, similar cases pushing, deviation prediction, remote trial, online judicial confirmation, performance evaluation (judgment evaluation), etc. The specific application examples are as follows.

3.2.1. Similar cases retrieval

Based on the core needs of judges in case handling, the courts in Beijing have innovatively constructed a 'Smart Judge' system serving unified judgment standards by using emerging technologies such as big data, cloud computing and AI and based on the Zhi-HuiYun Platform. Relying on the unified trial information resource database of the three-level courts in Beijing, 'Smart Judge' integrates multiple data resources such as judicial trial, judicial personnel, judicial administration and shared data, mines and analyzes the data resources, and automatically pushes the information such as case analysis, legal provisions, similar cases and judgments reference in the process of case handling, so as to provide unified and comprehensive trial norms and case handling guidelines for judges. 'Smart Judge' has access to multi-dimensional data support, automatically conducts the parties' information analysis, the trend analysis of this type of cases, comprehensive analysis of previous cases and the like according to the cases heard by the judge, as well as pushes all similar cases by relying on the legal rule database and the semantic analysis model. 'Smart Judge' also creates a whole process data service, which automatically extracts case information regarding the case filing stage, generates a 'case portrait', automatically generates a trial outline and record template regarding the trial stage, and

[25] Xia Cui, 'Towards Intelligentization: The Practical Path of Artificial Intelligence Embedded in Procuratorial Work Reform' (2021) 290 SS 132.

automatically generates judgment documents regarding the case closing stage, so as to realize the big data-driven service from case filing to case closing.[26]

The 'Enforcement AlphaGo' of Guizhou High People's Court is an 'enforcement big data application analysis system' with independent learning ability and can assist judges in handling cases. It is composed of Enforcement Think Tank + Senior Judge Database + Machine Artificial Intelligence Autonomous Learning. It uses AI and big data technology to deeply integrate various systems to form a unique system with autonomous learning ability to assist judges in decision-making through big data. When the presiding judge encounters a difficult case, the 'Enforcement AlphaGo' can automatically call similar cases and expert instructions from Enforcement Think Tank, generate more than two enforcement schemes and push them to the judge. The 37 process nodes of the enforcement case can have access to automatic case push, laws and regulations push, enforcement work specifications push, expert suggestions push, and videos push, etc., to make available intelligent services, so as to help judges quickly solve practical problems and improve enforcement efficiency.

3.2.2. Evidence standardization

In March 2006, the Zichuan District Primary People's Court of Zibo City, Shandong Province launched the computer sentencing software jointly developed with high-tech companies in the reform of sentencing standardization, realizing the application of AI in court sentencing.[27] Since 2016, Guizhou Province has taken the lead in trying to formulate the 'evidence standard guidelines' for the cases handled by public security organs, procuratorates and courts, and used big data to embed the element-oriented and structured evidence standards into the case-handling system, so that public security organs, procuratorates and courts can pay attention to the unified use of evidence and prevent wrongful conviction.[28] For another example, the Shanghai High People's Court developed the 'Shanghai intelligent case handling aided system for criminal cases' in 2018. By 'embedding the statutory unified evidence standard into the digital criminal case handling system of public security organs, procuratorates and courts', it tries to realize 'the unified evidence standard for the case handling personnel of public security organs, procuratorates and courts'. Specifically, this system should solve the problems such as inconsistent application of evidence standards in some significant, multiple and new types criminal cases. It requires what evidence should be collected and has the functions of inspection, check and supervision, so as to timely find flaws and contradictions in evidence, and make case handling personnel correct or explain.

---

[26] 'Beijing "Smart Judge" Promoting Similar Judgments for Similar Cases '(People's Court Daily,1 September 2017): <http://rmfyb.chinacourt.org/paper/html/2017-09/01/content_129653.htm> accessed 18 December 2021.

[27] Qiuhong Xiong,' Application of Artificial Intelligence in Criminal Proof' (2020) 34 CLR 79.

[28] 'Guizhou Political and Legal Organs Solidly Promote the Deep Integration of Technological Innovation and Judicial System Reform -- Accurate and Fair Case Handling Driven by Big Data '*People's Daily* (Beijing, 10 July 2017).

3.2.3. *Sentencing prediction (aided sentencing)*

Both the 'Legal Mirror System'[29] of Guizhou Province and the 'intelligent case handling aided system' developed by Shanghai 'Project 206' have the function modules of sentencing assistance, while the Hainan High People's Court has specially developed the 'standardized intelligent sentencing aided system'[30] to provide decision-making reference for judges to handle cases.[31]

3.2.4. *Text generation*

The court trial speech recognition system developed by the Suzhou Intermediate People's Court under entrustment by the Supreme People's Court can automatically transcribe speech into text, automatically distinguish the speakers and contents of the court hearing, and the judges, parties and other participants can see the transcribed text in real time.[32] In the trial operation of the system, the correct rate of speech recognition has reached more than 90%, and the clerk can finish the complete record of the court trial with only a small amount of correction. According to the comparative test, the court trial time is shortened by 20% ~ 30% on average, the court trial time of complex cases is shortened by more than 50%, and the integrity of court trial records reaches 100%.

3.2.5. *Deviation warning*

According to incomplete statistics, Shanghai, Jiangsu, Zhejiang, Guizhou, Yunnan and other provinces and cities have launched a trial aided system including the 'deviation early warning' function module.[33] Taking Jiangsu Province as an example, it has the first 'People's Court Justice Big Data Research Base' established by the Supreme People's Court nationwide (jointly built by Jiangsu High People's Court and Southeast University). Relying on the advantages of scientific research, the 'early warning platform for different judgments for similar cases' developed by the Research Base produces a sentencing algorithm through in-depth learning of many criminal documents, and automatically provides early warning for cases with great deviation, so as to provide technical support for unifying the judgment standard.[34] To be specific, when the judge determines the verdict result and completes the writing of the judgment document, the system will automatically capture the judgment document for intelligent analysis. Cases with high

---

[29] Xia Cui, 'Towards Intelligentization: The Practical Path of Artificial Intelligence Embedded in Procuratorial Work Reform' (2021) 290 SS 132, 137.

[30] 'Shanghai's Application of "Artificial Intelligence" in Case Handling to Prevent Wrongful Conviction, the Launch of China's First "Intelligent Case Handling Aided System' *Legal Daily* (Beijing, 11 July 2017).

[31] 'Let Modern Technology Better Help Judicial Reform -- Hainan Intelligent Sentencing System Operates "Faster, Better and More Cost-effectively"' *People's Court Daily* (Beijing, 9 December 2017).

[32] Guofeng Ding, 'The Construction of 'Smart Courts' in Jiangsu Injects New Impetus into the Modernization of Judicial Capacity' *Legal Daily* (Beijing, 20 March 2017) 1.

[33] Lusheng Wang,' Technical Barriers to the Development of Judicial Big Data and Artificial Intelligence' (2018) 20 CLR 48.

[34] 'Upgrading the Informatization Construction of Jiangsu 'Smart Courts' Injects New Impetus into the Modernization of Judicial Capacity '*Legal Daily (*Beijing, 20 March 2017).

deviation are automatically warned. The reasons for high deviation are explained to judges by using judicial big data visualization technology, or analyzing the distribution of similar cases and deviation status of judgment results.

3.2.6. Other applications of AI in the trial stage

In addition to the foregoing types of applications, the courts of Zhejiang Province have further promoted the 'Internet + trial' reform, conducting supervision through online traces and information disclosure.[35] The procuratorial organs of Jiangsu Province launched the 'procurator performance evaluation software', which realizes the automatic capture and calculation of relevant data, and establishes the digital personal files of procurators.[36] The courts in the Yangtze River Delta have established a professional judge meeting system of 'cross-region inquiry, pulse taking by expert and online prescription', which uses 'big data + AI' to gather judicial data resources in the Yangtze River Delta, analyzes regional judgment differences, law application, disputed issues and evidence citation, and promotes the cross-region 'similar judgments for similar cases'.

Based on the court trial stage, many Chinese scholars mainly discuss the application of AI in such links as evidence judgment (examination), aided sentencing, similar cases pushing, deviation prediction, remote trial, online judicial confirmation, and performance evaluation (judgment evaluation).

Evidence judgment (examination). Some scholars believe that the use of AI in the trial stage is basically the same as that in the procuratorial stage in terms of evidence validity and the probative force of a single evidence. The auxiliary function of AI in the judgment of proof standard should be mainly used in the trial stage, and the judgment of evidence in the trial stage has conclusive significance. Therefore, different requirements should be made for different stages of the trial. In AI systems, certain functional limitations should be imposed on the links used by the judge before the trial (including court trial preparation and pre-trial meeting), such as the discovery of flaws, defects, contradictions and judgments on whether the evidence meets the evidence specifications, and should not have the function of judging the probative force of single evidence or all evidence; in the court trial stage, AI shall not and cannot be used to assist in evidence judgment. the substantiation of court trial requires judges to form inner conviction during the court trial, and the principle of directness and verbalism should be implemented in the court trial, so that 'the investigation of factual evidence is conducted in the court and the judgment results are formed in the court'. However, the use of AI in the court trial process is bound to affect the judges' hearing and judgment of evidence, and will damage the authority and seriousness of the court trial as well. Therefore, judges should be prohibited from using AI at this stage; after the court trial, AI is used to assist the formation of inner

[35] 'New Highlights of Judicial Reform: Power and Responsibility Unification under the Judicial Accountability System': <http://llxfy.chinacourt.gov.cn/article/detail/2016/07/id/2042348.shtml> accessed 18 December 2021.

[36] Yonglian Zhuang, 'How to Build a Case Handling Performance Evaluation Mechanism for Procurator Quota System' (2017) 753 PPS 48.

conviction, but attention should be paid to inputting all evidence and cross-examination before and during the court trial into the system to avoid missing necessary evidence or information and thus affecting the judgment results. Moreover, conviction evidence and sentencing evidence should be separated as far as possible to avoid affecting the accuracy of AI-driven judgment.[37] Some scholars believe that evidence is the core of litigation, an important basis for restoring the facts of the case and an important basis for fair judgment. Evidence examination is an important part of the court trial. Evidence standardization is to summarize the experience of evidence authentication from many effective judgments through big data technology, transform the personal experience of multiple judges into a collective experience, and ensure the unity of evidence authentication standards.[38] In this regard, some scholars have proposed to establish a unified and electronic evidence standard, that is, to summarize the case handling experience through legal big data, and embed it in the digital case handling system of public security organs, procuratorates and courts, so as to standardize the judicial practice of public security organs, procuratorates and courts and their personnel.[39] In addition, some scholars suggest that the standing of human beings as decision-maker in judicial practice should not be shaken. AI can be used as an aid to supplement knowledge and support calculation, but it cannot be expected to become a 'vending machine' for judicial decision-making. If AI is to contribute to justice without prejudice, it should turn from 'evidence guidance' in a formal sense to 'evidence assistance'[40] in a substantial sense, and realize comprehensive upgrades based on proof principle, probability measurement based on evidence evaluation and cognitive monitoring based on holism. At the same time, human beings should not be shaken as the subject of judicial decision-makers, and the algorithm plays a supporting role rather than a dominating role, so as to avoid the uncontrollable negative effects of 'cognitive bias' hidden in AI on judicial practice.

Aided sentencing. Some scholars mention that the judicial application of intelligent sentencing algorithm not only promotes the structural transformation of China's traditional justice but also opens up the technical judgment path of 'similar judgments for similar cases'. This is mainly due to the subjective logic, quantitative normative logic and empirical normative logic of intelligent sentencing algorithm.[41] Some scholars have pointed out that simple sentencing considerations can be quantified, that is, they can be determined and calculated mechanically by computer programs, but in fact, in the aided sentencing system, the foregoing sentencing considerations need to be confirmed by the human

[37] Bo Zong,'Analysis on the Application of Artificial Intelligence in Criminal Evidence Judgment' (2018) 37 JNUPSL 68.

[38] Hui Zhu and Chenhui Liu, 'Research on the Application of Big Data in the Trial of Similar Cases' (2019) 20 JLA 47,54.

[39] Weimin Zuo, 'Some Thoughts on the Application Prospect of Legal Artificial Intelligence in China' (2018) 12 TLJ 108, 124.

[40] Shu Xie, 'How Can Artificial Intelligence "Unbiasedly" Help Criminal Justice -- From "Evidence Guidance" to "Proof Assistance"' (2020) 38 JNUPSL 109.

[41] Yujie Zhang, 'Judicial Application of Intelligent Sentencing Algorithm: Logic, Problems and Procedural Law Response' (2021) 81 OL 187.

brain (the judge), so the function of the computer is just a simple arithmetic operation. The real problem to be solved in sentencing is not to solve the calculation of punishment, but how to comprehensively consider and balance all factors affecting punishment (including personal and social factors), and finally present the most appropriate punishment for criminals. The punishment obtained through this procedure should reflect the comprehensive balance of social needs for crime retribution, prevention and suppression, correction and demand. Such a complex comprehensive balancing process cannot be undertaken by a programmed machine such as a computer, but should be undertaken by the human brain, that is, the judge.[42]

Online judicial confirmation. Some scholars have proposed three modes of online judicial confirmation, namely 'online reservation, on-site review', 'online reservation, written review', 'online reservation, video review'. In the 'AI + online judicial confirmation' mode, the AI-enabled machine independently reviews the judicial confirmation application from four aspects: first, whether the application materials are complete; second, whether the mediation agreement is reached by the parties voluntarily; third, whether the mediation agreement is enforceable; fourth, whether the electronic letter of commitment has been prepared.[43]

AI-enabled case division mechanism. Some scholars have mentioned the AI-enabled case division mechanism, that is, using AI technology to build an AI system applied to the court case division system to realize the automation and intelligentization of the case division system, that is, to study the basic theory, method and technology of how to apply computer software and hardware to simulate manual case division. At the same time, four modules are preset, namely case module, judge module, comparison module and output module. The setting items and variable values of each module are assigned by DelphiMethod. Through item-by-item comparison, the case division result is finally obtained.[44]

In addition, some scholars did not study a certain application of AI in court trial, but put forward a group of application types. Some scholars have proposed four forms of application of AI in smart courts, namely, the electronization and digitization of information, the intelligentization of case handling aided system, the prediction and supervision of judgment rendering, and the establishment of unified and electronic evidence standards.[45] Some scholars have pointed out that at present, the application of AI in court trial mainly focuses on the following three aspects: first, through intelligent speech recognition technology, it helps the whole process of court trial by trial records, case evaluation, document preparation and daily office work, so as to free trial assistant personnel from

[42] Qiuhong Xiong, 'Application of Artificial Intelligence in Criminal Proof' (2020) 34 CLR75,88.
[43] Mingliang Zhong, 'Practical Observation and Prospect of "Artificial Intelligence + Online Judicial Confirmation"' (2020) 15 JLA 122.
[44] Changwei Jin, 'Analysis on the Case Division Mechanism Driven by Artificial Intelligence' (2020) 76 JCUPSL 171.
[45] Weimin Zuo,' Some Thoughts on the Application Prospect of Legal Artificial Intelligence in China' (2018)12 TL 108,114.

recording or consulting affairs; second, through intelligent image and document recognition technology, it can realize the integration of sending, receiving and collecting electronic files, build smart trial big data, and free judges from simple case processing and cumbersome documents; third, through intelligent data analysis, it can realize judicial affairs management, evidence analysis, case reference, clerical error correction, etc., and assist judges in decision-making and judgment rendering.[46] Under the background of judicial big data, some scholars have discussed several important AI modules – similar cases recommendation, sentencing assistance and deviation warning from a technical perspective, analyzed their technical obstacles in judicial practice in detail, and proposed that similar cases recommendation, sentencing assistance and deviation warning are the most typical application modules in the development of judicial big data and AI. Their functions follow the technical path of map construction, plot extraction, similar case recognition, model training, sentencing prediction and deviation measurement.[47]

Of course, in promoting the application of AI technology in the trial stage, we should also pay attention to the following problems. Some scholars have pointed out that legal AI can only be a limited case-handling assistance means in the medium- and short-term in China, which is difficult to be applied to the core judicial work, i.e., judgment rendering, let alone to replace the thinking of human judges with technological means.[48] Some scholars have pointed out that the intelligent case handling system is exposed to the risk of discipline violation, exclusion and misjudgment, and further proposed that in order to effectively avoid the legitimacy risk caused by AI technology in the criminal trial field, we should establish the concept of power regulation, and regulate the intelligent case handling system from three aspects: the application mechanism (automatic judgment rendering), the participation mechanism (equalization of the defense), and the research and development mechanism (reliable decision-making), so as to protect the right of the accused to effectively participate in the intelligent system. In terms of the data, the defense lawyer of the accused can request to view, modify, correct and interpret the data related to their own rights and interests in the intelligent system.[49]

By comprehensively analyzing the articles and views of the above scholars, it can be found that at this stage, AI applied in the field of intelligent criminal justice only plays the role of auxiliary tools, and the results obtained from its analysis or technology are only a reference, the adoption of which depends on the judgment of judicial personnel. There are two different views on the function positioning of the judicial application of

[46] Xueqiang Gao, 'Chinese Justice in the Era of Artificial Intelligence' (2019) 49 JZU (HSSE) 229,237; Shuqin Zhang, 'Application of Artificial Intelligence in Trial' (2020) 49 JSNU (PSSE) 102,110.

[47] Lusheng Wang, 'Technical Barriers to the Development of Judicial Big Data and Artificial Intelligence '(2018) 20 CLR 46.

[48] Weimin Zuo, 'Some Thoughts on the Application Prospect of Legal Artificial Intelligence in China' (2018) 12 TLJ 108,124; Fuli Zhang and Haishan Zheng, 'Positioning, Prospect and Risk Prevention and Control of Artificial Intelligence Assisted Sentencing in the Era of Big Data' (2019) 283 GSS 92,100; Hongyang Luo and Xianglong Li, 'Ethical Issues in Intelligent Justice and Their Countermeasures' (2021) 1 PL 148,159.

[49] Chenshu Wei, 'Power Logic of Artificial Intelligence in Criminal Trial' (2021) 41 JXJU(SS) 147.

AI technology in the future: first, the application of AI technology only plays an auxiliary role at any time; [50] second, the application of AI in the judicial field may stand in a leading-role position in the future.[51]

3.3. Application of AI-driven evidence[52]

3.3.1. Application overview of AI-driven evidence

The deep integration of AI with intelligent justice is reflected in the field of evidence science, i.e., the emergence of AI-driven evidence. For example, in the second-instance criminal ruling concerning the crime of fraud committed by Yue Shanshan, the court held that 'Yue Shanshan provided a photo of Yang Wei (Wu Ziwei), and the investigation organ found out Geng, who was 95% similar to the photo through facial recognition technology, and Geng testified in court that she was the woman in the photo, but did not know Yue Shanshan and suspected that she had been secretly photographed.'[53] For another example, in the first-instance criminal judgment concerning the theft committed by Zhou Zhimin, the court held that

> the public security organ used the 'Hengyang static eagle eye facial recognition system' to compare the suspect images extracted from the theft scene at Dongliang Supermarket on September 19, 2018. The results showed that 16 people had a similarity of more than 70% with the targeted image, and it was found that the similarity of the fourth defendant Zhou Zhimin reached 69.41%. Viewed from the actual situation, 16 people had a similarity of more than 70% with the targeted image. Although there were many similar targets, the defendant Zhou Zhimin was listed as a key suspect because he was a native of Hengshan. However, the comparison result still cannot totally exclude other people from the suspect list, and is not enough to identify the defendant Zhou Zhimin as the real perpetrator of the theft.[54]

---

[50] Hongyang Luo and Xianglong Li, 'Ethical Issues in Intelligent Justice and Their Countermeasures' (2021) 1 PL 148,159.
Shuqin Zhang, 'Application of Artificial Intelligence in Trial' (2020) 49 JSNU (PSSE) 102,110; Yonglu Pan, 'Path Analysis of Artificial Intelligence Intervention in the Judicial' (2018) 3 OL109,118; Han Qin, 'Theoretical Reflection on AI-enabled Judicial System' (2021) 15 NLS 115,129; Xi Zheng, 'Application and Regulation of Artificial Intelligence Technology in Judicial Adjudication' (2020) 32 PULJ,647,696.

[51] Yujie Zhang,' Judicial Application of Intelligent Sentencing Algorithm: Logic, Problems and Procedural Law Response' (2021)3 OL 187,199; Mingliang Zhong, 'Practical Observation and Prospect of "Artificial Intelligence + Online Judicial Confirmation"' (2020) 15 JLA 122,130.

[52] The "AI-driven evidence" in this part refers to the evidence formed by the application of artificial intelligence; The judgment and analysis of evidence with the help of AI technology has been mentioned in the application of AI in examination, prosecution and trial.

[53] See the second-instance criminal ruling concerning the crime of fraud committed by Yue Shanshan, Case No.: (2020) Ji 02 Xing Zhong No. 210.

[54] See the first-instance criminal judgment concerning the theft committed by Zhou Zhimin, Case No.: (2020) Xiang 0423 Xing Chu No. 11.

This shows that the AI-enabled evidence conclusion represented by facial recognition technology is now concerned by judges and its acceptance has been taken into consideration.

However, the theoretical and practical circles in China have not paid enough attention to the AI-enabled evidence. At present, most of the existing studies focus on the topic of 'big data-driven evidence', but there are still disputes on the definition and type of big data-driven evidence. As for the definition of big data-driven evidence, some scholars, from the perspective of technical principles, introduced the three links to transform big data into evidence: the first step is to summarize and clean the data, the second step is to build an analysis model or machine algorithm, and the third step is to carry out operation to form an analysis conclusion; it is pointed out that big data-driven evidence is an analysis result or report based on massive electronic data.[55] On this basis, some scholars have further proposed that the big data-driven evidence has the dual structure of 'big data set' and 'big data report'.[56] Other scholars believe that big data-driven evidence is the evidence generated from filtering, summarizing, refining, and concluding massive data and is used in the court trial. At the same time, they point out that big data-driven evidence is different from 'analyzing and collecting evidence using big data technology'. The latter does not pose an obvious challenge to the traditional evidence rules, but the former will lead to an obvious conflict between big data-driven evidence and traditional evidence rules.[57] Some scholars, based on the methodological concept of big data, have pointed out that big data-driven evidence is a complex of case facts proving and analytical thinking, methods and technologies.[58] To sum up, it is not difficult to see that big data-driven evidence not only uses 'the conclusion formed by filtering, summarizing and refining massive data and then algorithm' as evidence but also includes 'directly using big data in the form of equal copies of data' as evidence. In this case, big data-driven evidence is closer to electronic evidence. Based on this, some scholars have pointed out that for data copies of big data, big data-driven evidence which is similar to electronic evidence can be examined according to electronic evidence examination rules and methods, which cannot reflect the particularity of such evidence. The uniqueness of big data-driven evidence lies in the part that draws conclusions through machine analysis, that is, AI-enabled evidence. The examination of this kind of evidence requires a new examination system. In other words, AI-enabled evidence is a machine opinion formed based on AI analysis that can be used to prove the facts of the case.[59]

[55] Pinxin Liu, 'On Big Data-Driven Evidence' (2019) 41 GLR 25.
[56] Yi Yuan, 'Attribute and Objective Verification Standard of Binary Physical Evidence of Big Data-Driven Evidence' (2021) 44 JSU(PSSE) 143.
[57] Fei Zheng and Guoyang Ma, 'Triple Dilemma and Way Out of Big Data-Driven Evidence Application' (2020) 28 JCU(SSE) 208.
[58] Hui Xu and Xiaodong Li, 'Research on Evidence Attribute Verification of Big Data-Driven Evidence' (2020) 36 JPPSUC (SSE) 50.
[59] Guoyang Ma, 'On the Examination of AI-driven Evidence in Criminal Procedure' (2021) CS 175.

3.3.2. Admissibility of cross-border criminal evidence

Network information technology has profoundly changed the external ecology and internal logic of criminal justice. The boundary between cybercrime and traditional crime is blurring, and electronic data has become a common and even key type of evidence in various crimes. The original criminal procedure system for traditional crimes based on a physical field system can hardly deal with such a large-scale crime transformation in time and effectively, and the dislocation between crime and crime governance is increasingly prominent. The cross-border criminal data collection is the exact embodiment of this misplaced relationship. Therefore, the request for assistance in the investigation and evidence collection between countries is undoubtedly the focus of the current international criminal judicial assistance, and it is also the key to making breakthrough progress in terms of international criminal judicial assistance.

The evidence validity under cross-border evidence collection often becomes the focus of litigation in courts.[60] Evidence validity is a legal issue, which refers to the qualifications and conditions for evidence to be admitted by the court stipulated by the law. The evidence validity is regulated and reflected through the rules of evidence. In criminal judicial assistance, because the parties or other litigation participants often do not appear in court, there are great differences in the legal systems between different places, and some evidence will be excluded due to the lack of evidence validity, which undoubtedly affects the effect of investigating crimes.

According to Paragraph 1 of Article 405 of The Interpretation of the Supreme People's Court on the Application of the Criminal Procedure Law of the People's Republic of China (**最高人民法院关于适用**〈**中**华人民共和国刑事诉讼法〉的解释), the court shall examine evidence materials obtained from abroad in terms of material sources, personnel providing or extracting materials and collection time. After examination, the evidence will be admitted generally if it can prove the facts of the case and comply with the provisions of The Criminal Procedure Law; however, if the source of the evidence is unknown or the authenticity thereof cannot be confirmed, it shall not be used as the basis for fact-finding. When judging cross-border evidence, the court should consider the requirements of hearsay rules and 'illegal evidence exclusion rules', reach a consensus through consultation, and establish corresponding supporting mechanisms to solve the existing problems. On the premise of ensuring litigation justice, the court should simplify the procedures of cross-border evidence collection, ensure the admissibility of relevant evidence, and further improve the efficiency of punishing cross-border crimes.

[60] Linlin Zhao, 'Analysis of Cross-border Evidence Collection in China's Interregional Criminal Justice -- Focusing on the Analysis of Evidence Validity' (2019) 5 JLA 120,127.

4. Protection of Fundamental Rights in the Application of AI

4.1. Fundamental rights infringed in the application of AI

The application of AI in the judicial field has improved judicial efficiency and judicial accuracy to a certain extent, but at the same time, the application of some technologies has also resulted in violations of citizens' fundamental rights, mainly reflected in the violations of citizens' rights to equality, privacy, communications freedom and confidentiality, specifically as follows:

4.1.1. The right to equality

The Constitution of China (宪法) stipulates that 'all are equal before the law'. Citizens should not be treated unfairly because of their nationality, gender, identity and social status. The combination of AI technology and justice not only brings convenience to judicial work but also brings unequal treatment caused by algorithm bias, which infringes on citizens' right to equality.

The application of AI technology in intelligent criminal justice infringes on the right to equality mainly in the trial stage. The algorithm deviation and algorithm black box of intelligent trial-aided systems such as sentencing assistance and similar cases pushing may lead to discrimination to varying degrees; specifically, the defendants who commit the same crime may be subject to different treatment (guilty bias) or unfair trial results; judicial informatization will make the court fully open to the public, and therefore external factors may affect the litigation justice.

Some scholars have pointed out that there will be deviations in the operation of the algorithm due to the algorithm's own factors or sudden errors, that is, algorithm bias, also known as algorithm discrimination, which refers to systematic and repeatable errors that can cause unfair and unreasonable results. The most common example is that the algorithm may produce different results for different people, or produce different results for two people with the same or similar conditions. If algorithm designers deliberately write programs with subjective judgment, algorithm manipulation will occur. The algorithm bias that damages the fundamental rights of the public mainly refers to the algorithm bias that damages the fundamental rights of unspecified subjects. The holders of these rights are uncertain, the intensity of right infringement is unknown, and it is difficult to contain the harmful consequences and for the injured individuals to obtain a remedy, which is mainly manifested in gender discrimination and racial discrimination.[61] Some scholars have pointed out that court informatization has changed the original litigation relationship, which has a certain impact on the rights of citizens involved in litigation, particularly in criminal justice. The defendant's defense rights based on the principles of presumption of innocence and equality between prosecution and defense may encounter

[61] Youhua Liu, 'Research on Algorithm Bias and Its Regulation Path' (2019) 40 LM 56.

difficulties due to the court informatization.[62] Some scholars have pointed out that judicial informatization will inevitably turn the court from semi-closed to fully open to the public so that the court has to consider the extrajudicial and extra-procedural factors emphasized by laymen, which will inevitably erode or reduce the fair trial right of the defendant and the parties. Since limitations shall be imposed on the media coverage of the court, it is even more necessary to limit the openness under judicial informatization. Therefore, the court should seek the opinions of the parties before making the court trial go online.[63]

4.1.2. The right to privacy

The right to privacy is a specific personality right, which refers to a personality right that a natural person may enjoy the peace of a personal life, as well as his personal information are protected according to law, and shall not be illegally disturbed, known, collected, utilized and disclosed by others. In China, although the right to privacy is mainly protected by civil laws such as The Tort Liability Law (侵权责任法), the right to privacy also has its constitutional basis, that is, the concretization of constitutional protection of citizens' personal dignity.[64] In China, AI technology's infringement on citizens' right to privacy is mainly caused by technical investigation and information collection and disclosure by the court.

In the process of litigation, the informatization of the court must involve the storage and use of the information of citizens involved in litigation, which inevitably concerns the personal information rights of relevant citizens. Therefore, the disclosure of case information based on the Internet will inevitably divulge the personal information of citizens; at the same time, relying on the case handling and management platform driven by modern technology and the trial aided system based on big data and AI technology, most of the information collection and use adopt the way of 'black-box operation', and there may also be the problem of illegal collection of personal information.[65] Some scholars, from the perspective of technical investigation measures, have pointed out that technical investigation is carried out with the help of modern technology without the knowledge of the target under investigation, which makes it possible to use technical investigation measures arbitrarily. At the same time, given the nature of the events or activities it actively intervenes in has not yet been determined as a criminal case, this may directly infringe on citizens' right to privacy. Therefore, its infringement on civil rights is even more serious than the traditional investigation means.[66] From the perspective of large-scale monitoring, some scholars have pointed out that China's public security organs are

[62] Xi Zheng,' Conflict and Coordination Between Court Informatization and Citizens' Criminal Procedure Rights' (2020) 42 JJU (PSS) 95,97.
[63]Xiaoxia Sun, 'On the Humanistic 'End' of Judicial Informatization' (2021) 39 LR 34.
[64] Bo Zong,' Legal Regulation of Large-scale Monitoring in Investigation' (2018) 159 JCL 24.
[65] Xi Zheng, 'Conflict and Coordination Between Court Informatization and Citizens' Criminal Procedure Rights' (2020) 42 JJU(PSS) 98.
[66] Dengke Xie, 'On the Protection of Privacy in Technical Investigation' (2016) LF 33,40; Shulin Yang,' On Procuratorial Supervision of Technical Investigation', JSMU 34 (HSSC)105,108.

currently equipped with a strong network monitoring capacity, which can realize the effective monitoring of network information such as online chat, web page content and even e-mail. Of course, citizen privacy will inevitably be involved in this process, making the investigation constitute a compulsory investigation measure.[67] Some scholars have pointed out that if it is used only for the purpose of ensuring judicial justice, can intelligent justice avoid trials that infringe on the parties' personality rights; however, intelligent justice may also infringe on the parties' right to privacy and the right to be forgotten in data collection and calculation. The infringement upon privacy in the era of big data has been discussed many times by the academic community because the calculation of algorithm technology has exceeded human cognition of their own information, which is an infringement upon human privacy. When collecting and processing evidence, intelligent justice should consider the protection of personal information right. Outdated information such as information that is no longer relevant to the identity of the parties, no longer effective and insufficient shall not be used as the basis of judicial trial, and the right to be forgotten of the parties should be respected. This protection of the right to personal information should be designed into the technology of algorithms to avoid the infringement upon the party's right to personal information. Whether it is out of the requirements of judicial fairness, or the protection of the parties' right to privacy and the right to be forgotten when collecting evidence, it is the protection of the parties' right to personality.[68]

4.1.3. The right to communications freedom and confidentiality

Article 40 of The Constitution of China stipulates citizens' right to communications freedom and confidentiality. When classifying the fundamental rights of citizens involved in the application of large-scale monitoring in investigations, some scholars have pointed out that the nature of communications freedom is different from that of communications confidentiality. The right to communications freedom is a right to freedom, which refers to the freedom of citizens to express their wishes through communication tools; the right to communications confidentiality is a right to privacy, which means that citizens express their wishes through letters, telephones, telegrams, faxes, mails, e-mails and the like, which shall not be illegally detained, hidden, opened, recorded, eavesdropped or otherwise obtained by anyone. Therefore, the right to communications confidentiality can be covered by the right to privacy.[69]

Some scholars, based on the legal regulation of German Telecom monitoring, have pointed out that the investigation organ may infringe on citizens' right to communica-

[67] Bo Zong,' Legal Regulation of Large-scale Monitoring in Investigation' (2018) 159 JCL 88.
[68] Hongyang Luo and Xianglong Li, 'Ethical Issues in Intelligent Justice and Their Countermeasures' (2021) 1 PL 148,159.
[69] Bo Zong,' Legal Regulation of Large-scale Monitoring in Investigation' (2018) 159 JCL 88.

tions confidentiality and personal information security when conducting Telecom monitoring.[70] Other scholars have studied China's procedural regulation of electronic evidence collection from the perspective of personal information protection, and put forward that the framework of citizens' 'personal information right' is the right to human dignity, communications confidentiality and freedom and protection against illegal search.[71]

It can be seen that the infringement upon citizens' rights to communications freedom and confidentiality mainly occurs in the application of AI technology in the investigation, such as telecommunication monitoring, network monitoring and e-mail detain against criminal suspects. In this sense, the application of AI technology in the investigation should comply with due process and the principle of Legality and Proportionality, guarantee the subject's right to be informed and establish a comprehensive supervision system.

4.1.4. The right to freedom of expression

The right to freedom of expression refers to the right enjoyed by citizens to use various media and ways to publicly publish and transmit their opinions, points, views and emotions, which are regulated, recognized and protected by law, without interference, restriction or infringement by any other person or organization. The right to freedom of expression mainly includes: freedom of speech, freedom of press and publication, freedom of artistic expression and freedom of assembly. Some scholars have pointed out that the use of large-scale monitoring in investigations has a direct and indirect impact on freedom of expression. The direct impact includes: the filtering and interception of specific information by investigation organs through large-scale monitoring will directly infringe on people's right to freedom of expression; indirect impact includes: if citizens know that the investigation organ can use large-scale monitoring without restriction, and can use the information so obtained as criminal evidence against them, or use such information improperly, it will inhibit citizens' motivation to express their opinions, demands and suggestions through various channels.[72]

To sum up, focusing on the research and analysis of infringement on specific rights, the application of AI to the field of criminal procedure may infringe on citizens' fundamental rights, mainly the right to equality and personality; specifically, the right to personality involves the right to privacy, personal information protection, communications freedom and confidentiality, personal freedom, and the right to protection against illegal search.

---

[70] He Huang, 'On the Legal Regulation of German Telecom Monitoring - Analysis Based on Fundamental Rights' 131 (2017) JCL 88,101.
[71] Yong Jiang, 'Procedural Law Turn of China's Electronic Evidence Collection Regulation from the Perspective of Personal Information Protection' 39 (2019) JJU (SS) 141,142.
[72] Bo Zong,' Legal Regulation of Large-scale Monitoring in Investigation' (2018) 159 JCL 89.

4.2. Protection of fundamental rights in the application of artificial intelligence

The application of AI in the field of criminal procedure has a positive significance. Therefore, some measures should be taken to regulate the application of AI technology in order to protect the fundamental rights of citizens.

In view of the infringement on the right to equality caused by algorithm black box and algorithm bias, a scholar proposed the introduction of a 'class action system'. The scholar believed that racial discrimination and gender discrimination caused by the use of algorithms may result in differential treatment for groups of specific races and different genders. Although The Constitution of China clearly stipulates that gender discrimination and racial discrimination are prohibited, it fails to specify the specific behavior mode and legal consequences; other laws and regulations only provide for principled provisions, without much actionability and operability. Such differential treatment is mostly reflected in resume screening and judicial prediction. Algorithm bias can be secretive, and algorithm-driven decisions are difficult to be understood by algorithm service recipients, resulting in the inability of algorithm service recipients to safeguard their legitimate rights and interests through private remedy. Given the use of algorithms is repetitive and universal, it is prone to repeated use by the public, and therefore the foregoing problem can be solved by introducing the class action system. Before filing a class action, you can first submit a written request to the algorithm user to explain the decision made. If the algorithm user revokes the decision and corrects it, the parties may settle the dispute. If the algorithm user refuses to explain the decision made, the group subject to discrimination can file a class action.[73]

In view of the inequality of litigation rights brought by court informatization, some scholars have proposed to ensure the equality of prosecution and defense through information isolation and information disclosure. The term 'information isolation' refers to shielding the information with an obvious tendency and not suitable for the judge to know. The term 'information disclosure' refers to the fact that the information unfavorable to the defense is fully disclosed to the defense.[74]

In view of the infringement upon fundamental rights and interests such as reasonable expectations of privacy in electronic data collection (online remote inspection, online extraction and electronic data freezing), some scholars have put forward three countermeasures: the categorization of electronic data collection based on fundamental rights, the constitutional adjustment of mandatory investigation measures in electronic data collection, and the establishment of illegal electronic data exclusion rules.[75]

In view of the infringement upon citizens' 'personal information right' in the process of electronic evidence collection, some scholars have pointed out that in the procedural

[73] Youhua Liu,' Research on Algorithm Bias and Its Regulation Path' (2019) 40 LM 65,66.
[74] Zheng Xi, 'Conflict and Coordination Between Court Informatization and Citizens' Criminal Procedure Rights' (2020) 42 JJU(PSS) 98.
[75] Dengke Xie,'On the Protection of Rights in Electronic Data Collection' (2020) 12 LAJ 14.

structure focusing on the protection of rights, all mandatory measures shall be subject to the due process requirement. Although there are special evidence collection techniques and carriers for electronic evidence collection, its procedural legal basis is still under the scope of due process. They also put forward three procedural improvement paths: the systematization of electronic evidence collection measures, the proportionality of electronic evidence collection procedures and the appropriate role of judicial review.[76]

In view of the protection of the right to privacy in technical investigation, some scholars have put forward three countermeasures: first, we should establish a judicial review system for the initiation of technical investigation and properly control the power against the right to privacy; second, we should refine the application standards of technical investigation and strengthen the reasonable expectation to privacy protection; third, we should clarify the procedural sanctions against illegal technical investigation and improve the institutional rigidity of privacy protection.[77]

In view of the infringement upon citizens' fundamental rights by using large-scale monitoring in investigations, some scholars have proposed that the existing investigation theories and norms must be revised, the case filing system should be reformed, and the target scope of technical investigations should be expanded; different regulations should be made according to the purpose and content of large-scale monitoring; the regulation of the use of large-scale monitoring in investigations should be carried out from two aspects: procedural norms and evidence rules; the former includes the scope of application, conditions of application, applicable subjects, approval procedures and implementation procedures, while the latter includes the exclusion rules of illegal evidence obtained by large-scale monitoring and the exclusion rules of unreliable evidence set according to the technological features of large-scale monitoring.[78]

Selected literature

'Beijing "Smart Judge" Promoting Similar Judgments for Similar Cases '(People's Court Daily, 1 September 2017): <http://rmfyb.chinacourt.org/paper/html/2017-09/01/content_129653.htm> accessed18 December 2021

'Guizhou Political and Legal Organs Solidly Promote the Deep Integration of Technological Innovation and Judicial System Reform -- Accurate and Fair Case Handling Driven by Big Data 'People's Daily (Beijing, 10 July 2017)

[76] Yong Jiang,' Procedural Law Turn of China's Electronic Evidence Collection Regulation from the Perspective of Personal Information Protection' 39 (2019) JJU (SS) 62,63.
[77] Dengke Xie,' On the Protection of Privacy in Technical Investigation' (2016) LF 39.
[78] Bo Zong,' Legal Regulation of Large-scale Monitoring in Investigation' (2018) 159 JCL 104.

'Let Modern Technology Better Help Judicial Reform -- Hainan Intelligent Sentencing System Operates "Faster, Better and More Cost-effectively"' People's Court Daily (Beijing, 9 December 2017)

'New Highlights of Judicial Reform: Power and Responsibility Unification under the Judicial Accountability System': <http://llxfy.chinacourt.gov.cn/article/detail/2016/07/id/2042348.shtml> accessed 18 December 2021

'Shanghai's Application of "Artificial Intelligence" in Case Handling to Prevent Wrongful Conviction, the Launch of China's First "Intelligent Case Handling Aided System' Legal Daily (Beijing, 11 July 2017)

'Upgrading the Informatization Construction of Jiangsu 'Smart Courts' Injects New Impetus into the Modernization of Judicial Capacity 'Legal Daily (Beijing, 20 March 2017)

Cai G, 'Origin, Development and Function of Criminal Evidence Standard Guidance' (2021) 306 SSS 187

Cui X, 'Towards Intelligentization: The Practical Path of Artificial Intelligence Embedded in Procuratorial Work Reform' (2021) 290 SS 132，137

Ding G, 'The Construction of 'Smart Courts' in Jiangsu Injects New Impetus into the Modernization of Judicial Capacity' Legal Daily (Beijing, 20 March 2017) 17.

Dong K, 'Evidence Standard: Connotation Reinterpretation and Path Prospect' (2020) 19 CLR 109,118

Gao T, 'Research on Quantitative Assessment of Social Risk of Arrest -- From the Perspective of Automated Decision-making and Algorithmic Regulation' (2021) 15 NLS 135

Gao X, 'Chinese Justice in the Era of Artificial Intelligence' (2019) 49 JZU (HSSE) 229,237

Huang H, 'On the Legal Regulation of German Telecom Monitoring - Analysis Based on Fundamental Rights' 131 (2017) JCL 88,101

Huang J, 'Negative List of Criminal Justice Artificial Intelligence' (2017) 10 EFV 85,94

Jin C, 'Analysis on the Case Division Mechanism Driven by Artificial Intelligence' (2020) 76 JCUPSL 171

Jiang Y, 'Procedural Law Turn of China's Electronic Evidence Collection Regulation from the Perspective of Personal Information Protection' 39 (2019) JJU (SS)62,63 141,142

Liu P, 'On Big Data-Driven Evidence' (2019) 41 GLR 25

Liu Y, 'Research on Algorithm Bias and Its Regulation Path' (2019) 40 LM 56,65,66

Luo H and Li X, 'Ethical Issues in Intelligent Justice and Their Countermeasures' (2021) 1 PL 148,159

Ma G, 'On the Examination of AI-driven Evidence in Criminal Procedure' (2021) CS 175

Sun D, 'Artificial Intelligence Assisted Accurate Prediction of Sentencing in China -- Taking Plea for Leniency Cases as the Applicable Field' (2020) 42 JJU 76,77

Sun Q,' Promoting the Deep Integration of Procuratorial Work and New Technology, Effectively Improving the Quality and Efficiency of Case Handling and Judicial Credibility' (2017) 752 PPS 7

Sun X, 'On the Humanistic "End" of Judicial Informatization' (2021) 39 LR 34

Pan Y, 'Path Analysis of Artificial Intelligence Intervention in the Judicial' (2018) 3 OL109,118; Han Qin, 'Theoretical Reflection on AI-enabled Judicial System' (2021) 15 NLS 115,129

Wei B, 'Difficulties and Paths of Integrating Judicial Artificial Intelligence into Judicial Reform' (2021) 43 MLS 4

Wei C, 'Power Logic of Artificial Intelligence in Criminal Trial' (2021) 41 JXJU(SS) 147

Wang L,' Technical Barriers to the Development of Judicial Big Data and Artificial Intelligence' (2018) 20 CLR 46,48

Wang R, 'Research on Judicial Supervision Mechanism of Big Data' (2021) 24 HUST (SSE) 132,136

Wang X and Tang L, 'Application and System Design of Artificial Intelligence in Preventing Wrongful Conviction' (2021) 42 LM 100

Wang Z,' Principle and Construction of Quantitative Model for Review of Social Risk Assessment of Arrest' (2016) 34 PLF 73,74

Wu S, 'The Dilemma and Transformation of China's Plea Bargain Mode -- from "Confirmation and Approval Mode" to "Negotiation and Review Mode"' (2020) 1 CS 154

Xie D, 'On the Protection of Rights in Electronic Data Collection' (2020) 12 LAJ 14

—' On the Protection of Privacy in Technical Investigation' (2016) LF 33,39,40

Xie S, 'How Can Artificial Intelligence "Unbiasedly" Help Criminal Justice -- From "Evidence Guidance" to "Proof Assistance"' (2020) 38 JNUPSL 109,117

Xiong Q, 'Application of Artificial Intelligence in Criminal Proof' (2020) 34 CLR75,79,88

Xu H and Li X, 'Research on Evidence Attribute Verification of Big Data-Driven Evidence' (2020) 36 JPPSUC (SSE) 50

Yang T, 'Rationality and Limit Analysis of Digital Evidence Standard -- Focusing on Shanghai "206" Intelligent System' (2020) 47 JSNU 45

Yang Y, 'Problems and Optimization in the Practice of Sentencing Suggestions in Plea for Leniency Cases' (2020) 312 AE 93

Yuan Y, 'Attribute and Objective Verification Standard of Binary Physical Evidence of Big Data-Driven Evidence' (2021) 44 JSU(PSSE) 143

Zhang F and Zheng H, 'Positioning, Prospect and Risk Prevention and Control of Artificial Intelligence Assisted Sentencing in the Era of Big Data' (2019) 283 GSS 92,100

Zhang M, 'Embracing the New Technology of Intelligent Voice and Creating a New Engine of Intelligent Procuratorial Work' (2017) 753 PPS 28

Zhang S, 'Application of Artificial Intelligence in Trial' (2020) 49 JSNU (PSSE) 102,110

Zhang Y, 'Judicial Application of Intelligent Sentencing Algorithm: Logic, Problems and Procedural Law Response' (2021) 81 OL 187,199

Zhao L, 'Analysis of Cross-border Evidence Collection in China's Interregional Criminal Justice -- Focusing on the Analysis of Evidence Validity' (2019) 5 JLA 120,127

Zheng F and Ma G, 'Triple Dilemma and Way Out of Big Data-Driven Evidence Application' (2020) 28 JCU(SSE) 208

Zheng X,' Conflict and Coordination Between Court Informatization and Citizens' Criminal Procedure Rights' (2020) 42 JJU (PSS) 95,97,98

——'Application and Regulation of Artificial Intelligence Technology in Judicial Adjudication' (2020) 32 PULJ,647,696

Zhong M, 'Practical Observation and Prospect of "Artificial Intelligence + Online Judicial Confirmation"' (2020) 15 JLA 122,130

Zhu Hand Liu C, 'Research on the Application of Big Data in the Trial of Similar Cases' (2019) 20 JLA 47,54

Zhuang Y, 'How to Build a Case Handling Performance Evaluation Mechanism for Procurator Quota System' (2017) 753 PPS 48

Zong B, 'Analysis on the Application of Artificial Intelligence in Criminal Evidence Judgment' (2018) 37 JNUPSL 61,68

——Legal Regulation of Large-scale Monitoring in Investigation' (2018) 159 JCL 24,88,89,104

Zuo W, 'Some Thoughts on the Application Prospect of Legal Artificial Intelligence in China' (2018) 12 TLJ 108, 124

AI SYSTEMS AND EVIDENCE LAW IN FINLAND*

*By Juhana Riekkinen** and Sofia Söderholm****

Abstract

In this report, we analyse how Finnish law applies to the use of AI in evidence-gathering and to AI-based evidence in criminal proceedings, and whether AI could be used to assess criminal evidence. As there are no statutes specifically on the use of AI in policing or criminal proceedings, these issues are governed by more general statutes, which largely adhere to the principle of technology neutrality. Evidence law is based on the free theory of evidence, which means that AI-based evidence is generally admissible. Predictive policing systems are currently not used in Finland, but the admissibility of AI-produced predictions would in any case be limited, as evidence that is not relevant to the facts of the case should be rejected. Further, unlawfully obtained AI-based evidence should be excluded if it might endanger the right to a fair trial. AI systems are not used to assess evidence in courts, and the use of fully automated systems is not permissible.

1 Introduction

Artificial intelligence ('AI') has numerous potential applications in the field of law enforcement and criminal justice. In particular, AI-based systems could be utilised in different ways in gathering, producing, and assessing evidence. The aim of this report is to provide an overview on the use of AI-based systems related to criminal evidence in Finland, and to examine the Finnish legal norms that apply to the use of AI in the context of evidence in criminal proceedings.

In Chapter 2, we analyse how Finnish law applies to the use of AI-based systems in gathering evidence. Then, in Chapter 3, we present the main features of Finnish evidence law norms and analyse what they mean for evidence that has been produced with the help of AI-based systems. In connection, although *predictive policing*[1] is still in its infancy in

* This report is mostly based on the Finnish responses to part 3 (Evidence Law) of the AIDP Questionnaire – Section III on AI and Administration of Justice: Predictive Policing and Predictive Justice. The original report concerning part 3 was prepared by Juhana Riekkinen. The report concerning part 1 (Predictive Policing) was prepared by Sofia Söderholm, who contributed to the final form of the present text, in particular in relation to questions relating to predictive policing.

** University Lecturer in Legal Informatics, University of Lapland, Faculty of Law (juhana.riekkinen@ulapland.fi).

*** Doctoral Researcher, Helsinki University, Faculty of Law (sofia.soderholm@helsinki.fi).

[1] For the concept of predictive policing, we refer to the much-cited definition by Walter L Perry, Brian McInnis, Carter C Price, Susan C Smith and John S Hollywood, *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations* (1st edn, RAND Corporation 2013) 30: 'Predictive policing is the application of analytical techniques – particularly quantitative techniques – to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions.' In addition, essential elements of predictive policing technology are exploitation of big data (e.g., Elizabeth E. Joh, 'Feeding the Machine: Policing, Crime Data, & Algorithms' (2017) 26 *Wm. & Mary Bill Rts. J.* 287) and AI.

Finland, we discuss AI-based predictive policing systems, which may directly or indirectly produce information with probative value. In Chapter 4, we assess the legality of automation of judicial decision-making and whether AI-based systems could play some role in assessing criminal evidence or supporting judges in this task. In each of these Chapters, we briefly discuss the current state of use of AI-based systems in these activities in Finland, and further point out some potential future developments and discussions at the national level. However, as public information on these matters is somewhat scarce, the main focus of this report remains on the legislation. Finally, in Chapter 5, we summarise our key findings.

2 Evidence-gathering through AI-based systems

2.1. Use of AI-based systems in practice

There is very little information publicly available on the use of AI-based systems for evidence-gathering purposes in Finland. Within law enforcement organisations, the National Bureau of Investigation ('NBI') has significant digital forensics capabilities. The NBI Forensic Laboratory assists and supports other law enforcement units by performing forensic analyses of various kinds, including digital forensics. Further, the NBI has a unit focused on cybercrime prevention and investigations (Cybercrime Centre). While the exact operational capabilities of these units are not public, it is highly likely that they have access to state-of-the-art digital forensics tools with AI-based features, including tools for technology-assisted review of documents and mobile device forensics.

In the private sector, digital forensics services are offered by several companies that operate in the cybersecurity/ICT field and major accounting firms, and it is likely that some of these offer services featuring the use of some AI-based tools for the purposes of evidence-gathering and analysis.[2] However, there are no statistics or research that would indicate to what extent these services are used by private companies or law firms in Finland.

2.2. Relevant normative frameworks

As we move to discuss the legal conditions of AI-assisted or AI-enabled digital forensics investigations, it should be noted that the European data protection framework fully applies to data processing by Finnish public organisations and private companies. Most police data processing activities are governed by the Law Enforcement Directive ((EU) 2016/680, 'LED')[3] and the national Act on the Processing of Personal Data in Criminal

[2] At least one Finnish digital forensics firm has previously advertised partnership with providers of forensic software with known AI-based features, such as OpenText (EnCase Forensic) and Cellebrite (UFED) (Difseco Oy, OpenText and Other Services, <https://difseco.com/other-services/> accessed 28 March 2022; content no longer available in May 2023).

[3] Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of

Matters and in Connection with Maintaining National Security (1054/2018, 'Act on the Processing of Personal Data in Criminal Matters').[4] This act grants the Police and other competent criminal justice authorities the permission to process personal data when it is necessary for performing their duties related to, *inter alia*, prevention, detection, investigation, and prosecution of criminal offences.[5] Besides providing the legal basis for processing, this act regulates and sets limits on data processing activities by these authorities. The Act on the Processing of Personal Data by the Police (616/2019) complements the Act on the Processing of Personal Data in Criminal Matters, and in part, the General Data Protection Regulation ((EU) 2016/679, 'GDPR'),[6] which applies as a *lex generalis* to police data processing that is beyond the scope of the LED. Public data processing activities beyond the scope of the LED and all private data processing activities are subject to the GDPR and the Data Protection Act (1050/2018), which complements the directly applicable EU regulation on the national level.[7]

The directly AI-related prohibitions on fully automated individual decision-making in the GDPR (Article 22) and the Act on the Processing of Personal Data in Criminal Matters (Section 13) do not prohibit evidence-gathering with the help of typical AI-assisted digital forensics tools. Evidence-gathering activities do not produce the kind of decisions with significant effects intended in these provisions, and such evidence-gathering is typically subject to human oversight. More generally, however, AI-assisted evidence-gathering and digital forensics investigations necessarily involve processing of personal data. Therefore, data protection law needs to be fully complied with, regardless of whether these investigations are performed by public authorities or private companies.

Beyond data protection legislation, there is currently no normative framework explicitly governing the use of AI-based systems for gathering evidence.[8] However, respecting fundamental rights, such as the rights to equality, privacy, and liberty and security, are all regulated in the Constitution of Finland (731/1999, 'Constitution'). Furthermore, Section

criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA.

[4] Finnish parliamentary acts are officially available in Finnish and Swedish. A collection of unofficial translations to other languages is available at Finlex, Translations of Finnish acts and decrees: <https://www.finlex.fi/en/laki/kaannokset/> accessed 28 March 2022.

[5] Act on the Processing of Personal Data in Criminal Matters, ss 4(1) and 1(1).

[6] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

[7] The Border Guard, the Customs and the Defence Forces have their own *lex specialis* data processing acts: Act on the Processing of Personal Data by the Border Guard (639/2019), Act on the Processing of Personal Data by the Customs (650/2019) and Act on the Processing of Personal Data by the Defence Forces (332/2019).

[8] If approved, the EU AI Act will be directly applicable in Finland.

21 of the Constitution constitutes a right to an effective remedy, fair trial, and good administration, which is always relevant when exercising public power.[9] According to Section 22 of the Constitution, public authorities must guarantee the observance of basic rights and liberties and human rights. Thus, all public officials, including police officers, are obliged by the Constitution to guarantee the fulfilment of these rights in their activities.

The principle of legality, enshrined in Section 2(3) of the Constitution, requires that the exercise of public powers shall be based on an Act. In the absence of a parliamentary act granting law enforcement authorities the power to use AI-based systems in evidence-gathering, it could be argued that the use of such systems is illegal. However, Finnish legislation generally adheres to the principle of *technology neutrality*, focusing more on functions and purposes than on specific technologies. This means that the use of AI-based systems by public authorities may be based on more general laws and provisions that do not specifically recognise AI, AI-related technologies, or individual AI-based systems. Therefore, to determine the legal limits of AI-enabled evidence-gathering by the Police, the general normative framework that enables law enforcement authorities to conduct criminal investigations should be considered. The bulk of this framework consists of the Criminal Investigation Act (805/2011), the Coercive Measures Act (806/2011), and the Police Act (872/2011).[10]

The conduct of criminal investigations is regulated in the Criminal Investigation Act. This *lex generalis* is complemented by the Coercive Measures Act, which governs the use of coercive measures that, *inter alia*, allow the Police to gather evidence of suspected criminal offences. These measures include different types of searches and seizure (Chapters 7 and 8), but also a multitude of covert investigatory powers for surreptitious monitoring of telecommunications and technical devices (Chapter 10). The Police Act contains provisions on covert measures similar to those regulated in Chapter 10 of the Coercive Measures Act. These powers may be used for the purposes of prevention and detection of crime and for civilian intelligence, as opposed to the Coercive Measures Act powers, which enable the Police to investigate criminal offences that have already been committed (or are suspected to have been committed).[11]

---

[9] Constitution, s 21: '(1) Everyone has the right to have their case dealt with appropriately and without undue delay by a legally competent court of law or other authority, as well as to have a decision pertaining to his or her rights or obligations reviewed by a court of law or other independent organ for the administration of justice. (2) Provisions concerning the publicity of proceedings, the right to be heard, the right to receive a reasoned decision and the right of appeal, as well as the other guarantees of a fair trial and good governance shall be laid down by an Act.'

[10] Additionally, *lex specialis* acts govern some aspects of crime prevention and criminal investigations by the Border Guard (108/2018), the Customs (623/2015), and in the Defence Forces (255/2014).

[11] Generally on investigatory powers in Finnish law from the viewpoint of digital investigations, see Juhana Riekkinen, 'Evidence of Cybercrime and Coercive Measures in Finland' (2016) 13 *Digital Evidence and Electronic Signature Law Review* 49, 50, 55–66.

Powers defined in the Coercive Measures Act allow the Police to gain access to, copy, or record computer data (including content data and metadata) from various sources, including devices used by suspects and third parties. In keeping with the principle of technology neutrality, the Coercive Measures Act does not define which kind of hardware or software tools or methods may be used for executing these investigatory measures. There are no explicit mentions of AI or AI-based tools, and the addition of specific AI-related rules seems unlikely in the near future.[12] The focus is on regulating decision-making procedures and conditions of access to potential evidence, and the further analysis of data that has been lawfully confiscated (along with a physical medium) or copied to be used as evidence is largely unrestricted. The Coercive Measures Act does not explicitly prohibit any analytical methods or tools, and it should not be interpreted as precluding the use of AI-based software in general. However, institutionalised legal principles such as *proportionality*, *minimum intervention*, and *sensitivity*,[13] and provisions safeguarding legal privileges[14] may limit the use certain tools, methods, or means subject to case-by-case analysis. In particular, this is relevant for searches targeting devices used by certain groups of professionals, such as lawyers, medical professionals, and journalists.[15] These principles and rules may even limit the use of non-AI-related digital forensics practices, such as the creation of forensic duplicates (bit-for-bit copies) of the entire contents of storage media.[16]

Private companies are not bound by the constitutional principle of legality. However, to be able to conduct (or to authorise a third party to conduct) forensic investigations, they must have lawful access to the potential evidence to be analysed. Generally, private parties may not conduct investigations with methods comparable to investigatory powers

[12] A working group set by the Ministry of Justice recently published a report concerning the needs to amend the Coercive Measures Act. This report only mentioned AI in relation to real-time biometric identification in connection with certain covert coercive measures. The working group opted not to prepare any draft provisions on this (largely due to uncertainty regarding the impact of the upcoming EU AI Act). Lauri Rautio et al., *Pakkokeinolain muutostarpeiden tarkastelu: Työryhmämietintö* (Ministry of Justice 2022) 65–66. In the Government Proposal that followed (HE 217/2022 vp), no AI-related provisions were discussed or proposed (law drafting documents are only available in Finnish and Swedish).

[13] Coercive Measures Act, c 1 ss 2–4. Criminal Investigation Act, c 4 ss 4–6 and Police Act, c 1 s 2 define similar principles.

[14] Coercive Measures Acts, c 7 s 3 contains prohibitions on confiscation and copying of privileged material, with references to provisions on right or duty not to testify in CJP, c 17. Provisions on 'special' searches which are likely to involve privileged material are located in Coercive Measures Act, c 8, and may apply to searches of data contained in a device (through a reference in c 8 s 28).

[15] These 'special' searches are subject to specific conditions and procedural rules, including the appointment of an independent representative, who is tasked with supervising the search procedure and making sure that no privileged material is searched or copied.

[16] Alternatively, it can be argued that forensic imaging is permissible regardless of the contents, and that the provisions that prohibit copying of privileged material should in these cases be interpreted as exclusionary rules that forbid further analysis and evidentiary use of any such material that is included in the forensic duplicate. Black-letter law and law drafting materials do not provide clear answers, and although there is some recent case law concerning the practicalities of 'special' searches targeting devices with privileged data, the legal situation remains unclear. See Juhana Riekkinen, *Sähköiset todisteet rikosprosessissa* (Alma Talent 2019) 240–247.

defined in the Coercive Measures Act or the Police Act, regardless of whether AI-based tools are used or not. Private investigations targeting devices and computer data that are not under the lawful control of the investigating party may trigger criminal liability. Potentially applicable provisions of the Criminal Code (39/1889) include computer break-in[17] and message interception[18].[19]

If the private party conducting the investigation has access to a device on which potentially relevant data are stored, they need to consider data protection obligations. For private entities, the law imposes no general duty to investigate crime, and therefore they do not have a similar general legal basis for processing crime-related personal data as the Police do. However, Article 6(1)(f) of the GDPR recognises the legitimate interests of the controller or a third party as a legal ground for processing, and Article 9(2)(f) allows the processing of even sensitive 'special categories of personal data' for the establishment, exercise, or defence of legal claims. While these provisions provide a legal basis for data processing in most scenarios where digital forensics investigations are performed, the investigating party needs to adhere to all of the data processing principles defined in Article 5 (including *purpose limitation* and *data minimisation*) and other duties and obligations specified in data protection law. This may limit the use of data-intensive AI-based tools.

If the potential evidence contains personal data of employees, the employer's evidence-gathering activities may be further restricted by the Act on the Protection of Privacy in Working Life (759/2004). Processing of e-mails and other data related to electronic communications is also subject to the Act on Electronic Communications Services (917/2014), which contains provisions implementing the ePrivacy Directive (2002/58/EC)[20]. In general, parties to electronic communications are entitled to process their own messages and traffic data, and may also give consent to other parties to engage in such processing (the consent of one party is sufficient).[21] The aforementioned act also regulates the conditions under which communications service providers and 'corporate or association subscribers'[22] may process traffic data for the purposes of investigating suspected misconduct and criminal offences.[23] As a result of these rules, the permissibility of large-scale AI-based document review targeting the contents of employee e-mail accounts is highly

[17] Criminal Code, c 38 ss 8–8a.

[18] Criminal Code, c 38 ss 3–4.

[19] As elaborated later, criminal acts by private parties may trigger exclusion of evidence obtained by such means in a subsequent trial. Compared to unlawful evidence-gathering by public authorities, criminal acts by private parties are less likely to trigger the exclusionary rule under CJP, c 17 s 25(3).

[20] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications).

[21] Act on Electronic Communications Services, s 136.

[22] This is defined as 'an undertaking or organisation which subscribes to a communications service or an added value service and which processes users' messages, traffic data or location data in its communications network' (Act on Electronic Communication Services, s 3(41)).

[23] Act on Electronic Communications Services, ss 143, 145a and 146–156.

questionable at best, and usually clearly illegal. AI-based review and analyses limited to communications metadata may be permissible.

2.3. Informational rights of the defendant

While there are no specific procedural rules concerning AI-based systems and information relating to the use of such systems, Finnish law grants the defendant informational rights allowing them wide access to information relevant to their case. These rights could be interpreted to cover some information relating to methods and tools of evidence-gathering, including if and how AI-based systems have been used in the investigation and how these systems operate. The informational rights of the defendant are governed by the European Convention on Human Rights ('ECHR'), the Constitution, the Criminal Investigation Act, and the Act on the Openness of Government Activities (621/1999, 'Openness Act').[24]

The principles of *audiatur et altera pars* and *equality of arms* are acknowledged in Finnish law. While these two principles are closely linked and sometimes even considered to be one and the same, they can be distinguished from each other.[25] *Audiatur et altera pars* is a foundational principle of procedural law: each party should have a chance to be heard. This includes a chance to present evidence, as well as to challenge and comment on evidence presented by other parties, which necessitates access to such evidence. The principle is incorporated in numerous provisions in different parliamentary acts, and it applies in virtually all court proceedings and in administrative decision-making.

Instead, *equality of arms* is more specifically a principle applicable in criminal proceedings, largely concretised by the case law of the European Court of Human Rights ('ECtHR') concerning Article 6(3) of the ECHR. From the point of view of informational rights and evidence, Finnish commentators have emphasised that this principle requires that the defendant is granted access not only to prosecution evidence but also to material which has not been named as evidence by the prosecution, but has surfaced during the investigation and may help the defence case. It should be possible for the defendant to gather evidence from the same 'pool of potential evidence' that the criminal justice authorities have access to, including sources that have been left out from the official police

[24] Data protection law (GDPR, Article 15 and Act on the Processing of Personal Data in Criminal Matters, s 23) contains further informational rights that may, under specific circumstances, enable the data subject to receive information that may be relevant in the context of criminal proceedings and thus complement the informational rights discussed here.

[25] See Laura Ervo, Oikeudenkäynnin oikeudenmukaisuusvaatimus: Käsikirja lainkäyttäjille (WSOYPro 2008) 133–136.

protocol.[26] Mere technical possibility of access is not sufficient to guarantee true participation in the proceedings, and the defendant should also be provided with adequate time and facilities as required by Article 6(3)(b) of the ECHR.[27]

The informational rights of parties, including defendants, are governed by Chapter 4, Section 15 of the Criminal Investigation Act. As a main rule, parties have the right to information on matters that have led to or become apparent in the criminal investigation, and any documentation of the criminal investigation that may affect or could have affected the consideration of their matter. This right exceeds the general right of access to public documents,[28] applying also to documents and information that are to be kept secret under the Openness Act. However, there are some exceptions. Notably, the right may be denied if this is necessary to secure a very important public or private interest.[29] This exception may be applicable to some information regarding technical and tactical law enforcement capabilities, possibly including information concerning specific features of AI-based systems (e.g., digital forensics tools) used in criminal investigations. However, when considering such restrictions, 'consideration shall be taken in the assessment of the right of the party to a proper defence or otherwise to appropriately secure their right in the court proceedings'.[30] In practice, any restrictions must be evaluated in accordance with the requirements of ECtHR case law concerning Article 6(3) of the ECHR.[31] Of course, even law enforcement authorities may not have full access to all information regarding proprietary AI-based tools that they use, in which case such information may practically remain unavailable to the defendant, as well.

There is no published case law relating to evidence-gathering through AI-based systems, or the interpretation of informational rights of the defendant or equality of arms in relation to their use. Furthermore, there is practically no legal commentary that would specifically address these AI-related issues in Finland.

3. **Evidence produced by AI-based systems**

3.1. Use of AI-based systems for production of evidence

Data processing acts applicable to Finnish law enforcement authorities permit the use of facial recognition technology for the purposes of preventing, detecting, and investigating

[26] Markku Fredman, *Rikosasianajajan käsikirja* (2nd edn, Alma Talent 2021) 476.

[27] Generally on equality of arms in Finnish legal literature, see, e.g., Ervo (n 25) 155–157, 262–264, 291; and Matti Pellonpää, Monica Gullans, Pasi Pölönen and Antti Tapanila, *Euroopan ihmisoikeussopimus* (6th edn, Alma Talent 2018) 616–621.

[28] Openness Act, s 9. Access to documents in the possession of public authorities is also guaranteed as a constitutional right (Constitution, s 12(2)).

[29] Criminal Investigation Act, c 4 s 15(3).

[30] Criminal Investigation Act, c 4 s 15(4).

[31] Further, there are limitations to the right of access relating to covert investigations and the ways in which access to audio and video recordings may be granted (Coercive Measures Act, c 10 ss 60 and 62, Police Act, c 5 ss 58 and 60 and Criminal Investigation Act, c 9 s 7(2)). A detailed account on the defendant's right to information can be found in Markku Fredman (n 26) 466–515.

criminal offences. Authorities may compare faces extracted from, e.g., surveillance camera recordings to photographs in various police databases.[32] A specific automated facial recognition system (KASTU) has been developed for police use. The use of this system began in May 2020, and details about its features are not public.[33] The system is mainly intended to be used as a tool for directing investigations and finding likely matches to be confirmed by further analysis. It is not intended to be used (or particularly suitable) for producing evidence directly.[34]

As already discussed above, data protection law sets limits to evidence-gathering and investigations with the help of AI-based systems, and may prohibit or limit the use of certain kinds of facial recognition systems. For instance, current law arguably does not permit the use of real-time automated facial recognition systems in connection with live video streams and the coercive measures of technical observation and extended surveillance.[35] Furthermore, the use of the controversial *Clearview AI* facial recognition application has been deemed unlawful by the Finnish data protection authority. This application was trialled without a specific legal basis by the Child Abuse Material/Child Sexual Exploitation unit of the NBI in early 2020; approximately 120 queries were made during the trial.[36] The Finnish Data Protection Ombudsman issued a reprimand to the NBI for unlawful processing of personal data.[37] Apparently, the queries did not lead to any information generated by the Clearview AI application being used as evidence in Finnish courts.

Although Finnish law enforcement authorities have shown considerable interest in automated facial recognition and other AI-based technologies in recent years, as far as the authors are aware, they do not regularly employ any notable AI-based systems to produce evidence for the purposes of criminal trials. Some evidence produced with the help

[32] Act on the Processing of Personal Data by the Customs, s 14(2) specifically mandates such comparison through automated facial recognition. The Act on the Processing of Personal Data in Criminal Matters and the Act on the Processing of Personal Data by the Police do not specifically mention automated facial recognition, but regulate the use of special categories of data in police registries for various purposes, including prevention, detection, and investigation of crime. The use of biometric data is generally permitted only when it is necessary (ss 11 and 15 of these Acts, respectively).

[33] Simo Ortamo, 'Poliisi on saanut rikollisia kiinni kasvoja tunnistavan tekoälyn avulla ja haluaisi laajentaa valtuuksiaan – testasimme, miten kone toimii' *Yle Uutiset* (1 August 2020): <https://yle.fi/uutiset/3-11448002> accessed 28 March 2022.

[34] However, as elaborated below, there are no evidence law rules that would specifically bar such evidence, or any other type of AI-produced evidence.

[35] Rautio et al. (n 12) 65.

[36] 'Testing of facial recognition software by NBI reported to Data Protection Ombudsman' (9 April 2021): <https://poliisi.fi/en/-/testing-of-facial-recognition-software-by-nbi-reported-to-data-protection-ombudsman> accessed 28 March 2022.

[37] Data Protection Ombudsman, Decision, 20 September 2021, 3394/171/21. Further, the NBI were ordered to request the service provider to delete any personal data relayed to it by the NBI through the use of the Clearview AI software. See 'Police reprimand from Deputy Data Protection Ombudsman – police have initiated measures ordered' (28 September 2021): <https://poliisi.fi/en/-/police-reprimand-from-deputy-data-protection-ombudsman-police-have-initiated-measures-ordered> accessed 28 March 2022.

of AI-based systems may be proffered in individual court cases, but AI-produced evidence remains a largely unrecognised phenomenon in Finnish courts.

In the course of typical modern criminal investigations, large volumes of data and other materials are gathered. However, it may be difficult to pinpoint the data that ultimately has probative value as evidence, and some data points may only be valuable when connected to others. In the future, AI-based systems are likely to be used increasingly for the processing of large data masses in the hopes of locating or 'producing' evidence from the raw data. Indeed, in the Strategy and Action Plan for Tackling the Grey Economy and Economic Crime 2020–2023, one aim is to develop an AI for the processing of criminal investigation material.[38] According to the decision in principle taken by the Finnish Government the aim of the project is

> to develop a system for the Police that utilises IT evidence in a secure way, cross-links it with existing police databases (police information system databases), and thus finds unifying factors in the data that could not be found without the system. The purpose of the system is to analyse the data reproduced from several devices at the same time and to compare them with each other. The system to be developed in the project would be able to receive unformatted data and automatically convert it to a format that can be indexed and analysed.[39]

3.2. Applicability of general evidence law norms

In short, there are no specific rules concerning evidence gathered or produced by AI-based systems in Finnish law. AI-based evidence is not considered a separate class or category of evidence, and there are no specific conditions on its admissibility in a trial, nor specific rules on how it should be assessed by triers of fact.

Traditionally, Finnish law of evidence has strongly embraced the *free theory of evidence*, setting very few formal standards, conditions, and requirements for the admissibility, presentation, and evaluation of evidence. Most aspects of evidence in civil and criminal proceedings are regulated in Chapter 17 of the Code of Judicial Procedure (4/1734, 'CJP'). The latest complete renewal of Chapter 17, which largely upheld the foundational status of the free theory of evidence despite introducing some new statutory exceptions, came into effect on 1 January 2016 (amendment 732/2015). Questions relating to *electronic* or *digital evidence* were not emphasised in the drafting process, and evidence law remains largely technology neutral. As a consequence of these general characteristics, the admissibility of computer data as evidence has never presented particular legal problems in

[38] *Strategy and Action Plan for Tackling the Grey Economy and Economic Crime* <https://www.vero.fi/en/grey-economy-crime/prevention/torjuntaohjelma/> accessed 29 December 2021.

[39] Valtioneuvoston periaatepäätös kansalliseksi harmaan talouden ja talousrikollisuuden torjunnan strategiaksi ja toimenpideohjelmaksi 2020–2023 33.

Finland,[40] but many norms on aspects of evidence related to information and communication technology remain unclear.

Free introduction of evidence is an essential part of the free theory of evidence. Chapter 17, Section 1(1) of the CJP guarantees each party the right to present evidence to the court investigating the case, as well as the right to comment on each item of evidence presented in court. While there is no list of allowed or disallowed types of evidence, Finnish evidence law recognises and regulates five basic categories (or means) of evidence. The court may hear 1) parties, 2) witnesses, and 3) expert witnesses, and 4) documents and 5) objects of judicial view may be presented to the court as evidence.

All five categories of evidence may be used to relay AI-produced information to the court. For instance, parties and witnesses may tell the court how they conducted digital forensics investigations or otherwise produced evidence with the help of AI-based tools, and what results they obtained. Generally, parties and witnesses are heard orally in the courtroom, and written testimonies are not permitted.[41] However, if the person who has conducted an AI-assisted investigation is not a party and has certain qualifications, they can be classified as an expert witness,[42] in which case they give their evidence initially in the form of a written statement. Expert witnesses may be further heard and cross-examined orally in the courtroom.[43]

The category of documents includes any representations of textual or comparable content, regardless of technology, format, or medium used for storing this information. Raw data, notes, and documentation generated during AI-assisted investigations could be considered as documents that can be presented as evidence as such. Especially in the context of digital forensics reporting, however, there is some unclarity as to the difference between a written statement by an expert witness (in which case there should be the possibility of cross-examination in court) and a document presentable as evidence (in which case the opposing party has a right to comment on the document, but there may

[40] Finland is a party to the Council of Europe Convention on Cybercrime (ETS No 185, 2001), which it ratified in 2007. According to the Explanatory Report to the Convention on Cybercrime, para 141 (concerning Article 14, on the scope of procedural provisions), this 'Convention makes it explicit that Parties should incorporate into their laws the possibility that information contained in digital or other electronic form can be used as evidence before a court in criminal proceedings, irrespective of the nature of the criminal offence that is prosecuted.' This obligation did not necessitate any particular legislative amendments in connection with the ratification of the Convention.

[41] CJP, c 17 s 24. This section also defines some exceptions. For instance, the court may allow a written statement 'for a special reason'.

[42] While normal witnesses tell the court of their experiences, expert witnesses are heard regarding empirical rules requiring special knowledge as well as regarding their application to the circumstances that arise in the case (CJP, c 17 s 34). Expert witnesses are to be known to be honest and competent in their field, and they may not be connected with the case or a party in a manner that endangers their impartiality (s 35).

[43] CJP, c 17 s 36(2): 'An expert witness shall be heard in court in person if: 1) this is necessary in order to remove ambiguities, deficiencies or inconsistencies in their expert statement; 2) the court deems it necessary for another reason; or 3) a party requests this and the hearing would apparently not be meaningless.'

not be a person to cross-examine). However, if the court opts for the latter interpretation and the identity of the author (or a person who contributed to the creation of the document) is known, it may be possible to hear them as a (normal) witness.

Furthermore, interpretations may differ on whether a particular item of evidence should be considered a document or an object of judicial view. The basic difference is that the probative value of a document lies in the textual message or content fixed on some medium, whereas the probative value of an object of judicial view is tied to its external (physical) properties, which can be observed directly with the human senses (typically sight, hearing, or touch).

As has been traditionally noted, the very same sheet of paper can be both a document and an object of judicial view, depending on what a party intends to prove by presenting it. If the relevant fact can be proven by the informational content of the words written on a sheet of paper, the sheet should be considered a document. If the *factum probandum* can be deduced from the ink stains on the paper, the sheet should be considered an object of judicial view.[44] Similarly, an audio recording might be considered a document if the *factum probandum* relates to what has been said, whereas the recording could be considered an object if the *factum probandum* relates to the identity or the emotional state of the speaker, or the noises in the background. In law drafting materials and literature, visual representations of information, such as photographs, maps, and video recordings, are typically considered objects of judicial view (regardless of whether they are printed on paper or stored electronically).[45] Still, in court practice they might be included in the list of documents that have been presented as evidence.[46]

The traditional distinction between documents and objects is badly suited for many modern types of electronic evidence, as computer data that corresponds to textual content or other comparable 'static' information with evidentiary value may be (re)presented in various alternative ways. Computer systems excel in the dynamic and often seamless combination of textual information with other types of media, and the correct interpretation of a message may depend on visual and structural aspects of its representation. Luckily, the legal significance of this distinction is also limited, as documents and objects of judicial view are mostly subject to the same norms.[47]

---

[44] This example is also mentioned in the Government Proposal on the 2016 evidence law renewal (HE 46/2014 vp) 100.

[45] HE 46/2014 vp 100; and Pasi Pölönen and Antti Tapanila, *Todistelu oikeudenkäynnissä* (Tietosanoma 2015) 443–444, 449–450. Instead, for the purposes of the Openness Act and other laws concerning access to public documents, any photographs, maps, video recordings, and audio recordings clearly fall under the concept of document.

[46] In Sweden, differing views have been expressed concerning the classification of sound and video recordings (which is noteworthy, as Finnish and Swedish law of evidence traditionally share many similarities). See Jonas Ekfeldt, *Om informationstekniskt bevis* (Juridiska institutionen, Stockholms universitet 2016) 406–409.

[47] For further discussion of this distinction and electronic evidence, see Riekkinen (n 16) 378–383.

Judicial view may be directed at a physical object that is brought to the courtroom, or a virtual or digital object that is presented with the help of computer hardware and software.[48] An example of something that clearly falls into the category of judicial view in the digital context would be an interactive live demonstration of how a computer system or its user interface operates.[49] The technologies underlying the computer system are largely irrelevant for this classification, and attempts to interactively demonstrate the functioning of an AI-based system may thus also fall into the category of judicial view.

Chapter 17 of the CJP contains a relatively comprehensive set of procedural rules concerning hearings of parties and witnesses. In contrast, the presentation of documents and objects of judicial view is subject to minimal regulation: according to Section 54, they shall be presented to the extent necessary in the main hearing; the same applies to written statements by expert witnesses. Details are left to the discretion of the presiding judge, and the exact manner of presentation may also depend on the availability of technological equipment, such as presentation screens, in the courtroom. As stated in Section 39, copies of documents may be presented in the courtroom, unless the court orders a document to be presented in the original, typically in order to ensure or assess its authenticity and integrity.[50]

In practice, different types of evidence and means of presentation may be combined. This may even be necessary to guarantee that the evidence is vetted thoroughly and its meaning and probative value can be correctly understood and appropriately assessed. For example, raw input and output data processed by an AI-based system can be presented as documentary evidence, and an expert witness may clarify the functional principles of the AI-based system in question, interpret the meaning of the output, point out any potential weaknesses or sources of error, and give their expert assessment on the reliability of the information produced by the system in a written statement. Both the original documents and the written statement may be presented in the courtroom with the help of laptops and presentations screens, with parties highlighting and reading out loud relevant excerpts. After this, the expert may be heard and cross-examined in the courtroom. It is also possible to hear several (expert) witnesses concurrently.[51]

Chapter 17 of the CJP contains some general rules on admissibility. According to Section 8, the court shall reject evidence that, *inter alia*, concerns a circumstance that is not rele-

[48] Judicial view may also take place as a session outside of the courtroom, to allow judges to make direct sensory observations about a specific place, location, or environment.
[49] An example mentioned in HE 46/2014 vp 100 is 'an electronic registry, the operation of its administration software and ways of storing information into the registry' (with a reference to the Supreme Administrative Court case KHO 2009:39, on problems with an electronic voting system).
[50] Copies can be physical or digital; for digital documents, *originality* can be understood as integrity in the sense that the content and the format of the data have remained unchanged and unaltered.
[51] CJP, c 17 s 50(1).

vant in the case, is otherwise unnecessary, or can be replaced by evidence that is essentially more credible[52].[53] Sections 10–23 contain various evidentiary privileges, such as doctor-patient and lawyer-client confidentiality and the privilege against self-incrimination. These provisions give parties and witnesses either the right or the duty not to answer certain questions, or to refuse to testify entirely. If a person has a right or duty not to answer a question, they are not obliged to present documents or objects regarding the same issues, either.[54] For the most part, these privileges do not bear any particular relevance for AI-produced evidence, but some AI-produced data may, of course, fall under some of these privileges in specific circumstances (e.g., data produced by an AI-based medical device used to treat a patient).

Further limitations on admissibility are set in Section 25, which concerns exclusionary rules. Evidence obtained through torture (Subsection 1) or contrary to the privilege against self-incrimination (Subsection 2)[55] may not be used. Subsection 3 concerns unlawfully obtained evidence in general,[56] and it states as follows:

> In other cases the court may use also evidence that has been obtained unlawfully, unless such use would endanger the conduct of a fair trial, taking into consideration the nature of the case, the seriousness of the violation of law involved in the obtaining of the evidence, the significance of the method in which the evidence was obtained in relation to its credibility, the significance of the evidence in respect of the decision in the case, and the other circumstances.

Even unlawfully obtained evidence may thus be used in criminal cases, unless this would lead to a violation of the defendant's right to a fair trial. Consequently, the threshold for excluding AI-produced evidence, like any other evidence, is high. It is important to notice, however, that the applicability of Section 25(3) does not require that the evidence in question has been obtained through a criminal offence – any unlawfulness will do.[57] The possible unfairness of the trial may need to be considered in situations where the input data has been gathered illegally or without a legal basis, or where the use of the AI-based

[52] This can be understood as a form of the *best evidence rule* (although it is more like a principle than a strict rule). However, because the alternative evidence needs to be *essentially* more credible in order to warrant rejection of the proffered evidence, this provision is unlikely to lead to the rejection of any AI-related evidence in favour of evidence produced by a more reliable AI-based system, or in favour of unprocessed data that is not subject to possible sources of error introduced by AI-enabled processing. Exceptions to the right to present evidence should be construed narrowly, and a claim regarding more credible evidence is likely to lead to both of the competing pieces of evidence being presented and compared to each other.
[53] Furthermore, the court shall also reject evidence that can be replaced by evidence that is available with essentially less cost or difficulty, and evidence that despite appropriate measures could not be obtained.
[54] CJP, c 17 s 9(2).
[55] Beyond this one privilege specifically mentioned in s 25(2), the law is not perfectly clear on whether violations of other evidentiary privileges should automatically lead to exclusion of evidence, or if evidence erroneously presented in violation of these privileges should be subjected to the test set in s 25(3). HE 46/2014 vp 92–93 suggests the first interpretation.
[56] Sub-ss 1 and 3 apply to all cases in the general courts, whereas sub-s 2 applies only to criminal cases.
[57] HE 46/2014 vp 92–93.

system can otherwise be considered illegal or unlawful. For instance, a minor violation of a data protection principle (e.g., data minimisation or storage limitation), which might lead to administrative sanctions under the GDPR, would still be very unlikely to lead to exclusion, provided that such unlawful data processing ended up producing relevant evidence that can be considered reliable and credible.[58]

Some of the assessment criteria mentioned in Section 25(3) do not bear any particular relevance in relation to AI-based systems. AI-produced evidence can be proffered in all kinds of cases and have any level of significance in respect to the decision. Although credibility is a factor in the assessment, Section 25(3) notably *only* concerns situations where evidence is unlawfully obtained. In the absence of unlawfulness, any issues that give reason to question the credibility of the AI-produced evidence (whether they relate to the input data, the algorithm, or anything else) are to be considered when assessing the probative value of the evidence, but will not result in inadmissibility. On the other hand, as taking appropriate measures to ensure the integrity of personal data is a legal obligation of the controller and the processor under data protection law,[59] negligent data security practices that allow tampering with or corruption of input data or data processing operations might (theoretically) be enough to make Section 25(3) applicable.

Finally, it should be noted that the exclusionary rule in Section 25(3) is particularly aimed at deterring misconduct by public officials. Law drafting documents suggest that the exclusion of credible evidence is not a desirable result in case of private misconduct, even when this private action amounts to a criminal offence.[60] However, this position has not been fully endorsed in legal commentary,[61] and criminal actions by private individuals have led to exclusion of evidence in earlier case law.[62] None of these sources of law address situations in which the illegal activity relates to the production of evidence with the help of AI-based systems, however. In the authors' view, exclusion of evidence that has been produced by private entities using AI-based systems in an illegal or unlawful manner is very unlikely, but cannot be ruled out categorically.

---

[58] Cf. Oskari Paasikivi, 'Tietosuojasta vapaa todistelu? Todistelu siviiliprosessissa henkilötietojen suojan näkökulmasta' (Master's thesis, Helsinki University 2019) 30–31. Paasikivi, who discusses the relationship between data protection and evidence in civil proceedings, concludes that s 25(3) does not prevent the presentation of evidence that has been obtained in violation of data protection law, despite the broad meaning of 'unlawfulness' and the fact that the provision is formally applicable in civil proceedings.

[59] GDPR, Articles 5(1)(f) and 32 and Act on the Processing of Personal Data in Criminal Matters, ss 9, 31 and 32.

[60] HE 46/2014 vp 94. However, the Legal Affairs Committee stated that exclusion due to a criminal act by a private third party should only occur in 'extremely exceptional cases' (LaVM 19/2014 vp 21).

[61] See, e.g., Mikko Vuorenpää, 'Muutama huomio laittomalla tavalla hankitun todistusaineiston hyödyntämisestä' (2018) 99 *Defensor Legis* 306, 312–313. Antti Jokela, *Pääkäsittely, todistelu ja tuomio. Oikeudenkäynti III* (Talentum 2015) 341 argues that legal practitioners should be held to the same standards as public officials. See also Pölönen and Tapanila (n 45) 336–337; and Riekkinen (n 16) 352–356.

[62] Court of Appeal of Eastern Finland, Judgment of 12 May 2010 (R 09/506, I-SHO 2010:5).

3.3. Evidence produced by predictive policing systems

The central administrative authority of the Finnish Police, the National Police Board of Finland, has not announced the usage of AI-based predictive policing systems by either the local police departments or the NBI.[63] However, as a member of the European Union, Finland has implemented the Passenger Name Record ('PNR') Directive ((EU) 2016/681)[64] in the form of the Act on the use of Passenger Name Record data for the prevention of terrorist offences and serious crime (657/2019). Analysing PNR data to identify people who were not suspected before and making them a target of policing activities based on the analysis provided by the PNR system could be considered as a form of predictive policing.[65] Limited information is available on the workings of the PNR system in Finland.[66]

---

[63] Ministry of the Interior, *Finland's Strategy on Preventive Police Work 2019–2023* (Publications of the Ministry of the Interior 2019:11) 37: 'Major investments are being made in the development of automation and artificial intelligence in Finland and other countries and using such applications in different tasks is still in its initial stages.': <https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/161343/SM_11_19_Strategy%20on%20preventive%20police%20work.pdf> accessed 27 December 2021. See also Vesa Syngelmä, 'Ennustamisteknologioiden hyödyntämismahdollisuudet osana ennakoivaa poliisitoimintaa' (Master's thesis, Tampere University, 2021) (concluding that predictive technologies are not currently in use in Finland).

[64] Directive (EU) 2016/681 of the European Parliament and of the Council of 27 April 2016 on the use of passenger name record (PNR) data for the prevention, detection, investigation and prosecution of terrorist offences and serious crime.

[65] European Data Protection Supervisor, Request for an Opinion by the European Parliament, draft EU-Canada PNR agreement (Opinion 1/15) Hearing of 5 April 2016 Pleading notes of the European Data Protection Supervisor (EDPS): <https://edps.europa.eu/sites/default/files/publication/16-04-05_pleading_canada_pnr_en.pdf> accessed 27 December 2021; and Douwe Korff and Marie Georges, Passenger Name Records, data mining & data protection: the need for strong safeguards. Executive summary. Council of Europe. The Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (T-Pd) Strasbourg, 15 June 2015. <https://rm.coe.int/16806b1761> accessed 27 December 2021. However, it should be noted that in Case C-817/19 *Ligue des droits humains ASBL v Conseil des ministres* [2022] para 194 the Court of Justice of the European Union stated that the wording of the directive precludes the use of self-learning AI systems in the evaluation process and especially when determining the evaluation criteria. Thus, at least after the judgement, it is safe to say that self-learning AI should not be used in the aforementioned context.

[66] The preparatory work of the Act is the only source to assess the nature of the system. According to the Government Proposal on the PNR legislation (HE 55/2018 vp) 18, the PNR data should be used for the creation of threat assessments and risk profiles, which guides the authorities to target their activities towards the passengers that fit the profiles. In order to identify 'the unknown suspects' the Passenger Information Unit (which in Finland is formed by the Police, the Customs and the Border Guard) has predefined evaluation criteria to which the PNR data is compared to. During the legislative procedure, the national data protection authority characterised the PNR system as 'an AI type solution' (Opinion of the Data Protection Ombudsman for the Constitutional Law Committee on 10 September 2018: <https://www.eduskunta.fi/FI/vaski/JulkaisuMetatieto/Documents/EDK-2018-AK-203144.pdf>; and Opinion of the Office of the Data Protection Ombudsman for the Administrative Committee on 7 January 2019 <https://www.eduskunta.fi/FI/vaski/JulkaisuMetatieto/Documents/EDK-2019-AK-235519.pdf> both accessed 29 December 2021).

Beyond the PNR system, the usage of predictive policing in Finland is still in its infancy. Public debate on predictive policing in Finland has been almost non-existent and academic research is still scarce.[67] There are no national legal rules, other normative instruments, or soft law sources specifically concerning AI-based systems for predictive policing. Therefore, the relevant normative framework mostly comprises of constitutional requirements (including fundamental rights and the principle of legality), laws on the processing of personal data, and the Police Act. Arguably, existing provisions in the Police Act, such as the general provisions on the duties of the Police[68] or the power of preventing an offence or disturbance[69], do not form an adequate legal basis for the Police to act based on a predictive policing prediction.

Regardless, there are several indications that the Police is interested in using AI in its crime prevention analysis. The role of *intelligence-led policing* (operating on knowledge analysed from criminal intelligence) has already been emphasised on many occasions by the Police.[70] Based on recent reports provided by the Police, using AI systems seems to be the goal and the next step of the digitalisation of police operations. According to the National Police Commissioner, who is the head of the National Police Board of Finland, the Board is currently exploring the future technology the Police would be using in 2030.[71] However, so far, no material on the project has been made available to the public.

In 2018, the final report of the study project on the Status of Crime Prevention in Finland found the role of efficient IT systems in the analysis of data for crime prevention to be highly important, and emphasised the usage of AI and big data in the analysis.[72] The report stated that the elements of big data and AI should be implemented in processing

---

[67] Sofia Söderholm, *Potentiaalisen rikoksentekijän asema ja oikeus syyttömyysolettamaan ennakoivassa poliisitoiminnassa* (Legal Tech Lab 2020) <https://helda.helsinki.fi/handle/10138/331782> accessed 29 December 2021; and Syngelmä (n 63).

[68] For instance, Police Act, c 1 s 1(1) defines the duties of the Police, which include securing the rule of law, maintaining public order and security, and preventing, detecting and investigating crimes.

[69] Police Act, c 2 s 10(1): 'A Police officer has the right to remove a person from a scene if there are reasonable grounds to believe on the basis of the person's threats or other behaviour, or it is likely on the basis of the person's previous behaviour, that he or she would commit an offence against life, health, liberty, home or property, or would cause a considerable disturbance or pose an immediate danger to public order or security.'

[70] Mika Sutela, 'Tiedon, analyysin ja analytiikan hyödyntämisen tarve poliisissa – ilmeinen ja suuri?' (Official blog of the Police, 15 September 2019): <https://poliisi.fi/blogi/-/blogs/tiedon-analyysin-ja-analytiikan-hyodyntamisen-tarve-poliisissa-ilmeinen-ja-suuri-> accessed 29 December 2021; and Opinion of the National Bureau of Investigation on data processing in the Police on 10 January 2019: <https://www.eduskunta.fi/FI/vaski/JulkaisuMetatieto/Documents/EDK-2019-AK-236903.pdf> accessed 29 December 2021.

[71] Poliisiylijohtaja Seppo Kolehmainen muistutti poliisien valatilaisuudessa: Poliisin pysyttävä mukana muutoksessa (17 May 2019): <https://poliisi.fi/-/poliisiylijohtaja-seppo-kolehmainen-muistutti-poliisien-valatilaisuudessa-poliisin-pysyttava-mukana-muutoksessa> accessed 29 December 2021.

[72] Rikostorjunnan tila -selvityshanke and Tero Kurenmaa, *Rikostorjunnan tila -selvityshankkeen loppuraportti* (The publication series of the National Police Board of Finland 1/2018) 45: <https://poliisi.fi/documents/25235045/42553324/rikostorjunnan_tila_selvityshankkeen_loppuraportti.pdf> accessed 21 August 2023.

large datasets in *Vitja,* the police information system.[73] Furthermore, the Financial Intelligence Unit of the NBI undertook a project that explored the opportunities arising from the use of AI in the fight against money laundering and terrorist financing. The project was called *RANKKA*, and took place between 2020 and 2021.[74] More recently, the Ministry of Finance has set up a working group to develop digital tools to support national risk assessment work on money laundering and terrorist financing. The term of office of the working group is 13.10.2021–30.6.2024. According to the project description, 'the aim is to create two different tools to support national money laundering and terrorist financing risk assessment work: a digital data platform and a risk assessment tool for processing quantitative and qualitative data.'[75] Other Finnish law enforcement organisations have likewise started to consider using AI in their activities.[76]

Due to these developments, it is reasonable to ask whether information produced by predictive policing could, currently or in the future, be used as evidence in criminal proceedings. In essence, the current admissibility of any output by AI-based predictive policing systems in trials is subject to the general evidence law norms that have been described above. There is no categorical ban on evidence produced by predictive policing systems. Naturally, the admissibility of any such output data primarily depends on whether or not it may help to prove facts relevant to the case at hand. A prediction, no matter how well it can be justified statistically, likely bears no relevance in proving that the accused is guilty of a specific past offence. Therefore, if such evidence is proffered in support of the guilt of the accused, it should be rejected by the court under Chapter 17, Section 8 of the CJP. This applies regardless of whether the primary use of system itself

[73] It should be noted that *Vitja* includes *Poti,* the new intelligence system of the Police, and it is also used by other law enforcement authorities such as the Finnish Customs and the Finnish Border Guard. Poliisi panostaa rikosten ehkäisemiseen ja paljastamiseen (19 December 2018): <https://poliisi.fi/-/poliisi-panostaa-rikosten-ehkaisemiseen-ja-paljastamiseen> accessed 27 December 2021.

[74] Projects and top-up funding of the Police are available at <https://poliisi.fi/en/projects-and-complementary-funding>. The aim of the project was 'to produce a study of technological solutions related to artificial intelligence and digitalisation in general that are applicable in the context prevention, detection and investigation of money laundering' (*Annual report of the Financial Intelligence Unit* (2020) 39: <https://poliisi.fi/documents/25235045/67733116/2020-Rahanpesun-selvittelykeskus-vuosikertomus-2020.pdf> accessed 29 December 2021).

[75] The project website <https://vm.fi/hanke?tunnus=VM141:00/2021> accessed 29 December 2021.

[76] The Finnish Border Guard has initiated a development project of surveillance techniques called *RAVAKE*. The aims of the project are to modernise the surveillance systems of land borders and sea areas, and the solutions used in maintaining and managing situational pictures. In the second phase of the project (2022–2024) the aim is to replace the Border Guard Information System by introducing a centralised data warehouse, which is to be used for management, analysis, and exploitation of data. In addition, the opportunities created by AI are to be exploited effectively (*Annual report of the Border Guard* (2020) 15: <https://raja.fi/documents/44957406/64377821/Tilinp%C3%A4%C3%A4t%C3%B6s_2020.pdf> accessed 29 December 2021). The National Enforcement Authority Finland has undertaken two projects called *RATKE* and *Harmaa*, the aim of which is to increase the efficiency of data acquisition, data processing and decision-making necessary for enforcement through robotics and data analytics (Ulosottolaitoksen hankkeet RATKE ja Harmaa hyödyntävät uutta teknologiaa (21 December 2021): <https://ulosottolaitos.fi/fi/index/ulosottolaitos/ajankohtaista/verkkouutisetjatiedotteet/uutiset2021/ulosottolaitoksenhankkeetratkejaharmaahyodyntavatuuttateknologiaa.html> accessed 29 December 2021).

is lawful or not, and even in the event that the use of a specific predictive policing system is given a specific legal basis by a parliamentary act.

The use of predictive policing systems may indirectly lead to the discovery of documents, objects, or witness statements. If the primary use of such a system is unlawful, arguably any evidence obtained in an investigation started due to a prediction might be considered unlawfully obtained. If a coercive measure has been performed based on a mere prediction, in the absence of the applicable prerequisites and/or procedure for this measure, the situation is more clear: any evidence produced by this measure should be considered unlawfully obtained.[77] The same applies to all police interventions on an individual's rights that are not based on a specific legal provision. If relevant evidence has been obtained as a consequence of the unlawful use of a predictive policing system, the Section 25(3) exclusionary rule should be considered. Although *Fernwirkung* of the exclusionary rule is a possibility, if the link between the unlawful action and obtaining a credible item of evidence is weak or indirect, exclusion is not a likely outcome.[78]

As a distinction to the use of predictive data as evidence of facts relating to past events or to obtain further evidence, predictive AI tools could also be used in courts to predict the future conduct of the accused. Due to the general sentencing rules and principles in Chapter 6 of the Criminal Code,[79] for example a recidivism risk score should generally bear no relevance for sentencing. A possible exception to this relates to so-called combination sentences.[80] In determining whether the prerequisites for a combination sentence are fulfilled, a risk score (or some other comparable prediction) could be used as evidence in support of the dangerousness of the offender,[81] as an addition to the psychiatric expert statement that must be procured by the court. Similarly, a risk score might be relevant when deciding on the conditional release of a prisoner.[82] However, since there is currently no clear legal basis for the use of an algorithmic system for these purposes, risk

[77] Arguably, in some situations a prediction might contribute to fulfilling the standards of proof that serve as prerequisites for investigative measures and investigative powers. The law defines different thresholds for different situations and powers (e.g., 'there are grounds to suspect', 'it is probable').

[78] The 'fruit of the poisonous tree' doctrine in its strictest form is not observed in Finland. Documents and objects obtained 'indirectly' through torture, however, should always be excluded under CJP, c 17 s 25(1). See, e.g., Pölönen and Tapanila (n 45) 232, 239, 248; and Riekkinen (n 16) 315–316.

[79] Punishments are primarily based on the offence, not the offender.

[80] A combination sentence is a specific criminal sanction meant for dangerous recidivist offenders. According to Criminal Code, c 2c s 11(1), it consists of (fixed-term) unconditional imprisonment and a one-year supervision term that immediately follows the prison term. The person serving the sentence is not entitled to conditional release or probationary liberty under supervision. The provisions on combination sentences entered into effect on 1 January 2018 (amendment 800/2017).

[81] One of the prerequisites, as per Criminal Code, c 2c s 11(2)(3) is that 'the perpetrator is deemed, on the basis of circumstances related to the offences and an examination required in [CJP, c 17 s 37(3)], to be particularly dangerous to the life, health or liberty of another person'.

[82] Criminal Code, c 2c ss 9 and 10. Concerning offender risk assessments in the Finnish penal system, see Annakaisa Pohjola, Vaarallinen rikoksentekijä? Tutkimus rikoksentekijän vaarallisuuden arvioinnista rikosoikeudellisessa seuraamusjärjestelmässä (Suomalainen Lakimiesyhdistys 2017).

scores generated by authorities could be considered unlawfully obtained. Further, without a high degree of transparency and proof that the specific algorithmic system produces statistically accurate and bias-free results, the credibility of any risk score would be by default impaired. Arguably, reliance on risk scores would in many situations endanger the fairness of the proceedings and trigger the exclusionary rule.[83]

3.4. Evaluation of AI-produced evidence

As regards evaluation of evidence, the Finnish system grants broad discretion to judges. There are no formal or categorical rules concerning the reliability, weight, or probative value of certain types or means of evidence. Chapter 17, Section 2(2) of the CJP states:

> 'The court, having considered the evidence presented and the other circumstances that have been shown in the proceedings, determines what has been proven and what has not been proven in the case. The court shall consider the probative value of the evidence and the other circumstances thoroughly and objectively on the basis of free consideration of the evidence, unless provided otherwise in law.'

Free consideration does not mean freedom to make arbitrary decisions or freedom from the general principles of scientific knowledge, logic, and reasoning. The court is also obligated to explain its reasoning on matters of evidence in the written judgment.[84]

Finnish legal commentary offers little insight into how AI-produced evidence should be evaluated in criminal cases. Indeed, there is little that can be said on a general level. The specific circumstances of the case, of the type of AI-produced evidence, and of each item of evidence need to be carefully considered. However, one of the authors has argued for a general 'auxiliary questions' framework to assist triers of fact in assessing electronic evidence. This model places the emphasis on the origins of the data as well as the informational process that leads to the evidence being presented in a court of law. The aim is to identify or rule out different kinds of sources of error that relate to different aspects of

[83] The use of algorithm-based risk assessment systems in the context of combination sentences has been evaluated by Anita Kritsos, 'Algoritmisten päätöksentekojärjestelmien soveltaminen rikoksentekijän vaarallisuutta koskevassa tuomarin päätöksenteossa' (Master's thesis, Helsinki University 2019). Kritsos concludes that although the free theory of evidence could be seen to provide a base for the use of algorithmic evidence, the use of such systems in supporting danger assessments cannot be currently lawful due to problems relating to transparency, lack of a specific legal basis, ethical issues, discriminatory effects, and due process issues.

[84] Criminal Procedure Act (689/1997), c 11 s 4(1): 'The reasons for the judgment shall be stated. The statement of reasons shall indicate the factors and the legal reasoning on which the decision is based. The statement shall also indicate the basis on which a contentious issue has been proven or not proven.' CJP, c 24 s 4 contains a similar obligation applicable to civil cases.

digital data and to the processing of the data in question. The non-exhaustive list of auxiliary questions additionally serves as a checklist that may help parties in supporting their own evidence and challenging evidence presented by other parties.[85]

Applying this model, and in accordance with the general principles on the burden of proof and equality of arms,[86] a party wishing to introduce AI-produced evidence would need to support their evidence by presenting information about the functioning of the AI-based system in general and in the particular case, and about subsequent processing of the data and measures taken to guarantee its integrity.[87] Unless the opposing party and the court are supplied with information that makes it possible to test the reliability of the system and of the data, such evidence should not be given significant weight, especially when the AI-produced evidence is proffered by the prosecution in criminal cases. AI-produced evidence should not simply be presumed reliable and trustworthy, and the presumption of innocence must be guaranteed. If AI-produced evidence is presented in support of the innocence of the defendant, the requirements of providing supporting information should not be interpreted to be as stringent. Still, in practice, any information that enables the court to rule out sources of error that could diminish the credibility of potentially exonerating evidence will certainly help the defence case.

4. **Evidence assessed through AI-based systems**

In Finland, judges are not known to use any AI-based systems to assess criminal evidence or its probative value. There is no legal basis for the use of such systems. The recently introduced case and document management system of the general courts, *AIPA*, contains no such functionality, nor does any other official information system currently or previously used by the courts. Consequently, there is no case law regarding decisions or judgements where such systems would have been (openly) used to assess evidence.

Under current law, it is clear that a person's guilt may not be determined by an AI-based system, and the introduction of any AI-based decision-making in criminal cases, especially in questions relating to evidence or culpability, seems highly unlikely even in the long term. Introduction of such a system would most likely require a constitutional amendment, as AI-based decision-making in such matters could be seen to contradict the provisions of the Constitution on procedural rights and protection under law (Section 21) and the independence of courts (Section 3(3)).[88] In general, Finnish discussion on au-

[85] See Juhana Riekkinen, 'Auxiliary Questions for Evaluating Electronic Evidence' (2019) *Jusletter IT*; and Riekkinen (n 16) 527–530.

[86] In criminal cases, the burden of proof is on the plaintiff regarding all circumstances on which their request for punishment is based, and the applicable standard of proof is 'no reasonable doubt' regarding the guilt of the defendant (CJP, c 17 s 3).

[87] This can also be described as meta-level evidence relating to the reliability of the primary evidence.

[88] Further, Courts Acts (673/2016), c 9 s 1(1) states: 'Judges exercise judicial powers independently and are, in this activity, subject only to the law.'

tomated decision-making in the public sector has predominantly focused on administrative decision-making in fields such as taxation,[89] and the automation of complex judicial decision-making is yet to be seriously discussed.[90]

From a legal point of view, the use of automated tools to support human decision-making is not as problematic as fully automated decision-making.[91] In the context of sentencing, a simple rule-based software tool, which can be used to analyse the criminal records of defendants, is reportedly already in use in several Finnish courts.[92] In the context of evidence, it could be argued that the use of AI-based support systems might help judges to map out the relations between different items of evidence, to structure their reasoning on matters of evidence, and consequently, to write better and more logically sound judgments. Further, as judges have broad discretion in evaluating evidence, it could be argued that as long as the judgment openly elaborates on how AI-based systems have been used to help in assessing the evidence, or at least describes the logic utilised by the AI-based system as understood by the human decision-maker, this would be permissible.[93] For the moment, the availability of easy-to-use and proven-to-be-reliable AI-based evidence management or decision support tools seems scarce, and therefore their adoption by Finnish judges—especially in the absence of parliamentary or other high-level institutional approval—seems unlikely in the near future.[94]

---

[89] See, e.g., Jorma Kuopus, *Hallinnon lainalaisuus ja automatisoitu verohallinto* (Lakimiesliiton Kustannus 1988); and more recently, Hanne Hirvonen, 'Automatisoitu päätöksenteko julkisella sektorilla' (2018) *Oikeus* 47(3) 302; and Tuomas Pöysti, 'Kohti digitaalisen ajan hallinto-oikeutta' (2018) 116 *Lakimies* 868, 892–895.

[90] Already Kaarle Makkonen discussed computational modelling of judicial decision-making in his dissertation *Zur Problematik der juridischen Entscheidung: eine strukturanalytische Studie* (University of Turku 1965). More recently, use of AI in the courts has been discussed by Riikka Koulu, Risto Koulu and Sanna Koulu, *Tuomarin roolit tuomioistuimissa* (Alma Talent 2019) 178–188, 191; and Sanna Luoma, 'Artificial Intelligence Improving the Delivery of Justice and How Courts Operate' in Riikka Koulu and Laura Kontiainen (eds), *How Will AI Shape the Future of Law* (Legal Tech Lab, University of Helsinki 2019).

[91] Nevertheless, even simpler forms of automation and digitalisation in the courts may bring about issues of legal significance, some of which have been pointed out by Riikka Koulu, 'Digitalisaatio ja algoritmit – oikeustiede hukassa?' (2018) 116 *Lakimies* 840, 847.

[92] Juha Terho, 'Automaattinen päätöksenteko ratkaisuna konkurrenssin katkeamiseen liittyviin ongelmiin' (2022) 103 *Defensor Legis* 106. According to Criminal Code, c 7 s 6, the court may need to consider earlier sentences of imprisonment in sentencing. The interpretation of this provision has been clarified by the Supreme Court (KKO 1972 II 5 and KKO 2004:130), and the tool seeks to model and automatise this 'algorithm' determined in case law. The tool itself, *Konkurrenssikone*, is available at GitHub: <https://github.com/konkurrenssikone> accessed 28 March 2022.

[93] Machine learning approaches typically suffer from limited explainability (and various biases) that would be unacceptable in criminal proceedings. Adoption of support tools based on machine leaning is effectively prevented by the legal obligation to provide reasoning concerning the basis on which a contentious issue has been proven or not proven (Criminal Procedure Act, c 11 s 4(1)).

[94] The absence of a legal basis in a parliamentary act could be seen as problematic in regard to Courts Acts, c 9 s 1(1) and the principle of legality. Simple visualisation tools that do not provide any conclusions or numerical values but allow for easier structuring of relationships between individual items of evidence and *facta probanda* could be the most realistically adoptable type of support software.

A further argument against the likelihood of adoption of software tools for evaluation of evidence is the fact that Bayesian and other mathematical theories of evidence (the logic of which can easily be expressed in code)[95] seem to have gained very limited acceptance among Finnish judges and other legal professionals, although they have been discussed in domestic literature for decades.[96] As mathematical models are not generally relied on, judges would probably be somewhat reluctant to accept probabilities, likelihood ratios, probative values, or any other numerical values calculated by a software tool. Moreover, as the case law of the Supreme Court of Finland does not approach the definition of the standard of proof in criminal cases in terms of mathematical probabilities, but instead by focusing on alternative hypotheses or explanations,[97] such numerical values would be, ultimately, of limited use without a wider reform of law of evidence.

5. **Conclusion**

Finnish legislation on coercive measures and other police powers does not specifically address the question of which tools can be used in evidence-gathering. In addition to the Coercive Measures Act and related legislation, limits to the use of AI-based systems are defined by data protection legislation and the constitutional principle of legality. Notably, both the Coercive Measures Act and data protection legislation rely heavily on principles, which influence the evaluation of the lawfulness of individual actions. Still, the nature of these norms makes it rather difficult to specify in advance which AI-based tools or methods might be considered lawful in a particular investigative scenario. As there is no case law and only scarce academic research or legal commentary, the legal situation is somewhat unclear.

In a similar manner, the use of AI-produced evidence in trials must be assessed against general, mostly technology neutral rules and principles. Finnish evidence law is based on the free theory of evidence, which means that AI-based evidence is generally admissible, and AI-produced information can be communicated to the court in many alternative ways, as long as the right to a fair trial and associated principles, such as *audiatur et altera pars* and *equality of arms*, are guaranteed. However, AI-produced output that is not relevant to the facts of the case should be rejected by the court, and unlawfully obtained evidence should be excluded if its use might endanger the right to a fair trial. In particular, exclusion might be required when authorities knowingly use an AI-based system unlawfully, the system does not function in a transparent way, and/or the system is prone to bias or errors. Again, lack of case law and scholarship focused on these issues means that the legal situation is far from settled.

---

[95] This is not to say that software tools would necessarily need to be limited to mathematical models.

[96] See, e.g., Hannu Tapani Klami, Minna Gräns and Johanna Sorvettula, *Law and Truth: A Theory of Evidence* (Finnish Society of Sciences and Letters 2000).

[97] See, e.g., KKO 2013:96, para 6. The approach of the Finnish courts is largely characterised by non-mathematical theories of evidence, such as the hypothesis model proposed by Christian Diesen, *Bevisprövning i brottmål* (Juristförlaget 1994) 120–151.

In Finland, evaluation of evidence is generally left to the free discretion of the judges. Currently, AI-based systems are not used in courts to assess evidence. While the use of fully automated systems is not lawful and is likely to remain so, AI-based software tools could play some role in supporting and/or assisting the judges in analysing evidence.

Selected literature

Annual report of the Financial Intelligence Unit (2020) <https://poliisi.fi/documents/25235045/67733116/2020-Rahanpesun-selvittelykeskus-vuosikertomus-2020.pdf> accessed 29 December 2021

Annual report of the Border Guard (2020) <https://raja.fi/documents/44957406/64377821/Tilinp%C3%A4%C3%A4t%C3%B6s_2020.pdf> accessed 29 December 2021

Diesen C, Bevisprövning i brottmål (Juristförlaget 1994)

Ekfeldt J, Om informationstekniskt bevis (Juridiska institutionen, Stockholms universitet 2016)

Ervo L, Oikeudenkäynnin oikeudenmukaisuusvaatimus: Käsikirja lainkäyttäjille [The requirement of fairness of the proceedings: Handbook for judicial decision-makers] (WSOYPro 2008)

European Data Protection Supervisor, Request for an Opinion by the European Parliament, draft EU-Canada PNR agreement (Opinion 1/15) Hearing of 5 April 2016 Pleading notes of the European Data Protection Supervisor (EDPS) <https://edps.europa.eu/sites/default/files/publication/16-04-05_pleading_canada_pnr_en.pdf> accessed 27 December 2021

Fredman M, Rikosasianajajan käsikirja [Criminal attorney's handbook] (2nd edn, Alma Talent 2021)

Government Proposal on amending the Coercive Measures Act and the Criminal Investigation Act (HE 217/2022 vp)

Government Proposal on the evidence law renewal (HE 46/2014 vp)

Government Proposal on the PNR legislation (HE 55/2018 vp)

Hirvonen H, 'Automatisoitu päätöksenteko julkisella sektorilla' [Automated decision-making in the public sector] (2018) Oikeus 47(3) 302

Joh E E., 'Feeding the Machine: Policing, Crime Data, & Algorithms' (2017) 26 Wm. & Mary Bill Rts. J. 287

Jokela A, Pääkäsittely, todistelu ja tuomio. Oikeudenkäynti III [Main hearing, evidence and judgment. Trial III] (Talentum 2015)

Klami H T, Gräns M and Sorvettula J, Law and Truth: A Theory of Evidence (Finnish Society of Sciences and Letters 2000)

Korff D and Georges M, Passenger Name Records, data mining & data protection: the need for strong safeguards. Executive summary. Council of Europe. The Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (T-Pd) Strasbourg, 15 June 2015. <https://rm.coe.int/16806b1761> accessed 27 December 2021

Koulu R, 'Digitalisaatio ja algoritmit – oikeustiede hukassa?' [Digitalisation and algorithmic decision making – jurisprudence at a crossroads?] (2018) 116 Lakimies 840

Koulu R, Koulu Risto and Koulu Sanna, Tuomarin roolit tuomioistuimissa [The roles of the judge in courts] (Alma Talent 2019)

Kritsos A, 'Algoritmisten päätöksentekojärjestelmien soveltaminen rikoksentekijän vaarallisuutta koskevassa tuomarin päätöksenteossa' [The application of algorithmic decision-making systems in judicial decision-making concerning dangerousness of the offender] (Master's thesis, Helsinki University 2019)

Kuopus J, Hallinnon lainalaisuus ja automatisoitu verohallinto [The Rule of Law and Computerized Administration of Taxation] (Lakimiesliiton Kustannus 1988)

Luoma S, 'Artificial Intelligence Improving the Delivery of Justice and How Courts Operate' in Riikka Koulu and Laura Kontiainen (eds), How Will AI Shape the Future of Law (Legal Tech Lab, University of Helsinki 2019)

Makkonen K, Zur Problematik der juridischen Entscheidung: eine strukturanalytische Studie (University of Turku 1965)

Ministry of the Interior, Finland's Strategy on Preventive Police Work 2019–2023 (Publications of the Ministry of the Interior 2019:11) <https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/161343/SM_11_19_Strategy%20on%20preventive%20police%20work.pdf> accessed 27 December 2021

Opinion of the Data Protection Ombudsman for the Constitutional Law Committee on 10 September 2018 <https://www.eduskunta.fi/FI/vaski/JulkaisuMetatieto/Documents/EDK-2018-AK-203144.pdf> accessed 29 December 2021.

Opinion of the National Bureau of Investigation on data processing in the Police on 10 January 2019 <https://www.eduskunta.fi/FI/vaski/JulkaisuMetatieto/Documents/EDK-2019-AK-236903.pdf> accessed 29 December 2021

Opinion of the Office of the Data Protection Ombudsman for the Administrative Committee on 7 January 2019 <https://www.eduskunta.fi/FI/vaski/JulkaisuMetatieto/Documents/EDK-2019-AK-235519.pdf> accessed 29 December 2021

Ortamo S, 'Poliisi on saanut rikollisia kiinni kasvoja tunnistavan tekoälyn avulla ja haluaisi laajentaa valtuuksiaan – testasimme, miten kone toimii' [The Police have apprehended criminals using a facial recognition artificial intelligence and would like to expand their powers – we tested how the machine works] Yle Uutiset (1 August 2020) <https://yle.fi/uutiset/3-11448002> accessed 28 March 2022

Paasikivi O, 'Tietosuojasta vapaa todistelu? Todistelu siviiliprosessissa henkilötietojen suojan näkökulmasta' [Evidence free from data protection? Evidence in civil proceedings from the perspective of data protection] (Master's thesis, Helsinki University 2019)

Pellonpää M, Gullans Monica, Pölönen Pasi and Tapanila Antti, Euroopan ihmisoikeussopimus [European Convention on Human Rights] (6th edn, Alma Talent 2018)

Perry W L, McInnis B, Price C C, Smith S C and Hollywood J S, Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations (1st edn, RAND Corporation 2013)

Pohjola A, Vaarallinen rikoksentekijä? Tutkimus rikoksentekijän vaarallisuuden arvioinnista rikosoikeudellisessa seuraamusjärjestelmässä [Dangerous offender? A Study on offender risk assessment within the Finnish penal system] (Suomalainen Lakimiesyhdistys 2017)

Police reprimand from Deputy Data Protection Ombudsman – police have initiated measures ordered' (28 September 2021) <https://poliisi.fi/en/-/police-reprimand-from-deputy-data-protection-ombudsman-police-have-initiated-measures-ordered> accessed 28 March 2022

Poliisi panostaa rikosten ehkäisemiseen ja paljastamiseen [The Police invest in crime prevention and detection] (18 December 2018) <https://poliisi.fi/-/poliisi-panostaa-rikosten-ehkaisemiseen-ja-paljastamiseen> accessed 27 December 2021

Poliisiylijohtaja Seppo Kolehmainen muistutti poliisien valatilaisuudessa: Poliisin pysyttävä mukana muutoksessa [The National Police Commissioner Seppo Kolehmainen reminded at the Police swearing-in: The Police have to stay on top of the change] (17 May 2019) <https://poliisi.fi/-/poliisiylijohtaja-seppo-kolehmainen-muistutti-poliisien-valatilaisuudessa-poliisin-pysyttava-mukana-muutoksessa> accessed 29 December 2021

Pölönen P and Tapanila A, Todistelu oikeudenkäynnissä [Evidence in the trial] (Tietosanoma 2015)

Pöysti T, 'Kohti digitaalisen ajan hallinto-oikeutta' [Towards administrative law in the digital era] (2018) 116 Lakimies 868

Rautio L et al., Pakkokeinolain muutostarpeiden tarkastelu: Työryhmämietintö [Examination of the needs for amending the Coercive Measures Act: Report of the Working Group] (Ministry of Justice 2022)

Report of the Legal Affairs Committee (LaVM 19/2014 vp)

Rikostorjunnan tila -selvityshanke and Tero Kurenmaa, Rikostorjunnan tila -selvityshankkeen loppuraportti [The final report of the study project on the Status of Crime Prevention in Finland] (The publication series of the National Police Board of Finland 1/2018) <https://poliisi.fi/documents/25235045/42553324/rikostorjunnan_tila_selvityshankkeen_loppuraportti.pdf> accessed 21 August 2023

Riekkinen J, Sähköiset todisteet rikosprosessissa [Electronic Evidence in Criminal Procedure] (Alma Talent 2019)

Riekkinen J, 'Auxiliary Questions for Evaluating Electronic Evidence' (2019) Jusletter IT

Riekkinen J, 'Evidence of Cybercrime and Coercive Measures in Finland' (2016) 13 Digital Evidence and Electronic Signature Law Review 49

Strategy and Action Plan for Tackling the Grey Economy and Economic Crime <https://www.vero.fi/en/grey-economy-crime/prevention/torjuntaohjelma/> accessed 29 December 2021

Sutela M, 'Tiedon, analyysin ja analytiikan hyödyntämisen tarve poliisissa – ilmeinen ja suuri?' [The need for the Police to use information, analysis and analytics – obvious and great?] (Official blog of the Police, 15 September 2019) <https://poliisi.fi/blogi/-/blogs/tiedon-analyysin-ja-analytiikan-hyodyntamisen-tarve-poliisissa-ilmeinen-ja-suuri-> accessed 29 December 2021

Syngelmä V, 'Ennustamisteknologioiden hyödyntämismahdollisuudet osana ennakoivaa poliisitoimintaa' [The opportunities for using predictive technologies as a part of predictive policing] (Master's thesis, Tampere University 2021)

Söderholm S, Potentiaalisen rikoksentekijän asema ja oikeus syyttömyysolettamaan ennakoivassa poliisitoiminnassa [The Legal Status of a Potential Offender and her Right to the Presumption of Innocence in the Context of Predictive Policing] (Legal Tech Lab 2020) <https://helda.helsinki.fi/handle/10138/331782> accessed 29 December 2021

Terho J, 'Automaattinen päätöksenteko ratkaisuna konkurrenssin katkeamiseen liittyviin ongelmiin' [Automated decision-making as a solution to problems relating to how earlier sentences are taken into account] (2022) 103 Defensor Legis 106

Testing of facial recognition software by NBI reported to Data Protection Ombudsman' (9 April 2021) <https://poliisi.fi/en/-/testing-of-facial-recognition-software-by-nbi-reported-to-data-protection-ombudsman> accessed 28 March 2022

Ulosottolaitoksen hankkeet RATKE ja Harmaa hyödyntävät uutta teknologiaa [The projects RATKE and Harmaa of the National Enforcement Authority Finland make use of new technologies] (21 December 2021) <https://ulosottolaitos.fi/fi/index/ulosottolaitos/ajankohtaista/verkkouutisetjatiedotteet/uutiset2021/ulosottolaitoksenhankkeetratkejaharmaahyodyntavatuuttateknologiaa.html> accessed 29 December 2021

Valtioneuvoston periaatepäätös kansalliseksi harmaan talouden ja talousrikollisuuden torjunnan strategiaksi ja toimenpideohjelmaksi 2020–2023 [The Decision in Principle on the National Strategy and Action Plan for Tackling the Grey Economy and Economic Crime taken by the Finnish Government]

Vuorenpää M, 'Muutama huomio laittomalla tavalla hankitun todistusaineiston hyödyntämisestä' [A few observations on the admissibility of unlawfully obtained evidence] (2018) 99 Defensor Legis 306

AI SYSTEMS AND EVIDENCE LAW IN THE NETHERLANDS

By Maša Galič, Abhijit Das and Marc Schuilenburg [*]

Abstract

Digital evidence plays an increasingly important role in contemporary criminal proceedings in the Netherlands. Various types of AI-based systems are used for the production of evidence, including: Hansken, a tool for the gathering of data out of huge data sets, and CATCH, a facial recognition tool. Despite this increasing reliance of digital evidence, Dutch law (including the draft Code of Criminal Procedure, which is the result of the ongoing Modernisation project) has yet to implement any significant changes to rules relating to evidence. As such, the few rules that regulate the gathering of evidence do not fit the particular needs of digital evidence very well. This leads to several issues, including with the principle of equality of arms. Considering the way digital evidence is gathered – in fact, produced – and examined, the defence needs additional or broader rights in order to participate in determining what counts as relevant information in a particular case, to participate in searching for exculpatory evidence, and to question the validity and accuracy of the functioning of AI-based systems. Such rights are, however, slowly being developed through case law.

1 Introduction

Following the structure of the questionnaire, this part of the report is based on the distinction between evidence *gathered* and evidence *produced* by AI-based systems. However, we argue that such a distinction is misplaced. Contemporary AI-based systems, such as Hansken (described below) that are used to gather evidence in a case also produce data. Criminal investigations nowadays lead to huge data sets composed of multimodal data (i.e., unstructured data of different types, including text, photo, video, audio data). Consequently, traditional tools, developed for searching structured textual data, no longer suffice to find what one is looking for. For this reason, new and more complex AI-based systems needed to be developed. These new tools first need to interpret the data by themselves (e.g., a tool searching for images of drugs needs to be able to determine that a particular photo indeed represents drugs). Second, they need to be able to find relevant correlations (or links) between the numerous data points in the data set (e.g., resulting in a convincing time-line and scenario). This means that we are not dealing with simple gathering of data, but with complex production of data by such systems.

[*] Dr. Maša Galič (m.galic@vu.nl) is an Assistant Professor in Privacy and Criminal Procedure Law at the VU University Amsterdam; Abhijit Das is a PhD researcher at the VU University Amsterdam and Programme Director at The Democracy and Media Foundation (a.das@stdem.org); Prof. dr. Marc Schuilenburg (m.b.schuilenburg@vu.nl) is Professor of Digital Surveillance at the Erasmus University Rotterdam and Assistant Professor of Criminology at the VU University Amsterdam. The authors would like to thank prof. dr. Lonneke Stevens and Thomas van Lieshout for their help with writing the part of the report on evidence.

2 Gathering evidence through AI-based systems

2.1 The example of Hansken

The Netherlands Forensic Institute (NFI) has developed a digital forensic tool called 'Hansken' that can process large volumes of (seized) digital material in order to find relevant data points and the connections between them.[1] Hansken is used by several investigative bodies in the Netherlands, including the Dutch National Police for the purpose of criminal investigation and the Dutch Fiscal Information and Investigation Service for the purpose of fraud detection in tax investigations.[2]

Hansken is used to extract and process data from all types of digital devices, such as laptops, smartphones, hard-disks and even whole servers (e.g., in the case of the seized Ennetcom server).[3] At the moment the tool is said to have the capacity to process three terabytes of data per hour.[4] Hansken includes a wide variety of tools (software),[5] which can be used to analyse diverse file systems, extract files, carve unallocated space and create full text indexes, parse chat logs, browse history and e-mail databases.[6] These tools can be used to examine various types of structured and unstructured data that may be relevant for the investigation, including text (e.g., names, keywords, phone numbers, chat-messages, e-mails), photos, videos, various types of metadata, and location data.[7]

2.2 The normative framework for the use of AI-based systems for gathering evidence

2.2.1. The legal framework

In the current legal framework, there are no provisions that specifically deal with Hansken or similar AI-based technologies used for the purpose of gathering evidence in criminal investigations. Instead, existing provisions that were developed for the 'analogue'

[1] Merve Bas Seyyar and Zeno Geradts, 'Privacy Impact Assessment in Large-Scale Digital Forensic Investigations' (2020) 33 Forensic Science International: Digital Investigation 1, 4.
[2] Other national bodies that use them are: the Netherlands Food and Consumer Product Safety Authority and Human Environment and Transport Inspectorate.
[3] See e.g., 'Dutch Police Seize Encrypted Communication Network with 19,000 Users' (*Reuters*, 22 April 2016) <https://www.reuters.com/article/us-netherlands-cyber-idUSKCN0XJ2HQ> accessed 14 January 2022.
[4] Bas Seyyar and Geradts (n 1) 2.
[5] Examples of software include: UFED, EnCase, FTK, EXIF, HDFS, Map Reduce, Cassandra, HBase, Elastic Search and Kafka; see Harm van Beek and others, 'Digital Forensics as a Service: Game On' (2015) 15 Digital Investigation 20.
[6] ibid 21.
[7] Bas Seyyar and Geradts (n 1) 4.

world are used.[8] However, these provisions are few and mainly concern types of evidence admissible in court and very general requirements concerning the lawfulness and reliability of evidence.

Based on the broad wording of Article 339 of the Dutch Code of Criminal Procedure (CCP), almost any type of evidence is admissible in Dutch courts.[9] Nevertheless, when digital data are used as evidence, they are usually submitted in the form of written police statements that report the results of an investigation.[10] Concerning the lawfulness of evidence, Article 359a CCP provides for the possibility to attach consequences to the unlawful gathering of evidence. Depending on the circumstances, the judge can decide to decrease the severity of the punishment, to exclude the evidence or to declare the public prosecutor inadmissible in the prosecution. However, in practice evidence is hardly ever excluded and cases are not negatively affected by unlawfully obtained evidence.[11] As to reliability, Article 359(2) CCP states that when the prosecution or the defence argues that evidence submitted by the other party is unreliable, the judge needs to motivate their rejection of a 'plea against the use of unreliable evidence'.

While the CCP does not contain any concrete provisions concerning the assessment of expert evidence, the Dutch Supreme Court has developed criteria for assessing expert evidence. According to these criteria, if the reliability of expert evidence is disputed, the judge needs to examine whether the expert has the required expertise and, if so, which method(s) the expert used, why the expert considers that these methods are reliable, and the extent to which the expert has the ability to apply these methods in a professional manner.[12] Yet, Dutch courts (so far) have ruled that in relation to the use of Hansken there can be no reference to expertise, so that the data gathered with – or, rather, produced through – Hansken is not considered as expert evidence.[13] The only resort left to the defence to examine the reliability of the Hansken system is to request the investigatory judge to appoint an expert (according to Article 227 CCP), who would provide information on the functioning of Hansken.[14]

---

[8] Bart Custers and Lonneke Stevens, 'The Use of Data as Evidence in Dutch Criminal Courts' (2021) 29 European Journal of Crime, Criminal Law and Criminal Justice 25, 40.

[9] The provision lists the following types of evidence, which are admissible in court: what the judge perceives on their own, statements by suspect, statements by witnesses, statements by an expert, and written documents.

[10] Custers and Stevens (n 8) 36.

[11] ibid 36–37. This is due to a very restricted interpretation of Article 359a stemming from the case law of the Dutch Supreme Court. See, e.g., Supreme Court of the Netherlands, judgment of 19 February 2013, NJ 2013, 308.

[12] Supreme Court of the Netherlands, judgment of 27 January 1998, NJ 1984, 404; see also Custers and Stevens (n 8) 36.

[13] See e.g., District Court of Amsterdam, judgment of 19 April 2018, ECLI:NL:RBAMS:2018:2504 (case nr. 13/997097-16), para. 7.3.

[14] See e.g., District Court of Amsterdam, intermediate decision of 29 September 2020, ECLI:NL:RBAMS:2020:4764 (case nr. 26Marengo), p 16; District Court of Amsterdam, intermediate decision of 17 November 2020, ECLI:NL:RBAMS:2020:5585 (case nr. 26Marengo), p. 7.

There are hardly any content-related changes concerning evidence law in the latest version of the draft new Dutch Code of Criminal Procedure (draft CCP). Two developments, however, merit mentioning.

First, the draft CCP introduces a new provision, according to which the public prosecutor may order companies or institutions, which can 'reasonably be suspected of having access to certain data' relevant for the investigation, to process these data and then submit the result of this processing to law enforcement (Article 2.7.51(1) draft CCP). Google, Facebook and Apple are given as examples of companies that may be asked to perform such processing.[15] Simple types of processing of data needed to provide information (e.g., first finding a customer number in one system, and then using that customer number to find the name and address data in another system) do not fall under this provision (this is covered by the classic disclosure order). Instead, the legislator had a more complex type of processing in mind, where the analysis of data would lead to the creation of new data, thus potentially including analysis performed by AI:

> The power in this Article concerns operations that go beyond multiple searches, for example comparing all data in one dataset with all data in another dataset, in order to identify data that appear in both sets. The main feature of this power, which is distinct from the normal supply of data, is that the operation produces "new" data which are then supplied.[16]

According to the Explanatory Memorandum, the idea behind this provision is to protect the private life of individuals. This provision namely enables the limitation of the amount of data that is provided to law enforcement. As such, the police only receive the results of the data analysis performed by a company that collects the data.[17] However, another, more practical goal is clearly sought through this provision: limiting the influx of data for the police. By ordering certain third parties to perform the initial 'sifting' through data, the police receive a lesser amount of data already considered relevant. In this sense, the new provision aims at enhancing the efficiency of police work (this provision is further discussed in 3.2.4).[18]

The second development in the draft CCP, is the introduction of a special 'technical tool' (*technisch hulpmiddel*) assisting the investigatory judge in his task to sift the data protected by the legal professional privilege (LPP) out of the data set relevant for the criminal investigation. While not mentioned explicitly in the Explanatory Memorandum, this tool is understood as an AI-based system and is seen as a solution to the lack of practical resources and expertise of the investigatory judge to sift out privileged data from large

[15] 'Ambtelijke Versie Juli 2020 Memorie van Toelichting Wetboek van Strafvordering' (Ministerie van Justitie en Veiligheid, 30 July 2020) 442 <https://www.rijksoverheid.nl/documenten/publicaties/2020/07/30/ambtelijke-versie-juli-2020-memorie-van-toelichting-wetboek-van-strafvordering> accessed 14 January 2022.
[16] ibid 443.
[17] ibid 441.
[18] ibid 442.

digital data sets. A lot of trust is placed into this tool.[19] In the Explanatory Memorandum it is, for instance, assumed that the tool will enable the sifting of LPP-data, where the person conducting the sifting via the tool would not gain any knowledge into the LPP-data. This would allow the investigating officer to conduct the sifting, instead of the investigatory judge, who is the only authority that may gain knowledge of LPP-data (Art. 2.7.65(4) draft CCP).

However, the Explanatory Memorandum does not include much discussion of the actual functioning of this tool and whether this would actually be possible from a technical perspective. According to the Explanatory Memorandum, the functioning of the tool is very crude: the investigatory judge and officers compose a list of search terms, which can include telephone numbers and email addresses of a lawyer. On this basis, the tool would then sift out certain protected data. However, as Stevens and Galič point out, it remains completely unclear, how the tool will be able to determine, which communications stemming from this telephone number or email are actually protected by LPP.[20] Not every communication between a client and his lawyer (or a doctor), is namely protected by the privilege (e.g., a discussion about the Tour de France between the two would not fall under the privilege). On the basis of this description, the tool is likely to lead to a large number of false positives and false negatives.

2.2.2. Case law and defence rights: access to the data set, to the AI-tool and information concerning the functioning of the AI-tool

There are no provisions in the law (or lower types of legal instruments), which oblige the prosecution to provide the defence with information about a particular AI-based system used to gather evidence. Consequently, the case law of Dutch courts plays a key part in the development of defence rights in the context of gathering (in fact, producing) data through AI-based systems. Since 2018, there has been a surge of court cases concerning cryptophones (phones that use encryption for the purpose of anonymous communication), in which the Hansken system has been used in order to gather evidence from huge digital data sets. In 2016, a whole server was seized by the Dutch police in order to access the content of encrypted communications ('Ennetcom cases'). And in 2020, the EncroChat cryptophones of more than 30.000 users were hacked by the French police, acting in cooperation with the Dutch police ('EncroChat cases').

Dutch courts are generally rather reluctant to request information on the functioning of Hansken from the NFI or to provide such information to the defence. Courts also quickly reject motions questioning the reliability of the functioning of Hansken (and the evidence gathered through it) from the defence. In general, Dutch judges seem to consider that the functioning of this AI-based system is unproblematic. For instance, the Amsterdam court

[19] See Lonneke Stevens and Maša Galič, 'Bescherming van Het Professionele Verschoningsrecht in Geval van Doorzoeking van Een Smartphone: Het EHRM Eist Een Concrete Basis En Een Praktische Procedurele Regeling in Het Recht' (2021) 70 Ars Aequi 845.
[20] ibid 851.

stated in a 2018 judgment, that Hansken was merely used in order to *view* (not even to gather) the evidence already collected, so that no specific legal basis is needed for its use.[21] Judges also seem to have a largely uncritical belief into the proper functioning of Hansken, perhaps related to the fact that the system has been developed 'in house', rather than by a private actor with commercial interests in mind. This 'presumed correctness' can be seen in a judgment by the Gelderland court, which ruled with very brief reasoning that the incompleteness of the results due to a software update, had no bearing on the integrity of the results and that the defence did not manage to prove otherwise.[22] Such attitude of the judges has important consequences, as it reduces the possibility of the defence to question and test the reliability of evidence gathered in this way.

Nevertheless, based on Article 182 CCP, the defence has the possibility to request the investigatory judge to carry out certain additional investigative acts. This general provision is in principle broad enough so as to enable the defence to propose their own search terms for the purpose of sifting through the data set with Hansken, as well as to request access to the data set and Hansken itself.[23] Dutch courts have already recognised the right of the defence to propose additional search terms, with which the prosecution will then search the whole data set (where the court reserves the right to assess, whether the proposed search terms are of sufficient relevance).[24] In this context, it should be noted that in Dutch law, it is for the prosecution generally to determine what information is relevant in the case. Only this information will then form part of the case file (Article 149a CCP) and be made available to the defence (Arts. 30-34 CCP).[25] While the defence can request the prosecutor to add information to the case file (Art. 34 CCP; e.g., by proposing additional search terms, with which a data set is to be searched), the prosecutor – with approval from the investigatory judge – may deny this request, if they consider it unsubstantiated. However, substantiating such a request can be a difficult task for the defence when it comes to huge data sets. After all, such data sets are comprised of hundreds of thousands (or even millions) of data points, stemming from numerous persons, so that specifying what one is looking for might be compared to looking for a needle in a haystack. Thus, if the requirement to substantiate such a request is set too high, the defence

[21] District Court of Amsterdam, judgment of 19 April 2018, ECLI:NL:RBAMS:2018:2504 (case nr. 13/997097-16), para. 7.3.

[22] District Court of Gelderland, judgment of 26 June 2019, ECLI:NL:RBGEL:2019:2833 (case nr. 05/780092-17), p. 9.

[23] In the Ennetcom-Tandem case (District Court of Amsterdam, judgment of 19 April 2018, ECLI:NL:RBAMS:2018:2504 (case nr. 13/997097-16, para. 7.3), the Amsterdam court stated that the defence had the possibility to expand the Tandem data set by asking the investigatory judge to approve additional search terms (but the defence did not make use of this possibility).

[24] See e.g., Court of Appeal Amsterdam, intermediate decision of 8 July 2020, ECLI:NL:GHAMS:2020:1904 (case nr. 23-002697-19), p. 13.

[25] This arrangement will not change much in the modernisation process of the CCP. The provisions regulating this are still based on the assumption that we are dealing with physical (i.e., paper) documents, which include findings including the reporting and interpretation of a selection of those data, rather than digital data sets themselves.

may be largely excluded from participating in the process of determining what is relevant in the case (this issue and the requirements of Art. 6 ECHR are further discussed in section 2.3).

In order for the defence to participate in this process, direct access to both the data set as well as to Hansken is thus desirable. However, according to Art. 182(3) CCP this request needs to be justified. While the law itself does not specify how precise this justification needs to be, Dutch courts generally require rather concrete specification of what the defence is looking for and why. Initially, requests for access by the defence – both to the data set and the Hansken tool itself – were rejected by courts, considered to be mere 'fishing expeditions.'[26] This began to change in 2021, with courts recognising that the defence needs to be afforded with the opportunity not only to examine the evidence against the defendant, but also to search for exculpatory evidence in the data set gathered by the prosecution. Nevertheless, Dutch courts still grant different scopes of access to the secondary data set (that is, the data set resulting from the initial searches with the search terms proposed by the prosecution and the defence in the full data set gathered in the case) to the defence. Some courts still deny access to this data set, considering that the request of the defence for such access was not substantiated enough.[27] Other courts either grant access to those messages and other data directly pertaining to the accused person, or the whole secondary data set to which the prosecution has access.[28] Nevertheless, based on case law from 2018 to 2021, it seems that with time, courts are granting broader access to the secondary data set to the defence.

Another issue concerns the *form* of the access to the secondary data set. Again, courts are granting different types of access, something which is also changing with time. Defence lawyers are generally provided with an Excel and/or PDF file with the relevant data. In addition, courts increasingly grant access to the same data set via Hansken, but this can only take place during a scheduled appointment at the Netherlands Forensics Institute. According to the prosecution, this limitation is due to practical considerations, which is planned to change in the near future, therefore granting access to defence lawyers to the data set with the use of Hansken via their own computers (something that should indeed be possible, considering that Hansken functions as a cloud-based service).[29]

---

[26] See e.g., Court of Appeal Amsterdam, judgment of 14 December 2018, ECLI:NL:GHAMS:2018:4620 (case nr. 23-00107717), section 8 (concerning a large data set gathered through the means of a key-logger).

[27] See e.g., District Court of The Hague, judgment of 25 August 2021, ECLI:NL:RBDHA:2021:9368 (case nr. 09/095750-21).

[28] See e.g., District Court of Rotterdam, intermediate decision of 25 January 2021, ECLI:NL:RBROT:2021:396; District Court of Rotterdam, intermediate decision of 15 July 2021, ECLI:NL:RBROT:2021:6853, para. 4; District Court of Amsterdam, intermediate decision of 1 April 2021, ECLI:NL:RBAMS:2021:1507 (case nr. 26Marengo); District Court of Rotterdam, intermediate decision of 25 June 2021, ECLI:NL:RBROT:2021:6113.

[29] The NFI are already working on this possibility, as presented by Hans Henseler and Harm van Beek, 'Hands-on with Hansken' (presentation at Bijzonder Strafrecht Cybercrime Congres, Den Haag, 3 December 2021) <https://www.hansken.nl/latest/news/2021/12/08/hands-on-with-hansken-at-the-cybercrime-congress-2021> accessed 14 January 2022.

Hansken, which was developed with the values of security and transparency in mind, also provides for automatic logging of activity while searching for evidence in the mass of data. As such, it would be fairly easy – at least from a technological perspective – to grant the defence (or an expert acting on behalf of the defence) access to these logging data in order to check, whether the prosecution's search activity was done in accordance with the law (e.g., whether they also gathered exculpatory evidence, and whether the system was functioning properly). This right has, however, not yet been granted to the defence.

2.3 Legal commentary

There is quite some discussion among Dutch scholars on the way Hansken, and similar AI-based system for the gathering of evidence, affect the right to a fair trial, especially equality of arms. Scholars generally argue for broader access of the defence to the gathered data set (in particular, the secondary data set, which is the result of the initial search of the full data set searched with the AI-tool) and to the AI-tool itself.[30]

On the basis of recent case law of the ECtHR concerning large data sets and Article 6 ECHR,[31] Galič argues that the defence is entitled to broad access to the secondary data set, without a strict requirement to justify such access. While the defence generally needs to justify any further search activity it is requesting (so as to prevent fishing expeditions), the particular context of huge data sets calls for a looser standard. When searching an enormous data set with millions of data points, one generally does not – in fact, cannot – know what one is searching for until they actually find it. In the case of the Ennetcom server, which contained data of about 19.000 users (at least some of whom might in some way be related to the accused), the accused simply could not have a proper idea of what might be found there. A requirement to specify what is being searched for would thus severely underestimate the complexities of analysing huge and interconnected amounts of data. It also does not offer the defence a comparable opportunity to that of the prosecution, which can search this data set repeatedly in order to refine their search terms; that is, in order to refine what exactly they are looking for. This has a serious effect on the principle of equality of arms.[32]

Scholars also argue that the defence should have access to the AI-tool itself, as they can hardly efficiently and effectively search the data set without it. As such, adequate access to the secondary data set must include access to the tool. Schermer and Oerlemans have,

[30] Maša Galič, 'De rechten van de verdediging in de context van omvangrijke datasets en geavanceerde zoekmachines in strafzaken: een suggestie voor uitbreiding' (2021) 2 Boom Strafblad 41; Bart Schermer and Jan-Jaap Oerlemans, 'AI, Strafrecht En Het Recht Op Een Eerlijk Proces' (2020) 1 Computerrecht 14.
[31] In particular, the following two judgments from 2019: ECtHR, 4 June 2019, ECLI:CE:ECHR:2019:0604JUD003975715, app. no. 39757/15 (Sigurður Einarsson and others v. Iceland); ECtHR, 25 July 2019, ECLI:CE:ECHR:2019:0725JUD000158615, app. no. 1586/15 (Rook v. Germany).
[32] See e.g., Galič (n 30); Custers and Stevens (n 8).

for instance, proposed granting access to the tool via a 'data room', where the defence could easily – but in a controlled environment – search the data set with Hansken.[33]

Furthermore, Galič argues for an expansion of the right of the defence to test the reliability of evidence produced with AI-based tools.[34] For this purpose, she first argues for increased transparency concerning the use of the AI-tool (rather than transparency concerning the source code, which is not likely to become public in relation to Hansken and similar systems), such as access to the logging reports concerning the search activities that the investigatory officers performed on the data set(s). Hansken already provides for automatic logging of search activities, so this would be simple to implement from a technical point of view. Second, she proposes that AI-based systems such as Hansken should be considered as expert evidence, which allow for additional testing for the purpose of reliability and afford the defence with the right to counter-expertise.

3 Production of evidence through AI-based systems

3.1 The example of CATCH: a facial recognition system

The Dutch police use facial recognition software called CATCH (short for 'Centrale Automatische TeChnologie voor Herkenning'). CATCH compares an image (a still from a video or a photograph) with a large database of current or past suspects and convicted persons that the Dutch police has gathered (consisting of 2,2 million images of 1,3 million persons).[35] Under certain circumstances, images may also be compared with a database of facial images of foreigners (without any requirement of suspicion), which consist of approximately 7 million images.[36] As such, CATCH does not (yet) perform real-time facial recognition, where the video feed of a particular individual (or set of individuals) from a camera would in real-time be compared with images in a particular database. However, real-time facial recognition is likely to be used by the Dutch police in the near future.[37]

---

[33] Schermer and Oerlemans (n 30) 10; see also JH de Wildt, 'Een Blik over de Grenzen: Vertrouwelijkheid, Data Rooms En Confidentiality Rings' (2017) Sanctierecht & Onderneming.

[34] Galič (n 30).

[35] 'Antwoorden Kamervragen over Het Bericht "Gezichtendatabase van Politie Bevat Foto's van 1,3 Miljoen Mensen"' (Ministerie van Justitie en Veiligheid, 10 September 2019) 3 <https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/kamerstukken/2019/09/10/antwoorden-kamervragen-over-het-bericht-gezichtendatabase-van-politie-bevat-foto-s-van-1-3-miljoen-mensen/antwoorden-kamervragen-over-het-bericht-gezichtendatabase-van-politie-bevat-foto-s-van-1-3-miljoen-mensen.pdf> accessed 14 January 2022.

[36] 'Aanhangsel van de Handelingen: Nr. 584, 2019/2020' (Tweede Kamer, 2019) 1 <https://zoek.officielebekendmakingen.nl/ah-tk-20192020-584.html> accessed 14 January 2022.

[37] See e.g., Anton Mous, 'Gezichtsherkenning in real time vindt wél plaats in Nederland' (*Vpngids* 14 December 2021) <https://www.vpngids.nl/nieuws/gezichtsherkenning-in-real-time-vindt-wel-plaats-in-nederland/> accessed 14 January 2022.

CATCH may only be used for the purpose of investigation of crimes for which a prison sentence of four years or more is prescribed. However, this set of crimes includes relatively minor crimes, such as theft, (WhatsApp-)scam and car burglary. According to the police, the system is employed, 'if the (possible) identity of the person on an image carrier would substantially contribute to the prevention, detection or prosecution of criminal offences.'[38]

3.2 The normative framework for the use of facial recognition systems

3.2.1. The legal framework

There are no specific rules concerning the use of facial recognition systems or the evidence produced by such systems in the Netherlands (nor are any proposed in the modernisation project). Such evidence is regulated by general rules concerning the lawfulness and reliability of evidence as described in section 2.2. The evidence generated by such systems can be challenged in the same way as the evidence generated by the Hansken system.

As a consequence of the distinct regulation of the collection of data and the subsequent processing of data for law enforcement purposes (described in the part of the report on predictive policing in the Netherlands), the use of facial recognition systems is regulated only by legal rules for the creation of databases of facial images of persons and general data protection rules for their subsequent processing. As such, there is no specific legal basis for the use of facial recognition technology in the CCP (or elsewhere). Facial recognition is thus seen only as a 'regular' technique for the processing of personal data. In this legal vacuum, comparable to the one relating to predictive policing, the police use facial recognition technology on the basis of the general police task (Article 3 Police Act), in combination with the provisions on the general police tasks as found in Articles 141 and 142 CCP. This also means that the use of this system does not require an authorisation from the investigatory judge.[39] As already discussed, these general legal bases only suffice in cases, leading to a minor intrusion into privacy. It is thus doubtful, whether they may be used in relation to facial recognition, which is commonly considered as highly intrusive, especially considering that it involves the processing of biometric – that is, sensitive – personal data.[40]

---

[38] 'Centrale Automatische TeChnologie Voor Herkenning (CATCH) Jaarcijfers 2020' (Politie, 2020) <https://www.politie.nl/binaries/content/assets/politie/onderwerpen/forensische-opsporing/catch-jaarcijfers-2020-hr-online.pdf> accessed 14 January 2022.

[39] 'Aanhangsel van de Handleidingen, Nr. 3932, 2018/2019' (Ministerie van Justitie en Veiligheid, 13 September 2019) 5 <https://zoek.officielebekendmakingen.nl/ah-tk-20182019-3932.html> accessed 14 January 2022.

[40] Cf. Commissie modernisering opsporingsonderzoek in het digitale tijdperk, 'Regulering van opsporingsbevoegdheden in een digitale omgeving' (2018) <https://kennisopenbaarbestuur.nl/documenten/rapport-commissie-koops-regulering-van-opsporingsbevoegdheden-in-een-digitale-omgeving/> accessed 14 January 2022; see also Proposal for a Regulation of the European Parliament and of the Council laying

The legal basis for the collection of facial images (and the creation of a database) is found in Article 55c CCP. Paragraphs 1-4 of Article 55c CCP regulate the taking of photos and fingerprints of persons suspected of crimes, for which a prison sentence of four years or more is prescribed. According to the fourth paragraph of this provision, the images (and fingerprints) can be further processed for the purpose of prevention, detection, prosecution and adjudication of criminal offences. These data can be stored for a very long time, between 20 and 80 years.[41]

The legal basis for further processing is regulated by data protection law in the Police Data Act (PDA). Photographs that are used for facial recognition constitute biometric data and are as such 'sensitive personal data'. In line with EU data protection law, the processing of this type of data is regulated more strictly in the PDA. Processing is only permitted if it is 'unavoidable' (Art. 5 PDA) for the purpose pursued. This means that its processing must be substantiated in a particularly precise manner, including stricter limitations on storage. However, the Dutch police are struggling with these obligations. It was recently revealed that the police are not complying with its obligation to delete photos of persons who are no longer a suspect or were acquitted in subsequent proceedings.[42] In 2020, the police stated that they have deleted more than 200.000 images, but it remains unclear how many individuals have been removed from the database.[43]

3.2.2. Reliability and neutrality of AI-based systems producing evidence[44]

Specifically in relation to the CATCH facial recognition system, the reliability and neutrality of the technology are preserved in the guidelines for the use of the system, which require a 'double human verification' in the decision-making process.[45] The procedure of double human verification is designed to reduce the risk of false positives (i.e., incorrectly assumed matches) and to protect the rights of data subjects.[46] After the CATCH system performs the comparison between the images, it gives an overview of the faces with the most similarities, including scale scores. After the comparison, the AI-generated list of candidates is presented to a trained expert. If the expert believes that there is indeed a match with one of the candidates, the match is shown to two other experts who assess the match independently (it is unknown what kinds of experts are meant here and in which way they are trained). If the experts do not come to the same conclusion, the

down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts 2021 [COM(2021) 206 final].

[41] 'Aanhangsel van de Handleidingen, Nr. 3932, 2018/2019' (n 39) 2.

[42] 'Police Remove 218,000 Photos from Facial Recognition Database' (*Dutch news*, 23 July 2021) <https://www.dutchnews.nl/news/2021/07/police-remove-218000-photos-from-facial-recognition-database/> accessed 14 January 2022.

[43] ibid.

[44] For a general discussion, see description in relation to Hansken in sections 2.2 and 2.3.

[45] 'Kamerbrief over Gebruik Gezichtsherkenningstechnologie: Waarborgen En Kaders Bij Gebruik Gezichtsherkenningstechnologie' (Ministerie van Justitie en Veiligheid, 20 November 2019) 2–3 <https://www.rijksoverheid.nl/documenten/kamerstukken/2019/11/20/tk-waarborgen-en-kaders-bij-gebruik-gezichtsherkenningstechnologie> accessed 14 January 2022.

[46] ibid.

most conservative conclusion is reported.[47] Even when the experts come to the same conclusion, this only results in an 'indication' that the suspect matches the person on the image.[48] The use of CATCH therefore does not lead to claims of a definitive identification of the suspect.

This has been confirmed in a 2019 judgment of the Zeeland-West-Brabant District Court,[49] which concluded that the results of the CATCH system, even after they have been 'confirmed' by two human experts, alone do not suffice for a criminal conviction (further discussed in the following section); additional corroborating evidence is necessary. This requirement that AI-generated evidence is corroborated by other evidence thus indirectly guarantees the reliability and neutrality of such systems.

3.2.3. Case law

So far, there has been only one judgment concerning the use of facial recognition software.[50] In the abovementioned 2019 judgment, the Zeeland-West-Brabant court only briefly discussed the validity of evidence that was produced by it, stating:

> The court is of the opinion that in this case the "hit" on the suspect in the so-called CATCH system (Central Automatic Technology for Recognition) is insufficient to conclude – beyond reasonable doubt – that the suspect can be designated as the person using the ATM machine. The observation that two investigators saw that there were many similarities and no significant deviations, is not considered so convincing by the court that the "hit" can serve as a basis for a proven conclusion. As there is no other evidence besides the recognition that links the accused to any of the charges, the court is of the opinion that the accused should be acquitted.[51]

According to Dutch evidence law, one source of evidence does not suffice for a conviction (with the exception of a police officer personally observing a crime taking place; Art. 344(2) CCP). In regard to evidence *linking* the suspect to the offence, however, one source of evidence is sufficient, as long as other evidence of the crime exists, which is independent of the link between the suspect and the crime (e.g., money has been withdrawn from an ATM with a stolen bankcard). Despite the fact that the law does not require this, the Zeeland-West-Brabant court required corroborating evidence for the purpose of establishing the link between the suspect and the crime (e.g., eyewitness testimony or matching DNA at the scene). This means that the court did not consider AI-produced evidence

[47] 'Antwoorden Kamervragen over Het Bericht "Gezichtendatabase van Politie Bevat Foto's van 1,3 Miljoen Mensen"' (n 35) 5; see also 'Kamerbrief over Gebruik Gezichtsherkenningstechnologie: Waarborgen En Kaders Bij Gebruik Gezichtsherkenningstechnologie' (n 45) 2–3.
[48] District Court of Zeeland-West-Brabant, judgment of 17 May 2019, ECLI:NL:RBZWB:2019:2191 (case nr. 02-665274-18), para. 4.3.
[49] ibid.
[50] ibid.
[51] ibid., para. 4.3; translation by the authors.

through the CATCH system (despite the confirmation by humans) as sufficient in establishing the link between the suspect and the crime. In this way, the court indirectly ensured the reliability and neutrality of evidence produced by AI-based systems.

3.2.4. Information provided by AI-based systems used by non-investigative authorities

As already mentioned in section 2.2.1, the draft CCP introduces a new provision, on the basis of which the public prosecutor may order companies and institutions to process certain data and then provide only the 'results' to the police (draft Article 2.7.51 CCP). Based on the broad wording of the provision and the Explanatory Memorandum, it seems that non-investigative authorities (e.g., companies such as Google or Facebook) may indeed provide data to law enforcement that has been processed – that is, produced – through an AI-based system. While the Explanatory Memorandum does not speak specifically of AI techniques, it does state that advanced types of processing, which lead to the generation of 'new data', are meant here. This broad definition thus likely includes the use of AI.

The last two paragraphs of the provision provide for important safeguards in relation to the reliability of the data generated in this way. According to paragraph 3 of Article 2.7.51 CCP, the public prosecutor may require that the person carries out the processing in accordance with the instructions of the investigating officer. As the Explanatory Memorandum put it:

> 'This paragraph therefore offers the possibility of setting requirements for the execution, also with regard to the verifiability of the processing afterwards. One of the instructions of the investigating officer could be to describe the exact procedure of the analysis or to have the analysis checked or repeated by a second person. An instruction can also be that the analysis must take place in the presence and under the supervision of an investigating officer or another expert. In this respect, it will play a role whether the order is addressed to a large company that regularly carries out such analyses for the purpose of investigation or to a relatively small company that is perhaps considered less reliable. In the latter case, it is obvious that the investigation will play a major role, for example by supporting the analysis by supplying hardware and software.'[52]

On the one hand, this provision offers a safeguard that is badly needed in order to strengthen the reliability and transparency of the processing and the data generated through it. On the other hand, the Explanatory Memorandum suggests an assumption of validity and reliability, when the processing is performed by 'large companies' that have knowledge and experience with data analysis. Not only is such an assumption mis-

[52] Ambtelijke Versie Juli 2020 Memorie van Toelichting Wetboek van Strafvordering' (n 15) 443–444.

placed (e.g., algorithms used by large companies such as Facebook and Google have oftentimes been found biased),[53] it is also unclear what the role of the defence is in this regard. Do they have a say, when the public prosecutor is considering, whether and in which way to instruct the company in regard to the prosecutor? The Explanatory Memorandum does not include any discussion on this.

The power granted in paragraph 3 of the provision is further strengthened by the power in paragraph 4. Paragraph 4 states that companies and institutions may be ordered to provide information 'about the data to which they have access' and about 'the actions required to carry out the processing referred to in the first paragraph'. The possibility of the public prosecutor to ask questions in advance about the (composition of the) data set and the effort that a company must make to perform a certain analysis, namely enables the prosecutor to assess whether an order for data analysis is useful and, if so, which conditions (as referred to in the third paragraph) should be imposed.[54] As such, para. 4 is of particular relevance in regard to AI-systems used for data processing. Depending on the interpretation of this requirement – do the 'actions required to carry out the processing' include technical steps taken by the system? – the prosecution thus might have the power to request further information concerning the manner in which the AI-tool functions and processes the data. A further question, again, relates to the defence: do or could they have access to this information? Such access would surely be needed in order to create an adequate safeguard for the reliability of AI-generated data that might serve as evidence in criminal cases.

3.2.5. Regional and international agreements on the admissibility of evidence

Two regional instruments might be mentioned here. The first is the proposed EU e-Evidence Regulation,[55] which is intended to facilitate access to electronic evidence by European police and judicial authorities. The draft e-Evidence Regulation focuses on 'data cooperation' and seeks to provide an alternative to the existing mutual legal assistance framework. The second is the second protocol to the Budapest convention (Convention on Cybercrime) of the Council of Europe on enhanced international cooperation and access to evidence in the cloud.[56] Unfortunately, neither of these instruments seems to have touched upon a key problem: the quality – and, thus, admissibility – of what is to be

[53] See e.g., Michael Walker, 'Upheaval at Google Signals Pushback against Biased Algorithms and Unaccountable AI' (*The Conversation*, 10 December 2020) <https://theconversation.com/upheaval-at-google-signals-pushback-against-biased-algorithms-and-unaccountable-ai-151768> accessed 14 January 2022; Karen Hao, 'How Facebook Got Addicted to Spreading Misinformation' (*MIT Technology Review*, 11 March 2021) <https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/> accessed 14 January 2022.

[54] 'Ambtelijke Versie Juli 2020 Memorie van Toelichting Wetboek van Strafvordering' (n 15) 444.

[55] Proposal for a Regulation of the European Parliament and of the Council on the European Production and Preservation Orders for electronic evidence in criminal matters 2018 [COM(2018) 225 final].

[56] Second Additional Protocol to the Convention on Cybercrime on enhanced co-operation and disclosure of electronic evidence 2021 [CM(2021)57-final].

exchanged. To this date, the proposals do not contain a single provision on how to reliably collect, analyse and present the material. There are, however, calls for the EU legislator to incorporate human rights standards in a new harmonising instrument on admissibility of evidence in criminal matters, for example in a dedicated Admissibility Directive.[57]

4 Evidence assessed through AI-based systems

To the best of our knowledge, AI-based systems used for assessing evidence are not (yet) used in the Netherlands, nor is there any significant debate on the matter. The only realistic example in which AI-based systems would actually assess criminal evidence, can be found in deepfake detection systems for the purpose of detecting fake images, videos or audio files among evidence. While it is unknown, whether the police already use such systems, on what scale and for which purposes, it can nevertheless be said that the development of such systems to be used in law enforcement has certainly begun in the Netherlands.[58]

5 Conclusion

We examined two types of AI-based systems used for the production of evidence: Hansken, a tool for the gathering of data out of huge data sets, and CATCH, a facial recognition tool. Even though Hansken is commonly described as a tool for the gathering of evidence from huge data sets, we argue that such systems actually do more than merely gather evidence that already exists: they produce it. This is so, because the system first needs to interpret the data by itself (e.g., a system searching for images of drugs needs to be able to determine that a particular photo indeed represents drugs). Second, it needs to be able to find relevant correlations (that is, links) between the numerous data points in the data set (e.g., resulting in a convincing time-line and scenario). Consequently, we need to talk about production of evidence, both in relation to Hansken as well as the CATCH facial recognition system.

Despite the fact that digital evidence plays an increasingly important role in contemporary criminal proceedings, Dutch law (including the draft Code of Criminal Procedure, which is the result of the ongoing Modernisation project) has yet to implement any significant changes to its rules relating to evidence. As such, the few rules that regulate the gathering of evidence do not fit the particular needs of digital evidence very well. This leads to, for instance, issues with the principle of equality of arms. Considering the way digital evidence is gathered and examined, the defence needs additional or broader

[57] See e.g., Balázs Garamvölgyi and others, 'Admissibility of Evidence in Criminal Proceedings in the EU' (2020) 3 Eucrim: the European Criminal Law Associations' Forum <https://eucrim.eu/articles/admissibility-evidence-criminal-proceedings-eu/ > accessed 6 January 2023.

[58] See 'UvA En NFI Doen Onderzoek Naar Herkennen Deepfakes En Verborgen Berichten van Criminelen' (Universiteit van Amsterdam, 22 May 2021) <https://www.uva.nl/content/nieuws/persberichten/2021/05/uva-en-nfi-doen-onderzoek-naar-herkennen-deepfakes-en-verborgen-berichten-van-criminelen.html?cb> accessed 14 January 2022.

rights in order to participate in determining what counts as relevant information in a particular case, to participate in searching for exculpatory evidence, and to question the validity and accuracy of the functioning of AI-based systems. We can see that such rights are slowly being developed through case law.

Selected literature

'Ambtelijke Versie Juli 2020 Memorie van Toelichting Wetboek van Strafvordering' (Ministerie van Justitie en Veiligheid, 30 July 2020) 442 <https://www.rijksoverheid.nl/documenten/publicaties/2020/07/30/ambtelijke-versie-juli-2020-memorie-van-toelichting-wetboek-van-strafvordering> accessed 14 January 2022.

'Aanhangsel van de Handleidingen, Nr. 3932, 2018/2019' (Ministerie van Justitie en Veiligheid, 13 September 2019) 5 <https://zoek.officielebekendmakingen.nl/ah-tk-20182019-3932.html> accessed 14 January 2022.

'Aanhangsel van de Handelingen: Nr. 584, 2019/2020' (Tweede Kamer, 2019) 1 <https://zoek.officielebekendmakingen.nl/ah-tk-20192020-584.html> accessed 14 January 2022.

'Antwoorden Kamervragen over Het Bericht "Gezichtendatabase van Politie Bevat Foto's van 1,3 Miljoen Mensen"' (Ministerie van Justitie en Veiligheid, 10 September 2019) 3 <https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/kamerstukken/2019/09/10/antwoorden-kamervragen-over-het-bericht-gezichtendatabase-van-politie-bevat-foto-s-van-1-3-miljoen-mensen/antwoorden-kamervragen-over-het-bericht-gezichtendatabase-van-politie-bevat-foto-s-van-1-3-miljoen-mensen.pdf> accessed 14 January 2022.

Bas Seyyar M and Geradts Z, 'Privacy Impact Assessment in Large-Scale Digital Forensic Investigations' (2020) 33 Forensic Science International: Digital Investigation 1, 4.

Harm van Beek and others, 'Digital Forensics as a Service: Game On' (2015) 15 Digital Investigation 20.

'Centrale Automatische TeChnologie Voor Herkenning (CATCH) Jaarcijfers 2020' (Politie, 2020) <https://www.politie.nl/binaries/content/assets/politie/onderwerpen/forensische-opsporing/catch-jaarcijfers-2020-hr-online.pdf> accessed 14 January 2022.

Commissie modernisering opsporingsonderzoek in het digitale tijdperk, 'Regulering van opsporingsbevoegdheden in een digitale omgeving' (2018) <https://kennisopenbaarbestuur.nl/documenten/rapport-commissie-koops-regulering-van-opsporingsbevoegdheden-in-een-digitale-omgeving/> accessed 14 January 2022.

Custers B and Stevens L, 'The Use of Data as Evidence in Dutch Criminal Courts' (2021) 29 European Journal of Crime, Criminal Law and Criminal Justice 25, 40.

'Dutch Police Seize Encrypted Communication Network with 19,000 Users' (Reuters, 22 April 2016) <https://www.reuters.com/article/us-netherlands-cyber-idUSKCN0XJ2HQ> accessed 14 January 2022.

Galič M, 'De rechten van de verdediging in de context van omvangrijke datasets en geavanceerde zoekmachines in strafzaken: een suggestie voor uitbreiding' (2021) 2 Boom Strafblad 41.

Garamvölgyi B and others, 'Admissibility of Evidence in Criminal Proceedings in the EU' (2020) 3 Eucrim: the European Criminal Law Associations' Forum <https://eucrim.eu/articles/admissibility-evidence-criminal-proceedings-eu/ > accessed 6 January 2023.

Hao K, 'How Facebook Got Addicted to Spreading Misinformation' (MIT Technology Review, 11 March 2021) <https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/> accessed 14 January 2022.

Henseler H and Harm van Beek, 'Hands-on with Hansken' (presentation at Bijzonder Strafrecht Cybercrime Congres, Den Haag, 3 December 2021) <https://www.hansken.nl/latest/news/2021/12/08/hands-on-with-hansken-at-the-cybercrime-congress-2021> accessed 14 January 2022.

'Kamerbrief over Gebruik Gezichtsherkenningstechnologie: Waarborgen En Kaders Bij Gebruik Gezichtsherkenningstechnologie' (Ministerie van Justitie en Veiligheid, 20 November 2019) 2–3 <https://www.rijksoverheid.nl/documenten/kamerstukken/2019/11/20/tk-waarborgen-en-kaders-bij-gebruik-gezichtsherkenningstechnologie> accessed 14 January 2022.

Mous A, 'Gezichtsherkenning in real time vindt wél plaats in Nederland' (Vpngids 14 December 2021) <https://www.vpngids.nl/nieuws/gezichtsherkenning-in-real-time-vindt-wel-plaats-in-nederland/> accessed 14 January 2022.

'Police Remove 218,000 Photos from Facial Recognition Database' (Dutch news, 23 July 2021) <https://www.dutchnews.nl/news/2021/07/police-remove-218000-photos-from-facial-recognition-database/> accessed 14 January 2022.

Schermer B and Oerlemans J-J, 'AI, Strafrecht En Het Recht Op Een Eerlijk Proces' (2020) 1 Computerrecht 14.

Stevens L and Galič M, 'Bescherming van Het Professionele Verschoningsrecht in Geval van Doorzoeking van Een Smartphone: Het EHRM Eist Een Concrete Basis En Een Praktische Procedurele Regeling in Het Recht' (2021) 70 Ars Aequi 845.

'UvA En NFI Doen Onderzoek Naar Herkennen Deepfakes En Verborgen Berichten van Criminelen' (Universiteit van Amsterdam, 22 May 2021) <https://www.uva.nl/content/nieuws/persberichten/2021/05/uva-en-nfi-doen-onderzoek-naar-herkennen-deepfakes-en-verborgen-berichten-van-criminelen.html?cb> accessed 14 January 2022.

Walker M, 'Upheaval at Google Signals Pushback against Biased Algorithms and Unaccountable AI' (The Conversation, 10 December 2020) <https://theconversation.com/upheaval-at-google-signals-pushback-against-biased-algorithms-and-unaccountable-ai-151768> accessed 14 January 2022.

de Wildt J H, 'Een Blik over de Grenzen: Vertrouwelijkheid, Data Rooms En Confidentiality Rings' (2017) Sanctierecht & Onderneming.

SPECIAL REPORTS

THE PORTUGUESE CHARTER OF HUMAN RIGHTS IN THE DIGITAL AGE – BRIEF REMARKS ON ARTICLE 9

By Anabela Miranda Rodrigues and Eduardo A. S. Figueiredo***

Abstract

In 2021, the Assembly of the Republic approved the Portuguese Charter of Human Rights in the Digital Age, calling upon our country to occupy a leading position in the global movement of digital constitutionalism. In sum, the Charter consecrates a set of fundamental rights and freedoms of the human person that must be respected, protected and promoted by public and private entities in the cyberspace. Therefore, in the present article, we will reflect on the underlying political-legislative intention of this document, as well as some of the juridical problems it poses. Also, special attention will be given to the content of its Article 9, concerning the use of Artificial Intelligence and robots.

1 Introduction

The Portuguese Charter of Human Rights in the Digital Age (hereinafter 'the Charter') derives from two legislative projects submitted before the Assembly of the Republic – the Portuguese Parliament – in late 2020: the first was submitted by the *Socialist Party*, which referred to a 'Charter of Fundamental Rights', and the second by the *'People, Animals, Nature'* (PAN) *Party*, which referred to a 'Charter of Digital Rights'.[1] Both draft bills were then merged in one single project to be discussed by the parliamentarians, whose final text referred instead to a Charter of 'Human' Rights. It's worthy to point out that the designation of the Charter as a Charter of 'Human' Rights is somewhat anomalous, since the expression 'human rights' is commonly used in the international context, and it is doubtful that the rights enshrined in the Charter – if one wants to be truly rigorous – can

* University of Coimbra Institute for Legal Research (UCILeR), Integrated Researcher. Faculty of Law of the University of Coimbra, Full Professor. E-mail address: anarod@fd.uc.pt

** University of Coimbra Institute for Legal Research (UCILeR), Collaborator Researcher. Faculty of Law of the University of Coimbra, Assistant. E-mail address: eduardo.figueiredo@uc.pt

[1] On July 9, 2020, the Socialist Party submitted his legislative project – Draft Bill No. 473/XIV/1: 'Approves the Charter of Fundamental Rights in the Digital Era' – before the Assembly of the Republic. Subsequently, on September 11, PAN presented the Draft Bill No. 498/XIV/1: 'Approves the Charter of Digital Rights and a set of complementary measures that ensure the reinforcement of citizens' guarantees in the digital domain' (whose initial text was replaced on September 28). In fact, these were not the first legislative initiatives in this matter: on May 15, 2019, a draft bill (Draft Bill No. 1217/XIII/4: 'Approves the Charter of Fundamental Rights in the Digital Era') had already been presented by the Socialist Party, but expired due to the end of the legislature some months after (following the dissolution of the Parliament by the President of the Republic).

be configured as such. However, attention must be drawn to the fact that this is, in a certain way, a matter of convention[2], considering the example of notable cases, such as the UK Human Rights Act of 1998 or the Charter of Fundamental Rights of the European Union of 2000.

The Charter was almost unanimously[3] approved in April 2021, promulgated by the President of the Republic, and published at the national Official Gazette ('*Diário da República*') as Law No. 27/2021, of May 17. It has then entered into force on July 18 of the same year. One may note, however, that the early days of this diploma were quite troubled. On the 29th of July of 2021, the President of the Republic requested the Constitutional Court to verify the constitutionality of Article 6 of the Charter ('Right to protection against disinformation'), not only considering the lack of clarity of some of the concepts used by the legislator, but also the intense public debate that had been triggered in the meantime concerning its content. Almost one year later, on the 18th of May of 2022, the Portuguese Ombudsman ('*Provedora de Justiça*') also required the Constitutional Court to declare the unconstitutionality of that same legal precept (although referring exclusively to its §5 and §6), for violation of the principles of supremacy of the rule of law and proportionality when restricting fundamental freedoms (freedoms of expression, information and of the press)[4]. Subsequently, Law No. 27/2021 was amended by Law No. 15/2022, of August 11, which 'amputated' Article 6, maintaining solely one paragraph establishing that 'the Portuguese State ensures compliance with the European Action Plan against Disinformation, in order to protect society against natural or legal persons, *de jure* or *de facto*, who produce, reproduce or disseminate narratives considered to be disinformation' (§1). The other five paragraphs were, thus, fully revoked: (1) those defining the concepts of 'disinformation' and of 'proven false or misleading information' (§2, §3 and §4); (2) the one entitling the Portuguese Regulating Authority for the Media ('*Entidade Reguladora para a Comunicação Social*') to receive complaints and apply sanctions to those who violated any norm of this legal precept (§5); and (3) the one affirming that 'the State supports the creation of fact-checking structures by registered media outlets and encourages the attribution of quality seals by reliable entities endowed with the status of public utility' (§6). With this specific exception, the remaining content of the Charter continues intact until today.

[2] Domingos Soares Farinho, 'The Portuguese Charter of Human Rights in the Digital Age: a legal appraisal' [2021] 13 *Revista Española de la Transparencia* 88.

[3] There were no votes against the adoption of the act and only 14 abstentions.

[4] In his Decision No. 66/2023, of March 7, the Constitutional Court refused to analyze both the requests made by the President of the Republic and by the Portuguese Ombudsman considering that, at the date of the analysis, the referred norms had already been repealed.

The present article will firstly focus on the legal nature of the Charter, which will be analysed through two different perspectives: on the one hand, its underlying political-legislative intention; on the other hand, some of the juridical problems it poses. Secondly, special attention will be given to the content of Article 9, concerning the use of Artificial Intelligence (AI) and robots.

2 Legal Nature of the Charter

The Charter is not a text that intends to obey the ideological and historical schemes of the constituent power. In any case, the Explanatory Memorandum of the Socialist Draft Bill did not fail to map out various initiatives tending to affirm a (*multilevel*)[5] *digital constitutionalism*.[6] In fact, in the digital age, fundamental rights and freedoms are confronted with new threats and subject not only to public interference, but also to private determinations.[7] Therefore, it is of utmost importance for constitutionalism to develop 'new ways of limiting abuses of power in a complex system that includes many different governments, businesses, and civil society organisations',[8] with the primary aim of protecting the *human person* and her *intrinsic dignity*. In this sense, digital constitutionalism includes a plurality of normative instruments translating constitutional values in the digital society, both emerging in the state context (such as constitutional and ordinary law) and beyond (self-regulation of private companies).[9] The approval of the Charter can be, without any doubts, included in this on-going movement.

In what specifically relates with AI, one has to recognize that, although the strengthening of fundamental rights and freedoms of the human person may not be enough to solve the challenges posed by the emergence and exponential evolution of intelligent systems[10], such efforts cannot be underestimated or even dispensed, considering the press-

[5] Ingolf Pernice, 'Multilevel Constitutionalism and the Crisis of Democracy in Europe' [2015] 11/3 *European Constitutional Law Review* 544-546.
[6] Explanatory Memorandum of the Socialist Bill (*note 1*) point 2.
[7] Giovanni De Gregorio, Digital Constitutionalism in Europe: Reframing Rights and Powers in the Algorithmic Society (Cambridge University Press 2022) 2.
[8] Nicolas P. Suzor, *Lawless: The Secret Rules That Govern Our Digital Lives* (Cambridge University Press 2019) 113.
[9] Edoardo Celeste, Digital Constitutionalism: the Role of Internet Bill of Rights (Routledge 2023) 84.
[10] Consider, for example, the challenges posed by the lack of transparency of those systems (the so-called "black box problem"), in conjunction with the theoretical and practical difficulties in shaping and making effective a subjective *right to explanation*, as proposed by Thomas Wischmeyer, 'Artificial Intelligence and Transparency: Opening the Black Box', in Thomas Wischmeyer and Timo Rademacher (eds), *Regulating Artificial Intelligence* (Springer 2020) 95-96.

ing need of building balances between power and responsibility in the current 'algorithmic society'.[11] Therefore, the main aim of the Charter is to stress and solidify the undisputable role played by (*conventional* and *digital*) rights and freedoms in a progressive context of *digit(al)ization* of the 'Life-World' (*Lebenswelt*).

This 'rights approach' must, nevertheless, be complemented by another one focusing on the development of an 'architecture of reasoning and control', as rightly observed by *Gomes Canotilho*.[12] In this sense, the Charter will serve as the *legal basis* in light of which competent authorities and courts, making use of their organizational structures and procedural mechanisms, will be able to undertake effective control and oversight activities, placing the human person at the beginning and end of the AI system. Again, in the words of the Draft Bill presented by the Socialist party:

> In this bill, an attempt was made to set out a diversified and comprehensive list of rights, freedoms and guarantees, which *innovates, clarifies and also serves as the basis for a binding action program* for the authorities.[13]

2.1 Political legislative intention

The Charter tried to apprehend the key transformations triggered by the digital revolution, as well as the way they *irradiate to* – or even *colonize* – fundamental normative precepts.

Particular attention must be paid to its legal form and value. In fact, the Charter was approved as a national ordinary law, under the Constitution of the Portuguese Republic of 1976 (hereinafter 'the Constitution'). In consequence, the rights and freedoms it enshrines cannot be qualified as *fundamental in a formal sense*, but only in a *material sense*, in line with the content of Article 16, §1, of the Constitution.[14] According to the dominant thesis in Portuguese literature, this constitutional precept refers to an *open perspective of fundamental rights and freedoms*, thus making it possible to admit their consecration outside the text of the Constitution. Nevertheless, it is worthy to note, with *Gomes Canotilho* and *Vital Moreira,* that the rights and freedoms enshrined in national legislative acts 'can only be qualified as "fundamental" … when they assume the same relevance – first and

[11] Marc Schuilenburg and Rik Peeters (eds), *The Algorithmic Society: Technology, Power, and Knowledge* (Routledge 2021) 1.

[12] José Joaquim Gomes Canotilho, 'Sobre a indispensabilidade de uma Carta de Direitos Fundamentais Digitais na União Europeia/About the indispensability of a Charter of Fundamental Rights to the European Union' [2019] 31/1 *Revista do Tribunal Regional Federal 1ª Região Brasília DF* 69-70.

[13] Explanatory Memorandum of the Socialist Bill (*note 1*) point 3 (emphasis added).

[14] Article 16, §1 (Constitution): 'The fundamental rights enshrined in the Constitution shall not exclude any others set out in applicable international laws and legal rules.'

foremost, due to their ethical and legal roots in the juridical conscience – as the rights established in the Constitution'.[15] There is, however, a minority thesis that prefers to adopt a formal approach to Article 16, §1, of the Constitution, arguing that the precept must be qualified as an 'upgrade norm', through which 'the Constitution upgrades the hierarchical position of fundamental rights norms contained in ordinary law'.[16] Anyway, from a *macro* perspective, one can say that the political-legislative intention underlying the Charter and, at the same time, its most worth noting added value is to clear any doubts surrounding the fact that 'the norms that in the Portuguese legal order consecrate and protect rights, freedoms and guarantees are fully applicable in the cyberspace' (Article 2, §2, of the Charter).

In contrast, from a *micro* perspective, one must distinguish between *fundamental digital rights in a broad sense* and *fundamental digital rights in a strict sense*:

In the first case, we are referring to those fundamental rights (formally or materially shielded by the Constitution that present *dimensions of digit(al)ization or that can be transposed to the digital sphere*. In this sense, the Charter aims to assert and reinforce the normative material scope of rights, freedoms and values enshrined in our fundamental law and/or to delimitate their scope of protection in a digital environment, stressing some specificities that must necessarily be taken into account. For example, the elementary value of *dignity of the human person*, affirmed in Article 1 of the Constitution, remains untouchable also in the cyberspace. Also, consider the protection conferred to the 'right to personal identity, good name and reputation, image and word', as well as 'the right to moral integrity' in a digital environment;[17] or, in the specific case of AI and robotics, the guarantee that the use of AI should always be guided by the respect for fundamental rights.[18]

In the second case, we are referring to those rights which *incorporate specific digital elements*. Several examples can be advanced in this regard: the right of unrestricted access

[15] José Joaquim Gomes Canotilho and Vital Moreira, *Constituição da República Portuguesa Anotada – Volume I* (4th edn, Coimbra Editora 2014) 365-366.
[16] Domingos Soares Farinho (*note 2*) 89-90.
[17] Article 12, §1 (Charter): 'Everyone has the right to personal identity, good name and reputation, image and word, as well as their moral integrity in a digital environment.'
[18] Article 9, §1 (Charter): 'The use of artificial intelligence must be guided by respect for fundamental rights …'.

to the internet, in non-discriminatory terms;[19] the right to cybersecurity[20] (whose specificities are evident inclusively at the level of its designation in a digital context, namely if one considers that the Constitution only refers to a 'right to freedom and security');[21] or, in the case of AI, the right to data protection.[22]

In conclusion, the Charter sought to set out a comprehensive and diverse list of fundamental digital rights and freedoms[23] – in some cases, with a mere clarifying purpose; in other cases, in a truly innovative manner –, which will serve both as *ground* and *bound* for the activity of public and private actors in the context of the digital and algorithmic age. A clear example of what is at stake can be found in the field of AI and robots, namely considering the simultaneous consecration of principles that must lead to its creation and use, as well as a set of rights of the decisions' recipient.[24] We will come back to this point later.

2.2 Juridical problems raised by the Charter

Despite these several positive aspects, the Charter has also been the object of some criticisms. As already suggested above, its most polemic aspect was the consecration of a 'right to protection from disinformation' (Article 6, in its *original version*) in such terms that a considerable part of the doctrine agreed that the legislator had gone too far in the fight against this hideous phenomenon. In fact, many were the voices noticing that the referred legal precept was more of a restriction to the fundamental freedoms of opinion and expression than a 'new' individual right, being truly doubtful that such restriction even respected the Constitution (namely, §2 and §3 of Article 18).[25] Some even sustained

---

[19] Article 3, §1 (Charter): 'Everyone, regardless of ancestry, gender, race, language, territory of origin, religion, political or ideological beliefs, education, economic situation, social circumstances or sexual orientation, has a right to unrestricted access to the internet.'

[20] Article 15 (Charter): "Everyone has the right to security in cyberspace …".

[21] Article 27, §1 (Constitution): 'Everyone has the right to freedom and security.'

[22] Article 8 (Charter) addresses, in general, the right to privacy in a digital environment. Its §2 specifically foresees that "the right to data protection, including control over its collection, registration, organization, structuring, conservation, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, diffusion or any other form of availability, comparison or interconnection, limitation, erasure or destruction, is ensured by law.".

[23] Apart from the rights that have already been mentioned, other examples can be given: the right of freedom of expression and creation in a digital environment (Article 4), the rights to meet, demonstrate, associate and participate in a digital environment (Article 7), the right to a neutral internet (Article 10), the right to develop digital skills (Article 11), the right to oblivion (Article 13), rights in digital platforms (Article 14), the right to freedom of creation and protection of contents (Article 16), the right of protection against abusive geolocation (Article 17), the right to a digital will (Article 18), *etc.*

[24] Article 9 (Charter), *infra*.

[25] Article 18 (Constitution): 'The law may only restrict rights, freedoms and guarantees in cases expressly provided for in the Constitution, and such restrictions must be limited to those needed to safeguard other constitutionally protected rights and interests.' (§2); 'Laws that restrict rights, freedoms and guarantees

that the hidden intention of such norms was to bring censorship back to Portugal, an accusation that sounded as worrying as exaggerated...[26] Nevertheless, the truth is that this vivid juridical, political and social uproar led to the amendment of Article 6, reducing its content to a mere programmatic norm with (almost) no practical relevance.

But the criticisms do not stop here. On the one hand, some authors consider that some precepts of the Charter are *redundant*, which might raise problems related to their interpretation and application. In fact, the mere duplication of existing rights, only adding a reference to the digital environment, can provoke 'disturbances in the legal system', regardless of its 'public policy value'.[27] And note, this criticism is particularly relevant considering, *inter alia*, the significant efforts pursued by the doctrine and the jurisprudence to coherently transpose 'conventional' human and fundamental rights and freedom to the cyberspace.[28] Although one might always argue that, in this sense, the precepts of the Charter actively contribute to the success of such efforts, there is always a risk that, in the end, they end up being compromised by repetitive, vague and/or dubious norms.

On the other hand, the Charter is recurrently accused of facilitating nonconformities and conflicts between its norms and others referring to the same rights and freedoms contained, for example, in juridical instruments adopted by the *United Nations*, the *Council of Europe* or even the *European Union* (EU).[29] As an example, consider the *right to data protection* or the *right to oblivion*, simultaneously consecrated in §2 of Article 8[30] and Article 13[31] of the Charter and in several provisions (*v.g.*, Articles 5, 15-18 and 20) of the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data – also known as 'General Data Protection Regulation'

must have a general and abstract nature and may not have a retroactive effect or reduce the extent or scope of the essential content of the constitutional precepts.' (§3).

[26] Luís Menezes Leitão, 'A inconstitucionalidade da Carta Portuguesa de Direito Humanos na Era Digital' (*Portal da Ordem dos Advogados*, 14 June 2022) <https://portal.oa.pt/comunicacao/imprensa/2022/06/14/a-inconstitucionalidade-da-carta-portuguesa-de-direitos-humanos-na-era-digital/> accessed in 30 April 2023.

[27] Domingos Soares Farinho (*note 2*) 92.

[28] Diane Rowland, '"Virtual world, real rights?": Human rights and the internet', in Marco Odello and Sofia Cavandolli (eds), *Emerging Areas of Human Rights in the 21st Century: the role of the Universal Declaration of Human Rights* (Routledge 2011) 8-9.

[29] Domingos Soares Farinho (*note 2*) 93.

[30] Article 8, §2 (Charter): *supra*, *note 22*.

[31] Article 13, §1 and §2 (Charter): 'Everyone has the right to obtain support from the State in exercising the right to erasure personal data concerning them, under the terms and conditions established in the European and national applicable legislation.' (§1); 'The right to oblivion can be exercised posthumously by any heir of the holder of the right, unless the latter has decided otherwise.' (§2).

(GDPR).[32] In this context, it is of utmost importance not to forget the principle of the 'precedence' (also referred to as 'primacy' or 'supremacy') of EU law, which stipulates that, where a conflict arises between an aspect of EU law and an aspect of national law, the first must prevail (*Flaminio Costa v ENEL* C-6/64, CJEU 15 July 1964).

Last but not least, the Charter is generally criticized for mixing, with no apparent criteria, 'substantive, organizational, promoting and simply programmatic norms with political objectives and empty declarations'.[33]

Despite everything that was said, we do believe that the Charter contains undeniable *value*, in *normative* and *symbolic* terms. As suggested above, it draws attention to the essential role that fundamental rights and freedoms must play as a basic framework for the creation and use, now and in the future, of digital instruments in the various social domains – digit(al)ization is, in fact, pointed out as a *total social phenomenon*[34] – and, namely, within the framework of justice, including criminal justice.

3 Article 9 of the Charter

As suggested above, Article 9 of the Charter specifically addresses the use of AI and robots. Its paragraph 1 establishes that 'the use of AI must be guided by the respect for fundamental rights, assuring a fair balance between the principles of explainability, security, transparency and responsibility, considering the circumstances of each concrete case and establishing processes that aim to avoid any prejudice or other forms of discrimination.' In turn, paragraph 2 states that 'the decisions made by algorithms with considerable impact on the recipients' realm must be communicated to stakeholders, both being appealable and auditable under law.' Last but not least, paragraph 3 relates to the creation and use of robots and determines the observation of the 'principles of beneficence and non-maleficence, respect for human autonomy and justice, as well as all the

[32] An opinion of the *National Commission on Data Protection* (Opinion/2020/116) – which pronounces on the content of the Draft Bill No. 473/XIV/1, at the request of the Commission on Constitutional Affairs, Rights, Freedoms and Guarantees of the Assembly of the Republic – refers very emphatically to a risk of non-compliance of the provisions of the legislative project with the EU Law (II.1). It should be noted that, following this opinion, several rules of the Draft Bill were adjusted (Article 4 of the Draft Bill, now Article 5 – Guarantee of access and use) or eliminated (Article 7, §2 and §4, of the Draft Bill, now Article 8 – Right to privacy in a digital environment).

[33] José de Melo Alexandrino, 'Dez breves apontamentos sobre a Carta Portuguesa de Direitos Humanos na Era Digital' (*Comissão da Carteira Profissional de Jornalista*, 2 June 2021) <https://www.ccpj.pt/pt/informacao/contributo-do-professor-jose-melo-alexandrino-para-analise-da-carta-portuguesa-de-direitos-humanos-na-era-digital/> accessed 30 April 2023.

[34] Antoine Garapon and Jean Lassègue, *Justice Digitale* (puf 2018) 83, referring to Marcel Mauss, *Sociologie et anthropologie*, (puf 1973) 274, which defines it as a phenomenon that 'sets in motion the whole of society and its institutions'.

principles and values enshrined in article 2 of Treaty on European Union, namely non-discrimination and tolerance.'

The approach adopted to analyse the different norms of this legal precept will focus on justice in general and, where appropriate, criminal justice in particular.

3.1 The use of AI must be guided by the respect for fundamental rights

The idea that the use of AI must be guided by an imperative of respect for fundamental rights and freedoms is not new. In fact, it derives from the so-called 'human rights-based approach to science':

> A human rights-based approach recognises that science is a socially organised, human activity, which is value-laden and shaped by organisational structures and procedures. It asks how governments and other stakeholders can create and implement policies to ensure safety, health and livelihoods; to include people's needs and priorities in development and environmental strategies; and to ensure they participate in decision-making that affects their lives and resources.[35]

In consequence, the values and interests of the *human person* must always take precedence over those of *science* and *society*, thus ensuring that *human dignity* asserts itself as *ground* and *bound* for any scientific, technical and technological advances.[36] It cannot be different in the field of AI. In fact, human and fundamental rights should be strengthened by AI, not undermined.[37]

Of course, one might affirm that this statement sounds quite declamatory. That may well be true. Nevertheless, we still believe that it has a huge significance in the various fields of justice, alerting to the reinforced need of respecting and protecting certain fundamental rights and freedoms (such as the right of access to a court or the right to a fair trial, namely with regard to the right to be heard and to equality of arms) in this specific context.[38] Moreover, it does challenge public and private actors to implement measures that are able to reduce the impact of AI systems on those rights and freedoms, not only in

---

[35] S. Romi Mukherjee, 'Linking Science and Human Rights: Facts and Figures' (*SciDev.Net*, 18 September 2012) <https://www.scidev.net/global/features/linking-science-and-human-rights-facts-and-figures/> accessed 1 May 2023.

[36] Eduardo A. S. Figueiredo, Direito e nanobiotecnociência: reflexões na encruzilhada da inovação, do risco e da crise do(s) direito(s) (Almedina 2021) 169.

[37] Commissioner for Human Rights, *Unboxing Artificial Intelligence: 10 steps to protect Human Rights* (Council of Europe, May 2019) 6 <https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64> accessed 1 May 2023.

[38] Anabela Miranda Rodrigues, 'Inteligência Artificial no Direito Penal – a Justiça Preditiva entre a Americanização e a Europeização', in Anabela Miranda Rodrigues (ed), *A Inteligência Artificial no Direito Penal* (Almedina 2020) 33-34.

substantial terms but also in organizational and procedural terms (*v.g.*, by providing information and ensuring transparency, by granting independent and effective oversight, *etc.*).[39]

3.2 The call for a fair balance between the principles of explainability, security, transparency and responsibility

According to the *principle of security*, an AI system must be robust, secure and safe throughout its entire lifecycle, functioning appropriately and not posing unreasonable risks to its users.[40] Therefore, the observance of this principle, *v.g.* in the field of justice, depends not only on the reliability of the sources and integrity of the data that will be used to 'feed' the system but also on the traceability of datasets, processes and decisions, thus ensuring that it will operate in a controllable environment.

In fact, the abovementioned principle is closely linked to the *principle of transparency*, which demands accessibility, readability (comprehensibility) and external control of the algorithmic models used in data processing. And note, the idea of transparency of the AI systems – which is mainly *technical* – can only be realised if the software is able to justify its outcomes and results, in accessible and clear language (*principle of explainability*). There are, however, theoretical and practical barriers to the development of algorithmic systems that are truly transparent and explainable, which may boycott the effectiveness of a *subjective right to explanation,* as we have already stated above. [41] Also, one must be attentive to the fact that transparency is not, in itself, a sufficient solution to the huge imbalances generated by the use of IA in the field of justice: in order to safeguard the interests of those affected by an algorithmic judicial decision (for example, in a criminal proceeding, the defendants and the community as a whole), as opposed to the interests related to the administration of justice, it is necessary that the understanding of the algorithm and its operating mode does not remain a matter restricted to experts, also including the true addressees of the decision. In this sense, *algorithmic literacy* is also a priority.

Finally, the *principle of responsibility* addresses not only *individuals* and *private actors* – namely *companies* –, committing them to develop AI systems according to a *human-rights-by-design orientation;* but also *public bodies*, instructing them to create independent authorities, with functions of certification and periodic audit of AI systems.[42]

---

[39] Commissioner for Human Rights (*note 37*) 9-10.

[40] OECD, *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449, 22 May 2019) 1.4. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449#mainText> accessed 1 May 2023.

[41] In the same vein, see José Joaquim Gomes Canotilho (*note 12*) 69.

[42] Serena Quatroccolo, 'Intelligenza Artificiale e Giustizia: Nella Cornice Della Carta Etica Europea, Gli Spunti Per Un' Urgente Discussione Tra Scienze Penali E Informatiche' [2018] *La legislazione penale* 6-9;

3.3 The aim to avoid any prejudice or other forms of discrimination

Being know that AI systems affect people everywhere and are subject to a *risk of (explicit or implicit) bias*, the juridical canon of *non-discrimination* is expressly mentioned. This means public and private entities must guarantee that AI systems do not create, reproduce or aggravate unjustified discrimination, nor do they lead to deterministic evaluations. In addition, when a situation of algorithmic discrimination has been identified, this principle calls for the adoption of measures to limit or, where possible, to correct it, and even to raise awareness among the parties involved about the fact that they are being discriminated.[43]

Also, note that the risk of discrimination is higher when sensitive data are at stake, such as those revealing racial or ethnic origin, socio-economic background, political opinions, religious or philosophical beliefs, and trade-union membership. Even genetic data, biometric data, health-related data and data concerning a person's sex life or sexual orientation fall within this category. In this context, special measures to prevent, detect and fight algorithmic discrimination must be in force.

3.4 The decisions made by algorithms with considerable impact on the recipients' realm must be communicated to stakeholders, both being appealable and auditable under law

Paragraph 2 of Article 9 of the Charter consecrates the so-called 'under user control' *principle*, which aims to ensure that users of AI systems act as plainly informed subjects and have total control over the decisions they make.

If one assumes the user can either be the person using the algorithmic tool (a *judge*, for instance) or the recipient of the decision made by the algorithm (a *person under suspicion of the court*, for instance), this principle translates, for the first one, into the possibilities of re-examining the data used to produce the output and of not being necessarily bound by it, taking into consideration the particularities of the case *sub judice*; for the latter, into the right to be informed in advance of the use of the tool and the right of access to a court, in the conditions foreseen by Article 20 of the Portuguese Constitution and Article 6 of the European Convention on Human Rights (ECHR).[44]

Mitja Gialuz, 'Quando la Giustizia Penale encontra l'Intelligenza Artificiale: Luci e Ombre dei *Risk Assessment Tools* tra Stati Uniti ed Europa' [2019] *Diritto Penale Contemporaneo* 13; and also, Anabela Miranda Rodrigues (*note 38*) 34-36.

[43] Serena Quatroccolo (*note 42*) 5; Mitja Gialuz (*note 42*) 12-13; and also, Anabela Miranda Rodrigues (*note 38*) 34.

[44] Serena Quatroccolo (*note 42*) 9; Mitja Gialuz (*note 42*) 13; and also, Anabela Miranda Rodrigues (*note 38*) 36.

The implications of this principle in the field of criminal justice are multiple. In any case, the main idea to be retained is that AI systems should increase users' autonomy in any decision-making process and not reduce or even neutralize it.[45] Note, however, that, unlike Article 22 of the GDPR[46], Portuguese law[47] does not recognize any kind of *legitimising efficacy* to the data subject's explicit consent, namely in the context of a decision based solely on automated processing for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties.

3.5 The creation and use of robots

Paragraph 3 of Article 9 states that the creation and use of robots must respect four principles of *biolaw*[48] (which were traditionally developed in the field of *bioethics*):[49] beneficence, non-maleficence, autonomy and justice.

---

[45] To read more about automated individual decision-making, see A. Barreto Menezes Cordeiro, *Direito da Proteção de Dados à luz do RGPD e da Lei n.º 58/2019* (Almedina 2020) 148 f.; and also, Mafalda Miranda Barbosa, 'Dos *expert systems* aos *data systems AI*: impacto ao nível da proteção de dados' [2021] 45 *Julgar* 21 f.

[46] According to Article 22, §1: 'the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.' However, §2 states that §1 shall not apply if the decision *(b) is authorized by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject´s rights and freedoms and legitimate interests*; or (c) *is based on the data subject's explicit consent*. According to §3, in the cases referred to in point (c) of §2, *the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision*. Finally, according to §4, *decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point a)* [explicit consent of the data subject] *or (g)* [processing necessary for reasons of substantial public interest] *of Article 9(2) applies and suitable measures to safeguard the data subject´s rights and freedoms legitimate interests are in place.* See also, recital 71 of the GDPR.

[47] See Law No. 59/2019, of August 8, which transposed Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data. According to Article 11 of the diploma: *a decision based solely on automated processing, including profiling, which produces an adverse legal effect concerning the data subject or significantly affects him or her, is prohibited unless authorized by law, as long as there is a right of the data subject to obtain human intervention on the part of the controller* (§1); and *decisions to which paragraph 1 refers shall not be based on special categories of personal data mentioned in Article 6* (§2).

[48] Peter Kemp refers to the principles of autonomy, dignity, integrity (embracing both the ideas of beneficence and non-maleficence) and vulnerability (in a certain sense, complementing integrity and referring to justice). See Peter Kemp, 'The Idea of European Biolaw: Basic Principles', in Erick Valdés and Juan Alberto Lecaros (eds), *Biolaw and Policy in the Twenty-First Century: Building Answers for New Questions* (Springer 2019) 29-31; and also, Eduardo A. S. Figueiredo (*note 36*) 131.

[49] Tom L. Beauchamp and James F. Childress, *Principles of biomedical ethics* (4th ed, Oxford University Press 1994) 120 f.

i) The *principle of beneficence* ('do good') demands the creation and use of robots to pursue the exclusive aim of benefiting humanity, the human person and/or the environment (it's the so-called "beneficial AI")[50]. Such benefits must be defined by democratic means. Moreover, it requires the prevention of harm and the removal of harm-causing conditions that might exist – which means that this principle must be considered in conjunction with the *principle of non-maleficence*;

ii) The *principle of non-maleficence* ('do no harm') recognises the existence of considerable risks related to the creation and use of robots, calling for rigorous processes of *risk analysis* (involving *risk assessment*, *risk management* and *risk communication*). Also, this principle demands that human dignity and the fundamental rights and freedoms of individuals are fully respected and protected from any kind of threat. When damages are caused, every person must be able to rely on the rule of law, access to justice, the right to redress and the right to a fair trial;

iii) The *principle of autonomy* underlines the fact that the creation and use of robots must never impair human freedom. On the contrary, autonomous machines must capacitate and empower individuals to take the lead in any decision-making process that concerns or affects them. In practice, this means that the person must be informed in advance of the fact that a robot will be used and, eventually, have a word to say in relation to that fact – consider, for example, the right to refuse care from a robot.[51] Also, robots must be controllable, being difficult to accept that, in any circumstance, the machine might impose certain behaviour upon, or restrict, a person.[52] Last but not least, this principle defies us to develop mechanisms that are able to adequately and effectively solve the so-called 'responsibility or liability gap';[53]

iv) The principle of *justice* refers to questions as diverse as: (1) the need to grant equitable access to robots by all members of society (refusing any

[50] Luciano Floridi and Josh Cowls, 'A Unified Framework of Five Principles for AI in Society' [2019] 1/1 *Harvard Data Science Review* 5; and also, Luciano Floridi, Josh Cowls, Thomas C. King and Mariarosaria Taddeo, 'How to Design AI for Social Good: Seven Essential Factors' [2020] 26 *Science and Engineering Ethics* 1771.

[51] Nathalie Nevejans, *European Civil Law Rules in Robotics* (PE 571.379, October 2016) 21 <https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf> accessed 1 May 2023.

[52] Nathalie Nevejans (*note 51*) 21.

[53] In general, see Filippo Santoni de Sio and Giulio Mecacci, 'Four Responsibility Gaps with Artificial Intelligence: Why They Matter and How to Address Them' [2021] 34 *Philosophy & Technology* 1057.

kind of 'robotics divide');[54] (2) the need to prevent and fight any kind of discriminatory behaviour by robots; (3) the need to grant that the creation and use of robots is led by a general imperative of *sustainability*; (4) the need to ensure the sharing of benefits related to or derived from the creation and use of robots.

Apart from these principles, §3 of Article 9 also mentions the need to respect the 'principles and values enshrined in article 2 of the Treaty on European Union': human dignity, freedom, democracy, equality, the rule of law, respect for human rights, pluralism, non-discrimination, tolerance, justice and solidarity. In fact, this reference causes us some perplexity: on the one hand, there is no plausible reason to justify why the national legislator decided to expressly refer to EU law in this context, when Articles 1 and 2 of the Portuguese Constitution also mention the values in question; on the other hand, this reference seems totally unnecessary, considering that those values are already presupposed by the four biolaw principles mentioned in the first segment of the norm. And note: no one can argue that the first segment of the norm refers only to *bioethical principles* and the second to *juridical 'principles and values'*. The Charter is a legal document – therefore, all of its references are made (or, at least, should have been made…) in the realm of Law.

4 Concluding Remarks

Digit(al)ization has confronted us with a truly paradoxical scenario, with technological advancements showing a Janus-faced attitude in relation to fundamental rights and freedoms.[55] On the one hand, new technologies increase our chances of making effective several rights and freedoms (for example, the internet offers us new possibilities to communicate, access information, organise a meeting or a protest, profess a religion, *etc.*); on the other hand, they also increase the risk of those same rights and freedoms being violated.

In this context, the Charter – with its 'proto-constitutional discourse' [56] and clear connections to the global movement of *digital constitutionalism* – contributes to nourish the debate on how fundamental rights and freedoms must be thought of and elaborated in light of the mutated digital society – not only in *substantal*, but also in *organizational and procedural* terms. In a certain way, it also reaffirms that, online just as offline, law continues to

[54] Nathalie Nevejans (*note 51*) 24.
[55] Edoardo Celeste (*note 9*) 17.
[56] Edoardo Celeste (*note 9*) 46.

be society's most important medium to ensure *order*, *rule*, and *justice*, as well as to protect the *human person* and her *intrinsic dignity*.[57]

Thus, in spite of the juridical problems raised by the Charter (either due to redundancies, risks of generating normative conflicts, or even inconsistencies in the way it is structured), we do believe this legal document to be a clear sign of the huge efforts made by the Portuguese State to participate in 'the worldwide process of transformation of the internet into an instrument for the achievement of freedom, equality and social justice, as well as a space for the promotion, protection and free exercise of human rights, granting social inclusion in a digital environment' (Article 2, §1, of the Charter). Moreover, it will serve as a legal framework *par excellence* to guide, justify and limit public and private actions or omissions in the context of the digital age, also contributing to preventing undesirable phenomena of judicial activism, with courts being called upon to intervene in a context of 'splendid isolation' to protect fundamental rights and freedoms, namely due to the legislator's inertia.[58]

Lastly, in the specific field of AI and robots, the Charter contributed to the identification and consolidation of the several basic principles and values that must be observed, namely in order to ensure that our destiny continues to have a 'human face' and don't end up lost in the *metaverse*…

Selected literature

Alexandrino J M, 'Dez breves apontamentos sobre a Carta Portuguesa de Direitos Humanos na Era Digital' (*Comissão da Carteira Profissional de Jornalista*, 2 June 2021) <https://www.ccpj.pt/pt/informacao/contributo-do-professor-jose-melo-alexandrino-para-analise-da-carta-portuguesa-de-direitos-humanos-na-era-digital/> accessed 30 April 2023

Barbosa M M, 'Dos *expert systems* aos *data systems AI*: impacto ao nível da proteção de dados' [2021] 45 *Julgar* 13-33

Beauchamp T L. and Childress J F., *Principles of biomedical ethics* (4th ed, Oxford University Press 1994)

[57] Matthias C. Kettemann, The Normative Order of the Internet: a Theory of Rule and Regulation Online (Oxford University Press 2020) 305.

[58] Oreste Pollicino, *Judicial Protection of Fundamental Rights on the Internet: a Road Towards Digital Constitutionalism?* (Hart Publishing 2021) 207-208. Note the Author considers that 'the amplification of the judicial momentum does not seem to be destined to fall, but to expand its boundaries for interpreting and enforcing fundamental rights in the algorithmic society' (211).

Canotilho J J G and Moreira V, *Constituição da República Portuguesa Anotada – Volume I* (4th edn, Coimbra Editora 2014)

Canotilho J J G, 'Sobre a indispensabilidade de uma Carta de Direitos Fundamentais Digitais na União Europeia/About the indispensability of a Charter of Fundamental Rights to the European Union' [2019] 31/1 *Revista do Tribunal Regional Federal 1ª Região Brasília DF* 69-75

Celeste E, Digital Constitutionalism: the Role of Internet Bill of Rights (Routledge 2023)

Commissioner for Human Rights, *Unboxing Artificial Intelligence: 10 steps to protect Human Rights* (Council of Europe, May 2019) <https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64> accessed 1 May 2023

Cordeiro A B M, Direito da Proteção de Dados à luz do RGPD e da Lei n.º 58/2019 (Almedina 2020)

De Gregorio G, Digital Constitutionalism in Europe: Reframing Rights and Powers in the Algorithmic Society (Cambridge University Press 2022)

Farinho D S, 'The Portuguese Charter of Human Rights in the Digital Age: a legal appraisal' [2021] 13 *Revista Española de la Transparencia* 85-105

Figueiredo E, Direito e nanobiotecnociência: reflexões na encruzilhada da inovação, do risco e da crise do(s) direito(s) (Almedina 2021)

Floridi L and Cowls J, 'A Unified Framework of Five Principles for AI in Society' [2019] 1/1 *Harvard Data Science Review* 1-14

Floridi L, Cowls J, King T C and Taddeo M, 'How to Design AI for Social Good: Seven Essential Factors' [2020] 26 *Science and Engineering Ethics* 1771-1796

Garapon A and Lassègue J, *Justice Digitale* (puf 2018)

Gialuz M, 'Quando la Giustizia Penale encontra l'Intelligenza Artificiale: Luci e Ombre dei *Risk Assessment Tools* tra Stati Uniti ed Europa' [2019] *Diritto Penale Contemporaneo* 1-23

Hutler B, Rieder T N, Mathews Debra J H, Handelman D A and Greenberg A M, 'Designing robots that do no harm: understanding the challenges of Ethics *for* Robots' [2023] *AI Ethics*

Kemp P, 'The Idea of European Biolaw: Basic Principles', in Erick Valdés and Juan Alberto Lecaros (eds), *Biolaw and Policy in the Twenty-First Century: Building Answers for New Questions* (Springer 2019) 19-32

Kettemann M C, The Normative Order of the Internet: a Theory of Rule and Regulation Online (Oxford University Press 2020)

Leitão L M, 'A inconstitucionalidade da Carta Portuguesa de Direito Humanos na Era Digital' (*Portal da Ordem dos Advogados*, 14 June 2022) <https://portal.oa.pt/comunicacao/imprensa/2022/06/14/a-inconstitucionalidade-da-carta-portuguesa-de-direitos-humanos-na-era-digital/> accessed in 30 April 2023

Mukherjee S. R, 'Linking Science and Human Rights: Facts and Figures' (*SciDev.Net*, 18 September 2012) <https://www.scidev.net/global/features/linking-science-and-human-rights-facts-and-figures/> accessed 1 May 2023

Nevejans N, *European Civil Law Rules in Robotics* (PE 571.379, October 2016) <https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf> accessed 1 May 2023
OECD, *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449, 22 May 2019) <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449#mainText> accessed 1 May 2023

Pernice I, 'Multilevel Constitutionalism and the Crisis of Democracy in Europe' [2015] 11/3 *European Constitutional Law Review* 541-562

Pollicino O, Judicial Protection of Fundamental Rights on the Internet: A Road Towards Digital Constitutionalism? (Hart Publishing 2021)

Quatroccolo S, 'Intelligenza Artificiale e Giustizia: Nella Cornice Della Carta Etica Europea, Gli Spunti Per Un' Urgente Discussione Tra Scienze Penali E Informatiche' [2018] *La legislazione penale* 1-12

Rodrigues A M, 'Inteligência Artificial no Direito Penal – a Justiça Preditiva entre a Americanização e a Europeização', in Anabela Miranda Rodrigues (ed), *A Inteligência Artificial no Direito Penal* (Almedina 2020) 11-58

Rowland D, '"Virtual world, real rights?": Human rights and the internet', in Marco Odello and Sofia Cavandolli (eds), *Emerging Areas of Human Rights in the 21st Century: the role of the Universal Declaration of Human Rights* (Routledge 2011) 7-23

Schuilenburg M and Peeters R (eds), The Algorithmic Society: Technology, Power, and Knowledge (Routledge 2021)

Santoni de Sio F and Mecacci G, 'Four Responsibility Gaps with Artificial Intelligence: Why They Matter and How to Address Them' [2021] 34 *Philosophy & Technology* 1057-1084

Suzor N P., Lawless: The Secret Rules That Govern Our Digital Lives (Cambridge University Press 2019)

Wischmeyer T, 'Artificial Intelligence and Transparency: Opening the Black Box', in Thomas Wischmeyer and Timo Rademacher (eds), *Regulating Artificial Intelligence* (Springer 2020) 75-101

CROSS-BORDER ADMISSIBILITY OF AI-EVIDENCE

By Sabine Gless[*]

Abstract

AI is a universally relevant research topic as well as a technology permeating our daily lives. Smart devices monitor our doings - for safety and for other reasons. A modern car, furnished with AI-devices, which oversees drivers' actions, might know more about the drivers' habits and activities than their family, friends and neighbors.[1] *If read out, we could learn many things from these devices and the IT-tools used to read them out are scientific-based and ubiquitously employed. Therefore, AI-generated evidence should be easily transferable across borders and encounter minimal hurdles regarding admissibility in different jurisdictions. This report tests this hypothesis by analyzing the cross-border admissibility of AI-generated evidence. The analysis begins by explaining the terminology and emphasizing the need for cross-border cooperation to facilitate the exchange of evidence. Subsequently, it explores the increasing role of AI in the evidentiary process, particularly within forensic contexts. The emergence of AI-driven evidence presents both transformative potential and possible pitfalls, as we have observed with other science-based evidence in the past. The report introduces real-world examples of AI-based evidence, such as DNA sample testing, consumer product alerts, and facial recognition systems. It then delves into the specific challenges related to the cross-border admissibility of AI evidence. Despite the absence of specific regulations, the report finds that domestic jurisdictions possess tools to address the two main problems: reliability of evidence proffered in criminal trials and fair trial safeguards. In conclusion, the report advocates for the desirability of a 'universal code' to govern the admissibility of AI evidence.*

1 Science – a true universal evidentiary language?

In the movie 'Contact', Dr. Eleanor Arroway (a fictional astronomer) states that '[m]athematics is the only true universal language'. This line embodies a major theme of the film, as the plot focusses on the discovery of extraterrestrial life and the subsequent mission to establish first contact. Unsurprisingly, however, mathematics does not offer the easy universal communication that Dr. Arroway seeks and, on her mission, she is instead left struggling for an authentic understanding of truth and evidence, outweighing differences in space and time.

[*] Prof. Dr. iur. Sabine Gless, Chair for Criminal Law and Criminal Procedure Law, Faculty of Law, University of Basel, sabine.gless@unibas.ch. The author wishes to express her gratitude to MLaw Janneke de Snaijer for her magnificent assistance in finalizing the manuscript and the authors of the respective country reports as well as the General Rapporteur Juliette Lelieur for sharing information. Without this support, the special report could not have been submitted.

[1] Gless S, Lederer F and Weigend T, 'AI-based Evidence in Criminal trials' 2023/2024 TULSA LAW REVIEW forthcoming.

In some ways, Dr. Arroway's efforts remind us of the legal scholars who want to unite positions on truth seeking, determined by legal traditions and cultural differences, to find a universal code for evidence in criminal trials. The advancement of forensic science, based on scientific approaches shared across borders, has nurtured the hope for a global standard on how to obtain reliable evidence, process it and assign an evidentiary weight. However, as this report will highlight, we have a long – not yet clearly defined – road ahead to arrive there.

At the conclusion of 'Contact', Dr. Arroway finds herself occupying a position she opposed at the beginning of her story, believing, with absolute certainty, something which she does not have the ability to prove to others with a mathematical degree of accuracy attached to the result. Fact-finders in criminal cases often find themselves in the same position.[2] Law overcomes this issue by accepting judicial belief, as a sort of proxy for societal acceptance of truth, as a set of facts established by compliance to procedural rules and trust in judges.[3]

Yet, the idea to evade the evidentiary problem with a mathematic-based approach exists,[4] and must be considered, particularly, where forensic – i.e. science-based – evidence is involved.[5] The possible employment of AI in the evidentiary process might revivify this debate, as it is based on mathematical concepts and computer science and offers new opportunities to conceptualize uncertainties attached to different modes of proving facts. Based on the scientific approach taken,[6] AI-employment could open up new ways of sharing evidence across borders without a loss of information in the cooperation of different jurisdictions. For instance if the cross-border transmission is based on a certified procedure that generates an 'authenticated object of perception' (*authentifiziertes Wahrnehmungsobjekt)*.[7]

With the differing procedural rules for the obtainment and assessment of evidence in each domestic system, there is an ongoing problem that information obtained under the rules of one legal system cannot demand universal applicability when presented as evidence in another different legal system. This is even true in cases where forensic experts

[2] Legal scholarship even argues that a decision based on significant (and thus very high) numbers (numbering in billions or trillions) are beyond human capacity of information processing and ingenuity, cf. Gill P, Benschop C, Buckleton J, Bleka Ø and Taylor D, 'A Review of Probabilistic Genotyping Systems: EuroForMix, DNAStatistX and STRmixTM' (2021) 12 Genes 1559, 1587.

[3] Cf. Gless S, 'Could Robot Judges Believe? Epistemic Ambitions of the Criminal Trial as we Approach the Digital Age. A Comment on Sarah Summers "Epistemic Ambitions of the Criminal Trial: Truth, Proof, and Rights" (2023) 5 International Journal on Evidential Legal Reasoning 1-11.

[4] Hoyer A, 'Der Konflikt zwischen richterlicher Beweiswürdigungsfreiheit und dem Prinzip „in dubio pro reo"' (1993) 105 ZStW 523.

[5] Vuille J and Taroni F, 'Measuring Uncertainty in Forensic Science' (2021) 24 IEEE Instrumentation & Measurement Magazine 5.

[6] De Finetti B, 'Bayesianism: Its Unifying Role for Both the Foundations and Applications of Statistics' (1974) 42 International Statistical Review 117, 121.

[7] For more details on the concept of an 'object of perception' as a primary source for fact-finding see Gless S, *Beweisrechtsgrundsätze einer grenzüberschreitenden Strafverfolgung* (Vol. 2, Nomos 2006) 34-8.

use natural sciences (e.g. biology, chemistry and physics) to assist a court in establishing a fact-based examination and analysis of evidential material to report to the court. The question is whether and how the employment of AI in the production of evidence can overcome issues of admissibility or evidentiary value in domestic criminal proceedings caused by a transfer of evidence across jurisdictions.

This report will look at the possible uses of AI to generate or assist in the assessment of evidence for cross-border use in criminal trials and, thus, how it could be employed to overcome transnational barriers when using a certain piece of information as evidence.[8] The examples used in this report are (a) probabilistic genotyping (as it is done e.g. by STRmix in Canada[9] and in the U.S.), and (b) alerts generated by consumer products (like drowsiness detection systems used in modern cars to identify indicators of unfitness in human drivers). Furthermore, the report highlights issues concerning (c) facial recognition systems, including AI-driven enhancements of faces and (d) broader forensic tools ('digital forensics as a service') provided by e.g. the Hansken software in the Netherlands or Palantir software that has been used in Germany (before a judgement by the German Constitutional Court banned it in February 2023).[10]

2 Background and terminology

This report combines two perspectives: that of transnational cooperation among different jurisdictions, with the aim of sharing evidence for proof in a criminal trial (2.1) and that of AI employment based on a computer science assisted evidentiary process, especially with regards to forensic evidence (2.2).

2.1 Transnational cooperation with the aim of evidence sharing

Transnational cooperation between states, as well as between states and International Agencies, in the area of law enforcement has seen a considerable rise in recent decades at many different levels, including the exchange of AI-evidence. Two prominent examples being the Prüm Treaty[11] (that allows for the exchange of DNA data, fingerprints and traffic data, and might allow for exchange of the results of facial recognition[12]) and Eurodac[13] (that allows for the exchange of fingerprints of asylum seekers and 'stateless people').

---

[8] For a practical example of how generating and assessing evidence merge see the stories provided on STRMix' homepage, e.g. 'STRmix™ Will Interpret DNA Evidence for the St. Louis County Police Department' (posted on 25 January 2023, 9:00 am) <https://strmix.com/news/strmix-will-interpret-dna-evidence-for-the-st-louis-county-police-department/> accessed 28 August 2023.

[9] See the report on Canada, <https://www.penal.org/fr/2023>, A-03, accessed 30 November 2023, p. 86.

[10] Germany, Bundesverfassungsgericht BVerfG, Urteil des Ersten Senats vom 16. Februar 2023, 1 BvR 1547/19, 1 BvR 2634/20.

[11] Council Decision (EC) 2008/615/JHA on the stepping up of cross-border cooperation, particularly in combating terrorism and cross-border crime (2008) OJ L 210/1.

[12] See e.g. <https://www.telefi-project.eu/> accessed 28 August 2023.

[13] Council Regulation (EC) 2725/2000 concerning the establishment of 'Eurodac' for the comparison of fingerprints for the effective application of the Dublin Convention (2000) OJ L 316; Council Regulation

The new methods for transferring evidence gathered abroad are not matched by changes on the domestic level. Questions regarding the use of such evidence, with regard to reliability and fairness of fact finding, remain a purely domestic affair: The competent authorities of the jurisdiction in charge of deciding on the merits of a case must also decide on the admissibility of a certain piece of evidence.[14] The underpinning rationale lies in the differences of requirements that guarantee reliability and fairness in each jurisdiction, which are core requirements for the acceptance of a judgment.[15] Does such 'legal connotation'[16] of evidence also exist for forensic evidence? Or does its grounding in scientific analysis (like the biological method used in DNA sample testing) warrant a different handling?

It is obvious that investigatory findings gleaned within the scope of transnational cooperation are only of practical value to national law enforcement agencies if they can be used as evidence in the relevant national criminal proceedings. Thus, if evidence is obtained under the jurisdiction of a foreign legal system, its evidentiary value can be called into question because, in such situations, the national rules covering investigative procedures are not usually observed in the foreign country. A mathematical (or in other way) allegedly 'objectively substantiated' or 'authenticated' mode to certify evidence across borders could help to build trust and thus ease the cross-border use of evidence.

Such an approach may help us to avoid 'legal misunderstandings' or 'cultural noise' and distillate the 'authenticated object of perception'[17] mentioned above. While authentication detached from a normative framework of a jurisdiction may have seemed out of reach a few decades ago, today one could imagine evidence generated autonomously by a transnationally certified AI-tool. Such an option would allow for a process of fact-finding deemed reliable by not only one, but many legal communities. This could possibly be achieved with the AI-based evidence in the forms that we used as examples in this report: (a) probabilistic genotyping and (b) alerts generated by consumer products. The aim would be a true internationalization of AI-based evidence that uses scientific accreditation for the universal acceptance of its reliability.

2.2 AI in the evidentiary process (forensic evidence)

The employment of AI for the obtainment and, particularly for the production of evidence is not (yet) mainstream practice. Evidentiary proceedings are tailored for humans.

(EC) 407/2002 laying down certain rules to implement Regulation 2725/2000 concerning the establishment of 'Eurodac' for the comparison of fingerprints for the effective application of the Dublin Convention (2000) OJ L 62.

[14] Gless S, Internationales Strafrecht: Grundriss für Studium und Praxis (3rd edn, Helbing Lichtenhahn 2021) no. 267.

[15] Jackson J and Summers S, The Internalisation of Criminal Evidence: Beyond the Common Law and Civil Law Traditions (CUP 2012) 69-70.

[16] Gless S, 'Grenzüberschreitende Beweissammlung' (2013) 125 ZStW 573.

[17] See supra Fn. 7.

Yet, AI is increasingly used for probabilistic genotyping, in facial recognition systems and for extracting relevant information out of a pool of data.

2.2.1 *Chances and pitfalls*

What does the employment of AI for obtaining and producing evidence mean for domestic evidentiary proceedings?

Lawyers value science-based evidence for its great potential to provide new pools of information and thus a chance for more accurate fact-finding in criminal trials. Yet, they are aware of possible pitfalls. One basic question is the informative value of forensic evidence for fact-finding in a criminal trial. We already face this problem when deciding, for instance, the significance of a DNA sample from a crime scene for the question of guilt. A crucial question is how to handle the manifold sources of error and to assign a proper evidentiary value on a match, taking into account inherent methodological error rates etc.[18] These questions have been at the center of a controversial debate that, in a certain way, marks what will be called the 'evidentiary cycle' in this report (see infra 2.2.2).

AI promises an innovative method for matching more traces and information, for instance with forensic audio analysis[19] or with sampling movement patterns,[20] whilst offering a (high) probability value. However, as with all new methods, the issue arises of how to handle new sources of error and the assigning the adequate evidentiary value on a match, whilst considering the inherent methodological error rates. Furthermore, it is important to note that, for instance, judges struggle to put DNA-sample tests to a decisive test before accepting matches and this is likely to be exacerbated if a DNA-sample test is based on AI (and thus matches are not entirely traceable and explainable).

Before turning to the finer details, it is important to point out that fact-finding always needs a holistic check, usually provided by human oversight and common sense, before acceptance. The following, somewhat obscure, example shows how blind trust in certain scientific methods can mislead terribly. The 'Phantom of Heilbronn' (also referred to as the 'Woman Without a Face') had been the object of a police hunt for an unknown female serial killer whose existence was inferred from DNA evidence found at numerous crime scenes in Austria, France and Germany from 1993 to 2009. The only connection between the crimes was the presence of a DNA sample from a single female, which had been recovered from approx. 40 crime scenes (ranging from burglaries to murders). Eventually, investigators concluded that there was no 'phantom criminal', but that the DNA had

[18] Vuille J and Taroni F, 'Measuring Uncertainty in Forensic Science' (2021) 24 IEEE Instrumentation & Measurement Magazine 5.
[19] Raponi S, Oligeri G and Ali IM, 'Sound of guns: digital forensics of gun audio samples meets artificial intelligence' (2022) 81.21 Multimedia tools and applications 30387.
[20] Becker S, Heuschkel M, Richter S and Labudde D, 'COMBI: Artificial Intelligence for Computer-Based Forensic Analysis of Persons' (2022) 36 KI-Künstliche Intelligenz 171.

already been present on the cotton swabs used for collecting DNA samples; it belonged to a worker at the factory where they were produced.[21]

This rather bizarre case points to the more widespread issue of humans ignoring the possibility of mistakes when 'scientific methods' are involved; a similar attitude might be expected regarding the use of AI. The Australian case of Farah Jama illustrates this. He had been convicted of rape in Australia in 2008, based on a sample collected from the genital area of a woman who had been found unconscious in the restroom of a nightclub and could not remember anything but was informed that she had been raped. DNA was found in the sample that matched Jama. He was acquitted on appeal (in 2009), after a review board determined that the incriminating DNA was the result of the contamination of the test kit used to sample the traces found on the victim.[22] In the everyday practice of criminal courts there are seemingly less severe sources of error, like not providing adequate information about the methodology or including exact error rates, which possibly cause more damage overall.[23]

AI based forensic evidence promises improved accuracy, but may also hold special difficulties. Take, for instance, the inherent black box problem, i.e. an inability to explain a certain result due to the opaque nature of the machine learning technology and its processes.[24] Triers of fact will have to decide whether to trust an AI-generated statement that can only partially be explained by experts. This is particularly the case, if the evidence is generated by AI in consumer products, like driving assistants, and not by certified forensic evidentiary tools (like systems made for probabilistic genotyping).

2.2.2 *Evidentiary cycle*

It seems with the emergence of every new type of forensic technology, that courts face the question of whether the means of registration or documentation is reliable, accurate, and objective, creating what has been called an 'evidentiary life cycle'.[25] These cycles are of interest because, among other things, they show that science-based evidence does not provide a fixed standard that supplies a timeless and ubiquitous code for the assessment

[21] <https://de.wikipedia.org/wiki/Heilbronner_Phantom> accessed 28 August 2023.
[22] See Vincent FHR, 'Inquiry into the circumstances that led to the conviction of Mr Farah Abdulkadir Jama' (Victoria Department of Justice 2010) <http://netk.net.au/DNA/Jama.pdf> accessed 28 August 2023.
[23] Vuille J and Taroni F, 'Measuring Uncertainty in Forensic Science' (2021) 24 IEEE Instrumentation & Measurement Magazine 5; Biedermann A and Vuille J, 'Bewertung von DNA-Untersuchungsergebnissen aus der Sicht von Gerichten und Sachverständigen: Wie viel von unserer Wahrnehmung können wir "für wahr nehmen"?' (2011) 129 ZStrR 278.
[24] Gless S, 'AI in the Courtroom: A Comparative Analysis of Machine Evidence in Criminal Trials' (2020) 51 GJIL 195, 211.
[25] Gless S, 'AI in the Courtroom: A Comparative Analysis of Machine Evidence in Criminal Trials' (2020) 51 GJIL 195, 215.

of evidence. The rise and partial fall of DNA sample testing documents this phenomenon.[26] Furthermore, the advancement of science can lead to pitfalls for fact-finding: Today it is possible to derive a DNA profile from a single cell. But what does a single cell actually indicate about somebody's presence or activity at a crime scene?[27]

2.2.3 *AI-evidence*

In this report AI-evidence is understood as information proffered as evidence that has been generated autonomously by tools driven by AI, i.e. produced, at least partly, with elements of machine learning. This is with the provision that, though opening up new promising sources of information, the evidence cannot be entirely understood and explained by human experts.

AI-evidence can be produced by tools that are specifically designed for forensic use, in the area of probabilistic genotyping by STRMix in the U.S. (2.2.3.1) for instance. Or it can be produced by a consumer product, like drowsiness detection alerts recorded by modern cars (2.2.3.2).

One problem defining AI-evidence is that we lack a complete understanding of what is part of the presentation of evidence before a court and what is part of the process of assessing the evidence. AI-evidence illustrates this issue, as it blurs the borders between the generation of evidence and assisting in the assessment of (or even only providing leads to) evidence. For instance, STRMix both analyses DNA material and provides probability values for matches, combining the generation of evidence with an assessment of its reliability. Similarly, 'digital forensics as a service' (2.2.3.3) and facial recognition (2.2.3.4) extract what they deem relevant data and match it with other data assessed as relevant, thereby excluding irrelevant data points ('noise') and enhancing certain patterns, and, in doing so, already assess what is relevant evidence.

2.2.3.1 Probabilistic Genotyping

Forensic DNA sample testing is an impressive example of how technology not developed for use in legal proceedings, but through science and strictly based on biological methods, changed the evidentiary perspective in legal proceedings and transformed forensic

[26] Murphy E, 'The New Forensics: Criminal Justice, False Certainty, and the Second Generation of Scientific Evidence' (2007) 95 California Law Review 721; Chessman C, 'A "source" of error: computer code, criminal defendants, and the constitution' (2017) 105 California Law Review 101; Stiffelman B, 'No Longer the Gold Standard: Probabilistic Genotyping is Changing the Nature of DNA Evidence in Criminal Trials' (2019) 24 Berkeley Journal of Criminal Law 110; Buckleton J and others, 'A Review of Likelihood Ratios in Forensic Science Based on a Critique of Stiffelman "No longer the Gold standard: Probabilistic genotyping is changing the nature of DNA evidence in criminal trials"' (2020) 310 Forensic Science International 110251.

[27] Vuille J and Taroni F, 'Measuring Uncertainty in Forensic Science' (2021) 24 IEEE Instrumentation & Measurement Magazine 5 with reference to Cook R, Evett IW, Jackson G, Jones PJ and Lambert JA, 'A hierarchy of propositions: Deciding which level to address in casework' (1998) 38 Science & Justice 231 and providing an example of wrongful conviction based on such evidence.

evidence presentation. While met with suspicion when first introduced in criminal trials,[28] it now seems a prerequisite for a guilty verdict in certain proceedings.[29]

But possibly more interesting regarding the wish for a universal standard is that forensic experts – everywhere[30] – appear to stick to a purer scientific approach when working with DNA samples (than they do, for instance, when presenting fingerprint sample testing).[31] When declaring a match, experts use population genetics data to assign the probability that an unknown person in a given relevant population would match, although they were not the source of the trace.[32] Such reports of probability assignment empower the fact finder to decide whether the probative value of the evidence is sufficient to consider that the suspect was indeed the source of the crime scene trace and whether that is sufficient as proof in the relevant jurisdictions. However, as mentioned above, (see supra 2.2.2) DNA sample testing has been a prominent example of the running of one 'evidentiary cycle': from gold standard to just another form of circumstantial evidence.[33]

STRMix promises to take DNA analysis to another level using AI. Yet, digitizing DNA test sampling and taking them to probabilistic genotyping causes new challenges: If someone wish to challenge the reliability of such an approach in a meaningful way, the notorious black box-problem must be faced. Different from other suppliers STRMix addressed the criticism of missing transparency resulting from this problem[34] and does disclose its source code based on a 'non-disclosure agreement'.[35] However, access to source

[28] See for Switzerland Donatsch A, '"DNA-Fingerprinting" zwecks Täteridentifizierung im Strafverfahren' (1991) 109 ZStR 175; Walder H, 'Der Indizienbeweis im Strafprozess' (1991) 109 ZStR 299.

[29] See for Switzerland Donatsch A, '"DNA-Fingerprinting" zwecks Täteridentifizierung im Strafverfahren' (1991) 109 ZStR 175, 188-92; Walder H, 'Der Indizienbeweis im Strafprozess' (1991) 109 ZStR 299, 302-11.

[30] See ENFSI, 'Guideline for evaluative reporting in forensic science' (2015) <http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf> accessed 28 August 2023, 11.

[31] Vuille J and Taroni F, 'Measuring Uncertainty in Forensic Science' (2021) 24 IEEE Instrumentation & Measurement Magazine 5.

[32] For more details see Vuille J and Taroni F, 'Measuring Uncertainty in Forensic Science' (2021) 24 IEEE Instrumentation & Measurement Magazine 5.

[33] *United States v Gissantaner* (2019) 417 F. Supp. 3d 857 (W.D. Mich. 2019); Chessman C, 'A "source" of error: computer code, criminal defendants, and the constitution' (2017) 105 California Law Review, 101; Stiffelman B, 'No Longer the Gold Standard: Probabilistic Genotyping is Changing the Nature of DNA Evidence in Criminal Trials' (2019) 24 Berkeley Journal of Criminal Law 110; Buckleton J and others, 'A Review of Likelihood Ratios in Forensic Science Based on a Critique of Stiffelman "No longer the Gold standard: Probabilistic genotyping is changing the nature of DNA evidence in criminal trials"' (2020) 310 Forensic Science International 110251.

[34] Imwinkelried EJ, 'Computer Source Code: A Source of the Growing Controversy Over the Reliability of Automated Forensic Techniques' (2017) 66 De Paul Law Review 97.

[35] Gill P, Benschop C, Buckleton J, Bleka Ø and Taylor D, 'A Review of Probabilistic Genotyping Systems: EuroForMix, DNAStatistX and STRmixTM' (2021) 12 Genes 1559, 1591.

code alone does not solve the transparency (and fairness) problem, without also access to training data, processing algorithms etc.[36]

Although DNA sample tests are used to prove guilt, establish innocence, or 'raise a reasonable doubt' in many jurisdictions, it is controversial what can be proven by DNA profile correspondence between a sample taken from a suspect and one found at a crime scene, as well as what the scientific standards are for matching samples.[37]

The importance of sound scientific standards when using DNA as a lead or evidence across jurisdictions has been proven by the efforts of many countries and association of states.[38] In the European Union the accreditation of forensic laboratories has become a corner stone of the close cooperation of its member states. The EU Council has ensured the integrity of DNA sample testing and its results by adopting the decision 2008/616/JHA[39] and the application of the EN ISO/IEC 17025 standard,[40] regarding the operation of testing and calibration in all relevant laboratories. With the adoption of Council Framework Decision 2009/905/JHA on the accreditation of judicial expert laboratories, all laboratories have to be accredited.[41] Against this backdrop, the employment of certified AI-tools for using DNA samples could help to re-set a standard for a science-based forensic code and thus provide the basis for ubiquitous evidentiary principles. Yet, it is not a sure-fire success. As we all know, AI-based tools come with their own pitfalls, among them is the inability to explain their reasoning or their results.[42]

2.2.3.2 Drowsiness detection alerts

Looking at the recent developments, it is quite likely that in the future we will live in smart environments that constantly monitor human behavior and – in doing so – generate information that can be useful for fact-finding in criminal proceedings. That this is

[36] Butler J, Iyer H, Press R, Taylor MK, Vallone PM and Willis S, 'DNA Mixture Interpretation: A NIST Scientific Foundation Review (2021) NISTIR 8351–DRAFT, 75 <https://doi.org/10.6028/NIST.IR.8351-draft> accessed 28 August 2023.

[37] See for further details: Murphy E, 'The New Forensics: Criminal Justice, False Certainty, and the Second Generation of Scientific Evidence' (2007) 95 California Law Review 721.

[38] See e.g. DNA Working Group: Quality Assurance Programme For DNA Laboratories <https://enfsi.eu/wp-content/uploads/2022/03/ENFSI-QA-Programme-v-16.pdf> accessed 28 August 2023.

[39] Council Decision (EC) 2008/616/JHA on the implementation of Decision 2008/615/JHA on the stepping up of cross-border cooperation, particularly in combating terrorism and cross-border crime (2008) OJ L 210/2.

[40] ISO/IEC 17025, 'General Requirements for the Competence of Testing and Calibration Laboratories' (2017): <https://www.iso.org/ISO-IEC-17025-testing-and-calibration-laboratories.html> accessed 28 August 2023.

[41] Council Framework Decision (EC) 2009/905/JHA on the accreditation of forensic service providers carrying out laboratory activities (2009) OJ L 322, 14.

[42] Roth A, 'Machine Testimony' (2017) 126 Yale Law Journal 1972, 1988.

not a utopian idea, but can easily become courtroom reality is illustrated by a case reported in 2016 by Swiss news media:[43] It involves a human driving a modern car embedded with a drowsiness detection system, who ran over a scooter driver and injured her badly. The drowsiness detection system had repeatedly alerted the driver prior to the accident of driving errors made due to perceived fatigue, alerts that were ignored by the driver. The system collected data on the driver's steering movements, body tension, seating position, respiratory rate, and eyelid movements, evaluating these indicators for 'signs of drowsiness', and, on the basis of complex algorithms and elements of machine learning, chose to issue an alert to the driver.[44] The driver's failure to heed the warnings issued by the driving assistance systems was ultimately seen to indicate negligence, which led to the driver being fined.

Differing from forensic tools, like DNA sample testing tools or 'digital forensics as a service', drowsiness detection requires more trust in the system as the data surrounding the events antecedent to an accident are unreproducible.[45] Only the recorded alert can be proffered as a proof of the unfitness of the driver to steer a vehicle. Thus, the court deciding on the merits is in a difficult position.

2.2.3.3 Digital forensics as a service (e.g. Hansken, Palantir)

Forensic tools for information processing of large volumes of digital material have been produced by government authorities, like 'Hansken' provided by the Netherlands Forensic Insitute,[46] or private companies, like Palantir,[47] which also sells its services to governments.[48]

As it has been object to Dutch case law, 'Hansken' provides a good example for an AI-tool aimed at easing the search for the 'needle in the haystack' of digitized text messages, photos, GPS data, etc. It also brings new issues to evidentiary rules with its autonomous hunt for relevant data points and the self-determined connections between them. Hansken is used by several investigative bodies in the Netherlands, including the Dutch National Police for the purpose of criminal investigation and the Dutch Fiscal Information and Investigation Service for the purpose of fraud detection in tax investigations.[49] It can

---

[43] For details of the press coverage, see 'Ex-FDP-Chef Philipp Müller wegen fahrlässiger schwerer Körperverletzung verurteilt' (watson, 31 October 2016) <https://www.watson.ch/schweiz/gesellschaft%20&%20politik/627658357-ex-fdp-chef-philipp-mueller-wegen-fahrlaessiger-schwerer-koerperverletzung-verurteilt> accessed 28 August 2023.

[44] Gless S, Di X and Silverman E, 'Ca(r)veat Emptor: Crowdsourcing Data to Challenge the Testimony of In-Car Technology (2022) 62 Jurimetrics 285, 289.

[45] Gless S, Di X and Silverman E, 'Ca(r)veat Emptor: Crowdsourcing Data to Challenge the Testimony of In-Car Technology (2022) 62 Jurimetrics 285, 289-290.

[46] <https://www.hansken.nl/an-introduction-to-hansken/the-history-of-hansken> accessed 28 August 2023.

[47] <https://www.palantir.com/offerings/> accessed 28 August 2023.

[48] Report on Germany, in this volume, p. 124-127.

[49] Seyyar MB and Geradts ZJMH, 'Privacy Impact Assessment in Large-Scale Digital Forensic Investigations' (2020) 33 Forensic Science International: Digital Investigation 1, 4.

extract and process data from all types of digital devices, such as laptops, smartphones, hard-disks and even whole servers (e.g., in the case of the Ennetcom servers).[50] This shows how 'digital forensics as a service'[51] is used to examine various types of structured and unstructured data that may be relevant for an investigation, including text (e.g., names, keywords, phone numbers, chat-messages, e-mails), photos, videos, various types of metadata, and location data.[52] However, 'digital forensics as a service' can go further and even generate 'evidence', or at least leads to evidence, dependent on the domestic concept of what actually constitutes evidence.[53] Often, intense scrutiny of what is actually produced - and by whom – will be required when 'digital forensics as a service' is promised.[54]

2.2.3.4 Facial recognition

Facial recognition tools, though now widely used, remain highly controversial. The controversy is illustrated by the rather paradoxical position taken in the European Union's proposal for an AI Act[55]: It prohibits the use of biometric remote identification systems in public spaces by police and law enforcement authorities, by pointing out the possible chilling effect on the exercise of fundamental rights.[56] Yet, the proposal carries various opening clauses for the member states that allows for the use of remote biometric identification systems based on domestic legislation, aiming at the prevention of danger or the search for victims of crime, missing children and, in certain cases, fugitives. One specific instance is the facial recognition system CATCH, used by the Dutch police, which compares an image (a still from a video or a photograph) with a large database of current or past suspects, convicted persons or other specified individuals.[57]

2.2.4 *Interim conclusion*

The use of forensic evidence and, in particular, DNA sample testing illustrates the potential that science has for fact-finding – though this comes with chances and pitfalls that are sometimes hard to digest. Ultimately, legal proof is a normative concept concerned with the establishment of the existence or non-existence of facts 'to the satisfaction of a

---

[50] Report on the Netherlands, in this volume, p. 318.

[51] <https://www.hansken.nl/an-introduction-to-hansken> accessed 28 August 2023.

[52] Seyyar MB and Geradts ZJMH, 'Privacy Impact Assessment in Large-Scale Digital Forensic Investigations' (2020) 33 Forensic Science International: Digital Investigation 1, 4.

[53] The same is true for the very interesting tools explained in the Report on Germany, <https://www.penal.org/fr/2023>, A-02, p. 41 (ZAC-AIRA).

[54] See e.g. the explanations of the German Constitutional Court declaring declaring the use of *Palantir* surveillance software by police in the states of Hesse and Hamburg unconstitutional (Germany, Bundesverfassungsgericht BVerfG, Urteil des Ersten Senats vom 16. Februar 2023, 1 BvR 1547/19, 1 BvR 2634/20).

[55] Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act, AI Act) and amending certain union legislative acts (2021) COM/2021/206 final.

[56] Art. 5 para 1 lit. d of the Proposal AI Act (Fn. 54).

[57] Report on the Netherlands, in this volume, p. 325.

legal tribunal' that decides on the verdict.[58] Probability values or other numbers are alienated to evidentiary assessments.

It is important to note that legal scholars, in particular those engaged with legal epistemology, relying on 'doxastic logic', hope to make judges rely increasingly on probability values when acknowledging information for fact-finding. The question is whether courts could use these same probability values to explain why they believe an eye witness or expert's evidence, as they do when they determine the probability of a DNA sample found at a crime scene being an accurate match to the suspect's DNA.[59] Scholars of doxastic logic would hope for a more transparent and rational consideration of evidence. However, number-based approaches to rationalizing the assessment of evidence will have difficulty succeeding, with many jurisdictions being deeply rooted in the 'intime conviction' of judges.[60] In contrast, some prominent cases illustrated that judges' handling of statistics can go wrong, thereby indicating that a number-based approach does not hand a silver bullet to the fact-finder.[61]

However, the use of AI-evidence could possibly open up a new debate. Although, in the future, forensic experts might not be able to entirely explain the sample matching procedure performed by an AI-driven tool, but could nevertheless assign numbers to probability ratios for input and output (e.g. after conducting a DNA-sample test), numbers for error rates could still help judges to assess evidence. If probability value grids were established for the assessment of evidence, and if these would grow into binding benchmarks for a rational assessment of evidence (thus bind the fact-finder *de facto* in their evaluation of evidence), it appears a logical extension to first certify the AI-tools that generate (or assess) the evidence presented in court. In certain jurisdictions (like Germany or Switzerland), instruments designed as forensic tools (like breathalyzers or radar guns) are accredited beforehand (see infra 3.2.2). Attaching probability values to their results seems less problematic than relying on information or assessments generated by consumer products (e.g. drowsiness detection alerts issued by a car) that are not based on a certified procedure, rather the opposite: car producers will want to protect business secrets regarding driving assistants like drowsiness detections systems, which makes their certification for use as evidence generating devices even more complicated.

[58]Twining W, Rethinking Evidence: Exploratory Essays (2nd edn, CUP 2006) 293.

[59] For more details see Ross L, 'The Foundations of Criminal Law Epistemology' (2022) 9 ERGO 58.

[60] Cf. Bredin J, 'Le doute et l'intime conviction' (1996) 23 Droits 21, 22; Gless S, *Beweisrechtsgrundsätze einer grenzüberschreitenden Strafverfolgung* (Vol. 2, Nomos 2006) pp. 385-389.

[61] E.g. *People v Collins* (1968) 68 Cal. 2d 319, discussed by Finkelstein M and Fairley WB, 'A Bayesian Approach to Identification Evidence' (1970) 83 Harvard Law Review 489; *R. v Sally Clark* (2003) EWCA Crim 1020; Nobles R and Schiff D, 'Misleading Statistics Within Criminal Trials – The Sally Clark Case' (2005) 2 Significance 17.

3 AI-evidence transfer across jurisdictions

Can AI-evidence overcome the old problems of evidence transfer when information obtained under the rules of procedure of one legal system as evidence cannot demand universal applicability when presented as evidence in a system with another set of procedural rules? Or phrased more concretely: Can the results of probabilistic genotyping, 'digital forensics as a service' or drowsiness detection alerts be used in criminal proceedings everywhere? Or, does such evidence also encounter the difficulty that it is generated under a normative framework based on domestic procedural rules for the obtainment and assessment of evidence, which thus cannot demand universal applicability, and so falls short of guaranteeing reliable evidence or a 'fair trial' from the perspective of the legal community in whose name the judgment is eventually rendered?

3.1 Authenticated objects of perception

At first blush, one would expect less difficulties for the transfer of AI-evidence with regard to reliability, like when DNA sample test results and probabilistic genotyping findings are exchanged across borders. As such, when results are based on natural sciences (primarily biology, less so chemistry and physics) as well as on statistics and IT technology, all (hopefully) based on a state of the art method. For the examination and comparative analysis of the evidential material, one would hope for the conceptualization of something of an 'authenticated object of perception',[62] i.e. a piece of evidence of which the information value has been confirmed based on a standard, scientific process, like in DNA sample testing.

Such an 'authenticated object of perception' could be exchanged across borders, as it can demand universal applicability when presented as evidence in a system with different procedural rules than the one it has been obtained. This is because the evaluation of the information as legal evidence within a particular criminal trial and in the context of the event under investigation (i.e. whether somebody committed an offense, who the offender was, and how the offense was committed) only follows later in the criminal trial.

It is, however, important to note that the reliability of certain information based on a scientific evaluation, like a particular DNA sample test result, can be assessed differently in multiple jurisdictions. Initiatives like the EU's accreditation of forensic laboratories[63] are crucial to fending off irregularities.

Can we build a bridge from evidence to proof across jurisdictions based on science as a global language, and thus facilitate cross-border use of evidence by way of math, i.e. statistics and probabilities? Science-oriented lawyers and forensic experts might hope so. In particular, representatives of 'legal doxasticism', who aim to insert numbers into the

[62] See supra Fn. 7.

[63] Council Framework Decision (EC) 2009/905/JHA on the accreditation of forensic service providers carrying out laboratory activities (2009) OJ L 322.

process of evidentiary reasoning, would be particularly hopeful.[64] Using an AI-based tool to evaluate evidence (DNA sample testing) could ground evidence assessment on a rational foundation that is far stronger than a human judge's belief. Such tools could, ideally, be calibrated towards objectivity, thereby providing a stronger basis for rationality. But to go down this road, we must accept statistical or mathematical probabilities as a decisive reason, in a legal sense. If this view were to be taken, in the future, we will accept that an evaluation based on a statistical assessment should be as worthy as well-defined human belief. One important question for the future is: Can a certified AI-driven device generate an 'authenticated object of perception' and, if yes, under what conditions?

This question will reoccur, for instance, when 'DNA-hits', based on sample testing are entered in a database: Will they qualify as a hit in other states? Could a drowsiness alert recorded in a car be used as evidence in all countries with jurisdiction to prosecute a fatality? Could a hit by a facial recognition system be entered into the Prüm database and could a pattern carved out by a system providing 'digital forensics as a service' be presented as circumstantial evidence in other jurisdictions?

Practitioners as well as legal scholars harbor the hope that the process of fact-finding in their respective domestic criminal justice systems could be helped by the use of AI,[65] yet it remains unclear whether and how far impediments in the production of evidence in a national criminal proceeding caused by a transfer of evidence can be rectified. One reason for doubt is that the acceptance of evidence requires not only trust in accuracy,[66] but also in the fairness of the proceeding,[67] including the evidence obtainment, to achieve 'legal truth'.[68] It is not yet clear what 'fairness' with regard to AI-evidence implies (see below 3.2.3), but there will be a need for new defense rights to uphold equality of arms or the right to confront incriminating evidence.[69]

---

[64] Ross L, 'The Foundations of Criminal Law Epistemology' (2022) 9 ERGO 58, 66. A broader point is made by Buchak L, 'Belief, credence, and norms' (2014) 169 Philosophical Studies 1, when she argues that blame must be based on belief, and not on mere credence.

[65] Neiva L, Granja R and Machado H, 'Big Data applied to criminal investigations: expectations of professionals of police cooperation in the European Union' (2022) Policing and Society 1.

[66] Ho HL, 'The Legal Concept of Evidence' in Zalta EN (ed) *The Stanford Encyclopedia of Philosophy* (Winter 2021).

[67] Gless S, Beweisrechtsgrundsätze einer grenzüberschreitenden Strafverfolgung (Vol. 2, Nomos 2006) 195-229.

[68] See for two recent accounts, from a more normative perspective: Summers SJ, 'The Epistemic Ambitions of the Criminal Trial: Truth, Proof, and Rights' (2023) 4 Quaestio Facti. Revista Internacional Sobre Razonamiento Probatorio 249; and from a more technical perspective: Stoykova RR, 'Digital evidence: Unaddressed threats to fairness and the presumption of innocence' (2021) 42 Computer Law and Security Review 105575.

[69] Gless S, 'AI in the Courtroom: A Comparative Analysis of Machine Evidence in Criminal Trials' (2020) 51 GJIL 195, 222-5.

3.2 Are domestic jurisdictions prepared for a cross-border use of 'authenticated object of perception'?

After mapping the general issue, it might be interesting to examine whether domestic jurisdictions are prepared to exchange and use AI-evidence as a form of 'authenticated object of perception'. Using the information provided in legal scholarship and in this panels' country reports as anecdotal evidence, the yield is meagre: there is little information on the use of AI-evidence and its regulation.

3.2.1 Absence of regulation

The non-appearance of AI-evidence in most country reports and the discussion about the lack of specific rules in the Netherlands,[70] Germany,[71] or Italy[72] reveals the lack of adequate regulation of this important field. Today domestic jurisdictions' evidentiary rules are still deeply rooted in the analogue world.[73] It is welcome that supranational initiatives, like the Council of Europe's (CoE) Electronic Evidence Guide[74] do provide standards for digital evidence. However, such guides do not yet take the specific problems of AI-generated evidence into account. Yet, scholars point to the pitfalls for a robust truth-finding and an effective defense for nearly a decade.[75] In particular, the infringements on traditional defense rights (like the right to meaningfully confront incriminating evidence),[76] due – among other things – to the notorious black box problem[77] or trade secrete issues of private companies involved in design, training or production.[78] Even the employment of AI-based facial recognition has not been clearly banned, despite its precarious impact, as has been pointed out during the debate on the EU proposal for an AI Act.[79] No new legal rules for the evidentiary proceeding have been adopted (as is explained in detail in the Dutch country report), but data generated by AI are regarded as increasingly

[70] Report on the Netherlands, in this volume, p. 318 and 326.
[71] Report on Germany, <https://www.penal.org/fr/2023>, A-02, p. 40.
[72] Report on Italy, <https://www.penal.org/fr/2023>, A-01, p. 23.
[73] Cf. explanations on the admissible forms to present evidence in a German criminal proceeding, Report on Germany, <https://www.penal.org/fr/2023>, A-02, p. 3.2.1.
[74] <https://www.coe.int/en/web/cybercrime/-/iproceeds-2-launching-of-the-electronic-evidence-guide-v-3-0#> accessed 28 August 2023.
[75] Cf. Imwinkelried E J, 'Computer Source Code: A Source of the Growing Controversy Over the Reliability of Automated Forensic Techniques (2016) 66 DEPAUL L. REV. 97; Roth A, 'Machine Testimony' (2017) 126 Yale Law Journal 1972, 1988.
[76] Gless S, 'AI in the Courtroom: A Comparative Analysis of Machine Evidence in Criminal Trials' (2020) 51 GJIL 195, 222-5.
[77] Gless S, 'AI in the Courtroom: A Comparative Analysis of Machine Evidence in Criminal Trials' (2020) 51 GJIL 195, 211-2.
[78] Wexler R, 'Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System', (2018) 70 STAN. L. REV. 1343, 1377–1429.
[79] Cf. Veale M and Borgesius FZ, 'Demystifying the Draft EU Artificial Intelligence Act' (2021) 4 Computer Law Review International 9; Muller C and Dignum V, 'Artificial Intelligence Act: ALLAI Analysis & Recommendations' (2021), 13 f. <https://allai.nl/wp-content/uploads/2021/08/EU-Proposal-for-Artificial-Intelligence-Act-Analysis-and-Recommendations.pdf> accessed 28 August 2023.

important tools to support criminal justice activities. Facial recognition and DNA sample testing are just two examples, but both triggered controversial debates.[80]

AI- evidence is thus regulated by general rules concerning the lawfulness, reliability and 'fairness' of evidence.[81] These rules seem outdated, as is shown in the German country report for the lack of an adequate evidentiary form to present drowsiness alerts[82], and in the Dutch country report with regard to the application of rules adopted for traditional databases to facial recognition systems or the lack of rules to monitoring 'digital forensic as service', despite its risks for a fair trial.[83]

3.2.2 Safeguards for reliability

AI-evidence is welcomed because of its promise of increased accuracy.[84] This is crucial, as criminal verdicts must be based on a set of facts that is acceptable as truth in the relevant jurisdiction.[85] But AI-evidence is not foolproof and – as has been pointed out above – comes with its own issues, among them the so-called black box problem[86] and the impossibility to check the correctness of results, in particular when not all data is processed, but only data points assessed relevant by the producer of an AI-device.[87] Thus, one would expect to see new safeguards that protect reliability – which can possibly even build trust across borders.

However, apparently, no new regulation has been adopted in domestic criminal justice systems to remedy this, even though different approaches seem possible, like a) drawing on already established principles in each jurisdiction; b) agreeing on a common approach based on legally accepted principles, e.g. *Daubert* criteria; and c) establishing a common procedure and set of criteria for certification of AI-evidence.

3.2.2.1 Draw on established principles in each jurisdiction

If fact-finders want to draw on established principles for safeguarding reliability, a wide range of possibilities opens up – reliability (like fairness) can, for instance, be protected

[80] Neiva L, Granja R and Machado H, 'Big Data applied to criminal investigations: expectations of professionals of police cooperation in the European Union' (2022) Policing and Society 1.
[81] Report on the Netherlands, in this volume, p. 326-327.
[82] Report on Germany, <https://www.penal.org/fr/2023>, A-02, p. 44; see also Gless S and Weigend T, 'Intelligente Agenten als Zeugen im Strafverfahren?' (2021) 12 Juristenzeitung 612.
[83] Report on the Netherlands, in this volume, p. 326-327.
[84] Murphy E, 'The New Forensics: Criminal Justice, False Certainty, and the Second Generation of Scientific Evidence' (2007) 95 California Law Review 721.
[85] Turner J and Weigend T, 'Negotiated justice' in Sluiter G and others (eds), *International Criminal Procedure: Principles and Rules* (OUP 2013).
[86] Gless S, 'AI in the Courtroom: A Comparative Analysis of Machine Evidence in Criminal Trials' (2020) 51 GJIL 195, 198.
[87] See for detailed information using the example of modern cars: Gless S, Di X and Silverman E, 'Ca(r)veat Emptor: Crowdsourcing Data to Challenge the Testimony of In-Car Technology (2022) 62 Jurimetrics 285.

by excluding unreliable evidence.[88] However, as the Dutch and German country reports, rightly point out: In practice, exclusion of evidence is rare.[89] Yet, 'flanking duties' may provide a safety net for reliability. For instance, in the Netherlands a court is obliged, after prosecution or defense argue that evidence submitted by the other party ought to be excluded for unreliability, to motivate a rejection of such a plea (Art. 359 para 2 Dutch CPP).

3.2.2.2 Common methodological approach

Another approach to safeguarding reliability could be the development of a common approach that has to be complied with so that AI-evidence can be regarded trustworthy.

Here, the U.S. Supreme Court could be a forerunner with its *Daubert* criteria. To determine whether a forensic method is valid it looks at five factors: (1) whether the theory or technique in question can be and has been tested; (2) whether it has been subjected to peer review and publication; (3) its known or potential error rate; (4) the existence and maintenance of standards controlling its operation; and (5) whether it has attracted widespread acceptance within a relevant scientific community.[90] The *Daubert* criteria have been adopted as an approach in other countries – like Germany – with regard to AI-based evidence, however not yet in legal matters.[91] Apparently, the Dutch Supreme Court offers similar criteria for assessing forensic expert evidence.[92] Furthermore, professional associations (like forensic experts[93]) have come up with more specific criteria for different fields.[94]

---

[88] For the application in different jurisdictions, see country reports in: Gless S and Richter T, *Do Exclusionary Rules Ensure a Fair Trial? A Comparative Perspective on Evidentiary Rules* (Springer 2019).

[89] Report on Germany, <https://www.penal.org/fr/2023>, A-02, p. 44 with reference to Thaman SC and Brodowski D, 'Exclusion or Non-Use of Illegally Gathered Evidence in the Criminal Process: Focus on Common Law and German Approaches' in Ambos K and others (eds), *Core Concepts in Criminal Law and Criminal Justice* (Vol. 1, CUP 2020); Report on the Netherlands, in this volume, p. 319, with reference to Custers B and Stevens L, 'The Use of Data as Evidence in Dutch Criminal Courts' (2021) 29 European Journal of Crime, Criminal Law and Criminal Justice 25, 40.

[90] Daubert v Merrell Dow Pharmaceuticals, Inc. (1993) 509 U.S. 579.

[91] E.g. regarding determination of authorship of a text using an AI-tool for text comparison: Ehrhardt S, 'Autorenerkennung' in Müller E, Schlothauer R and Knauer C, *Münchener Anwaltshandbuch Strafverteidigung* (3rd edn, C.H.Beck 2022) 2890.

[92] Supreme Court of the Netherlands, judgment of 27 January 1998, NJ 1984, 404; see also Custers B and Stevens L, 'The Use of Data as Evidence in Dutch Criminal Courts' (2021) 29 European Journal of Crime, Criminal Law and Criminal Justice 25, 36.

[93] ENFSI, 'Guideline for evaluative reporting in forensic science' (2015) <http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf> accessed 28 August 2023; Association of Forensic Science Providers, 'Standards for the Formulation of Evaluative Forensic Science Expert Opinion' (2009) 49 Science & Justice 161; Forensic Science Regulator, 'Codes of Practice and Conduct, Development of Evaluative Opinion, FSR-C-118' (2021) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/960051/FSR-C-118_Interpretation_Appendix_Issue_1__002_.pdf> accessed 28 August 2023.

[94] Gill P and others, 'Assessing the Value of Forensic Biological Evidence - Guidelines Highlighting the Importance of Propositions' (2018) 36 Forensic Science International: Genetics 189; Royal Society and

It is expected that the rules around forensic evidence will play a special role. But it remains unclear how powerful these rules will be, as AI-evidence does not easily conform to the traditional concepts of forensic evidence. One could argue that AI-systems analyzing evidential material for fact-finding (like DNA sample testing by STRMix) or AI-systems assessing human conduct for evidentiary purposes (like drowsiness or lie detection) can be seen as expert evidence itself. Yet, Dutch courts rejected that view (at least for 'digital forensic as a service', like Hansken).[95] One could also argue that AI-evidence (and the systems which generated them) ought to be presented and explained by a forensic expert, and, for instance, Dutch courts support that view (request for appointment of an expert for the examination of the reliability of the 'Digital Forensics as a Service'-system under Art. 227 Dutch CCP[96]). However, such an approach meets its limits when humans cannot fully trace the reckoning and assessment of large data pools by such systems.[97]

Even if a common approach to AI-evidence is formalized with a set of criteria, in order to prove a fact as a single result it must be verified for proof in an individual case and here statistical numbers do not provide a robust answer.[98] The Dutch country report illustrates the importance of the 'human in the loop' here for verifying results of facial recognition systems (requirement of a 'double human verification').[99] This is illustrated by the Dutch country report explaining how the identification an AI-facial recognition system makes is handled by courts in the Netherlands.[100]

3.2.2.3 Certification

Certification might be a key, as it allows for general validation of a tool that facilitates the verification of a result achieved by the particular (certified) method. Certification is

---

Royal Society of Edinburgh, *Forensic DNA Analysis: A Primer for Courts* (Edinburgh 2017), <https://royalsociety.org/-/media/about-us/programmes/science-and-law/royal-society-forensic-dna-analysis-primer-for-courts.pdf> accessed 28 August 2023; Aitken C and Taroni F, 'Fundamentals of statistical evidence - A primer for legal professionals' (2008) 12 International Journal of Evidence and Proof 181.

[95] Report on the Netherlands, in this volume, p. 319 with reference to District Court of Amsterdam, judgment of 19 April 2018, ECLI:NL:RBAMS:2018:2504 [7.3].

[96] Report on the Netherlands, in this volume, p. 319 with reference to District Court of Amsterdam, intermediate decision of 29 September 2020, ECLI:NL:RBAMS:2020:4764 16; District Court of Amsterdam, intermediate decision of 17 November 2020, ECLI:NL:RBAMS:2020:5585 7.

[97] Biedermann A and Vuille J, 'Understanding the logic of forensic identification decisions (without numbers)' (2018) sui generis 397.

[98] Stoney DA, 'What Made Us Ever Think We Could Individualize Using Statistics?' (1991) 31 Journal of the Forensic Science Society 197; Cole SA and Biedermann A, 'How Can a Forensic Result Be a "Decision"? A Critical Analysis of Ongoing Reforms of Forensic Reporting Formats For Federal Examiners' (2020) 57 Houston Law Review 551; Biedermann A, Bozza S and Taroni F, 'Decision Theoretic Properties of Forensic Identification: Underlying Logic and Argumentative Implications' (2008) 177 Forensic Science International 120.

[99] Report on the Netherlands, in this volume, p. 327-328.

[100] Report on the Netherlands, in this volume, p. 328 translating from 'Kamerbrief over Gebruik Gezichtsherkenningstechnologie: Waarborgen En Kaders Bij Gebruik Gezichtsherkenningstechnologie' (Ministerie van Justitie en Veiligheid, 20 November 2019).

one step towards generating an 'authenticated object of perception'[101], (see 3.1) with information content that can be used universally. DNA sample testing might be working towards this or, possibly, blood alcohol testing. With regard to the latter, legal rules or case law often establish some rule of proof based on the outcome of such tests.[102] With the employment of AI, proper scientific validation of tools and methods, in addition to the fundamental understanding of its purpose and required elements, will become a lot more complex and, at the same time, more critical to actually understanding the information content.

During the last decade, forensic sciences have presented numerous guidelines for best practice that can provide the foundations for working towards certification in different areas.[103] The criteria that emerges regularly is that certification must be based on repeatability and reproducibility. It is important to note that certification will be a long journey, and while validation of certain methods might be easily manageable, verification of a certain tool will need a lot more resources[104] and with each technological change, this verification must be repeated.[105] In order to facilitate the cross-border circulation of evidence produced by AI systems, we would, however, not only need validation of a method but also of a specific AI system and its results.

3.2.3 Safeguards for a fair trial

AI-evidence can also raise fairness issues, for instance with regard to the infringements on defense rights[106] or fair trial rights[107] (as guaranteed by Art. 6 para 3 ECHR), as well as considerations of undue invasions of privacy.[108]

As has been pointed out above (see supra 2.2.1 and 2.2.3) AI-evidence comes with issues linked to the so-called black box-problem and the inability of AI experts to explain its processes. These issues can become problematic if a defendant wishes to confront incriminating AI-evidence. It is apparent from this panel's country reports that the manner in which evidence is obtained, presented and evaluated in a criminal proceeding reflects

[101] See supra Fn. 7.
[102] Laschewski G, 'Atemalkoholanalyse und Strafverfahren – unvereinbar? – Eine aktuelle Bestandsaufnahme' (2009) 22 NZV 1.
[103] See supra Fn. 93.
[104] Guo Y, Slay J and Beckett J, 'Validation and verification of computer forensic software tools – Searching function' (2009) 6 Digital Investigation (Supplement) 12.
[105] Hall SW, Sakzad A and Choo KKR, 'Explainable artificial intelligence for digital forensics' (2022) 4.2 Wiley Interdisciplinary Reviews: Forensic Science 1434.
[106] Gless S, 'AI in the Courtroom: A Comparative Analysis of Machine Evidence in Criminal Trials' (2020) 51 GJIL 195, 222-224.
[107] For instance the presumption of innocence, for more details see e.g. Stoykova RR, 'Digital evidence: Unaddressed threats to fairness and the presumption of innocence' (2021) 42 Computer Law and Security Review 105575.
[108] For a detailed analysis on how to translate the Constitutional protection into the digital and information age see Schulhofer SJ, *The Fourth Amendment in the Twenty-first Century. More Essential Than Ever* (OUP 2012), 115–43.

particular acknowledgment of the legal position of the defendant, starting with the concept of 'fair hearing', and together works towards the right of those charged with a crime to receive a meaningful explanation of the verdict.[109] Yet, no domestic jurisdiction seems to have an adequate answer when it comes, for instance, to a meaningful right to confront incriminating AI-evidence.[110]

Only very slowly do we see the cautious beginnings of the building of adequate rights for confronting AI-evidence, for instance with the obligation of the prosecution to provide the defense with information about a particular AI-based system used to gather evidence.[111] Important cases fueling the current legal debate are those where authorities' cracked into cryptophones and exposed users relying on anonymity of encrypted communication for their business. In particular after the 'Ennetcom' and 'EncroChat cases', Dutch courts still seem generally reluctant to share information on the functioning of the relevant AI-tools with the public or the defense, obviously assuming that there is no questioning the reliability of the functioning of these tools.[112] Such an attitude curtails the defense right to question and test the reliability of evidence. Dutch courts do, however, acknowledge certain new rights of the defense, like the right to propose additional search terms, with which the prosecution should then search the whole data set.[113]

Overall, the debate is framed by an analogue world that thinks with paper documents in mind, rather than the digital pools of data that are autonomously analyzed by AI-tools (the blind spots of which are difficulties to detect for humans). This is illustrated by the courts' handling of requests for access to evidence by the defense. As the Dutch Report explains, their courts traditionally require rather concrete specification of what the defence is looking for and why, which is difficult if one doesn't know the design, training data or source code of a tool. Therefore, defense motions to receive access to the data set and the AI-tool were rejected labelled as mere 'fishing expeditions'.[114] However, more

[109] See e.g. *Berhani v Albania* App no 847/05 (ECHR, 27 May 2010) [12]; *Khamidov v Russia* App no 72118/01 (ECHR 15 November 2007) [107]; *Ajdarić v Croatia* App no 20883/09 (ECHR 13 December 2011) [47-52]; *Anđelković v Serbia* App no 1401/08 (ECHR 9 April 2013) [26-29].

[110] See in detail Gless S, 'AI in the Courtroom: A Comparative Analysis of Machine Evidence in Criminal Trials' (2020) 51 GJIL 195, 225-46.

[111] See the Report on the Netherlands, <https://www.penal.org/fr/2023>, A-04, p. 50 with reference to case law on the 'Ennetcom cases'.

[112] Report on the Netherlands, in this volume, p. 319 on 'EncroChat cases' with reference to District Court of Amsterdam, judgment of 19 April 2018, ECLI:NL:RBAMS:2018:2504 [7.3] and District Court of Gelderland, judgment of 26 June 2019, ECLI:NL:RBGEL:2019:2833 9.

[113] Report on the Netherlands, in this volume, p. 322 with reference to Court of Appeal Amsterdam, intermediate decision of 8 July 2020, ECLI:NL:GHAMS:2020:1904 13.

[114] Report on the Netherlands, in this volume, p. 323 with reference to Court of Appeal Amsterdam, judgment of 14 December 2018, ECLI:NL:GHAMS:2018:4620 section 8.

courts seem to realize that the defense needs to be afforded with a meaningful opportunity not only to examine the evidence against the defendant, but also to search for exculpatory evidence in data sets harvested by the prosecution.[115]

German courts have dealt with this problem, too. In a 2020 case involving digitized radar guns, the Federal Constitutional Court held that the right to a fair trial in principle includes a right to obtain access to all relevant raw and/or measurement data that have been stored for the purpose of the investigation, even if they were not included in the case file.[116] The Court has recognized a 'right to raw data' based on Article 2 in conjunction with Article 20 of the German Basic Law[117] and emphasized the importance of being able to trace the machine's data processing operations.[118] Even before the 2020 landmark decision, some courts had argued that defendants must be able to investigate whether there exist any doubts about the viability of the accusation; if they cannot do so, the factual basis of the conviction would ultimately be shielded from meaningful verification.[119] Yet, German jurisprudence does not grant a general right to know all raw data; for instance, if authorities use digital tools for traffic monitoring that do not record raw data, defendants cannot challenge the measuring tool's result on these grounds.[120]

A right to access all information to meaningfully defend oneself against evidence generated by AI is supported by many legal scholars. They argue for broader access for the defense to AI-tools, as well as harvested data pools (including, secondary data sets, which are the result of an initial search of the full data pool).[121] The approach is also in

---

[115] Report on the Netherlands, in this volume, p. 323 with reference to District Court of The Hague, judgment of 25 August 2021, ECLI:NL:RBDHA:2021:9368; District Court of Rotterdam, intermediate decision of 25 January 2021, ECLI:NL:RBROT:2021:396; District Court of Rotterdam, intermediate decision of 15 July 2021, ECLI:NL:RBROT:2021:6853 [4]; District Court of Amsterdam, intermediate decision of 1 April 2021, ECLI:NL:RBAMS:2021:1507; District Court of Rotterdam, intermediate decision of 25 June 2021, ECLI:NL:RBROT:2021:6113.

[116] Bundesverfassungsgericht [BVerfG] [Federal Constitutional Court], 2 BvR 1616/18, Nov.12, 2020, https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/DE/2020/11/rk20201112_2bvr161618.html (Ger.).

[117] GRUNDGESETZ FÜR DIE BUNDEREPUBLIK DEUTSCHLAND [GG] [BASIC LAW], https://www.gesetze-im-internet.de/englisch_gg/: Art. 2 subsec. 1 and Art. 20 subsec. 3.

[118] Bayerisches Oberstes Landesgericht [BayObLG] [Bavarian Higher Regional Court], Dec. 9, 2019, 202 [ObOWi] 1955/19 (Ger.), <https://www.gesetze-bayern.de/Content/Document/Y-300-Z-BECKRS-B-2019-N-31165?hl=true> (reversing prior decisions denying an obligation to disclose such data due to the assumption that calibrated and regularly monitored devices produce valid findings).

[119] Oberlandesgericht Saarbrücken [Higher Regional Court of Saarland] Sep. 3, 2019, Ss Rs 34/2019 (43/19 OWi): <https://dejure.org/dienste/vernetzung/rechtsprechung?Text=Ss%20Rs%2034/2019>, <https://www.burhoff.de/asp_weitere_beschluesse/inhalte/5294.htm> (Ger.).

[120] Bundesverfassungsgericht [BVerfG] [Federal Constitutional Court] Beschl v. 20.06.2023, Az. 2 BvR 1167/20.

[121] Galič M, 'De rechten van de verdediging in de context van omvangrijke datasets en geavanceerde zoekmachines in strafzaken: een suggestie voor uitbreiding' (2021) 2 Boom Strafblad 41; Schermer B and Oerlemans JJ, 'AI, Strafrecht En Het Recht Op Een Eerlijk Proces' (2020) 1 Computerrecht 14.

compliance with the ECtHR case law[122] and the scholarly assessment of what can be called a 'fishing expedition'. Faced with a large pool of data that can only be analyzed with an AI-tool, the principle of equality of arms requires that the defense gets access to the AI-tool and the data pool to conduct meaningful preparation to rebut charges.[123] As far as privacy is concerned, criminal procedural codes try to limit the intensity of privacy invasion, based on an approach of private sphere levels.

3.2.4 Interim conclusion

Based on the information provided in legal scholarship and this panel's reports it seems that domestic jurisdictions are not prepared to exchange and use AI-evidence as a new form of 'authenticated object of perception' that can be exploited with little cultural noise. The reason being that, not only are evidentiary proceedings still rooted in the analogue world, but also that AI-evidence – different from predictive policing or legal tech – seems not to have reached courtrooms yet. It seems likely, however, that criminal justice systems have difficulties to acknowledge regarding the specific issues of AI-evidence used for proof in criminal proceedings.[124]

It will have to be seen whether the possibilities opened up by AI to generate universally acceptable science-based evidence will come to fruition. If '[m]athematics is [a] true universal language', as Dr. Arroway claimed,[125] the traditional assertion that a certain piece of information that is obtained and processed compliant with the relevant procedures of a certain jurisdiction ought to lose weight when presented in another domestic procedural system.

4 A universal code?

Does AI-evidence open up a route for Dr. Arroway's vision of a universal code for evidence in criminal trials that allows us to – at least partially – transfer evidence across borders without having to think about what principles should govern cross-border evidence collection?[126]

This special report advocates testing the water for a universal understanding of science based 'authenticated objects of perception' approach. The aim is that certified information does not get lost in translation when information is transferred across jurisdictions. AI-based evidence opens a path based on scientific methods that evolve in various

[122] *Sigurður Einarsson and others v Iceland* App. no. 39757/15, ECLI:CE:ECHR:2019:0604JUD003975715, (ECtHR 4 June 2019); *Rook v Germany* App. no. 1586/15, ECLI:CE:ECHR:2019:0725JUD000158615 (ECtHR 25 July 2019).

[123] Custers B and Stevens L, 'The Use of Data as Evidence in Dutch Criminal Courts' (2021) 29 European Journal of Crime, Criminal Law and Criminal Justice 25.

[124] For a detailed analysis of the characteristics of evidence and proof in criminal procedure see: Ho HL, 'The Legal Concept of Evidence' in Zalta EN (ed) *The Stanford Encyclopedia of Philosophy* (Winter 2021).

[125] See supra 1.

[126] For a detailed analysis of the numerous obstacles for information transfer in traditional evidence cooperation see Gless S, *Beweisrechtsgrundsätze einer grenzüberschreitenden Strafverfolgung* (Vol. 2, Nomos 2006) 107-52.

countries in step with each other. The results of AI-driven devices should be movable across borders, like gadgets and their manuals or software with their outcomes. This is the idea behind the feeding of results of facial recognition into European wide databases.[127] Yet, as things stand today, it is too early to determine whether AI-evidence will lead to the creation of a universal code for (certain) evidence.

It could do so, if we are inspired by the possibilities, but also address the pitfalls, including the possibility that fact-finders might rely blindly on AI-evidence. Furthermore, a transfer of AI-evidence across borders will require an international standard for certification that aims at validating a certain method and providing a grid for verifying specific tools. This is an enormous task.

There is also the huge challenge of different jurisdictions adopting domestic legislation which enables them to feed AI-evidence generated abroad into their fact-finding process in a straightforward manner. It is not for nothing that the rules governing the admission and assessment of evidence in many criminal justice systems remain rather vague. In the end, it lies with fact-finders and wider society to believe evidence for it to be used for fact-finding.

This brings us back to the age-old debate about the belief necessary for establishing facts in criminal proceedings – and in real life – and the movie 'Contact'. At the end of the film, Dr. Arroway is questioned by a panel as to what kind of proof she can provide for her conviction in extra-terrestrial life. One panel member asks incredulously: 'Doctor Arroway, you come to us with no evidence, no record, no artefacts. Only a story that to put it mildly strains credibility. Over half a trillion dollars was spent, dozens of lives were lost. Are you really going to sit there and tell us we should just take this all... on faith?' And she eventually answers: 'Yes. As a scientist, I must concede that, I must volunteer that.'

Selected literature

Aitken C and Taroni F, 'Fundamentals of statistical evidence - A primer for legal professionals' (2008) 12 International Journal of Evidence and Proof 181

Association of Forensic Science Providers, 'Standards for the Formulation of Evaluative Forensic Science Expert Opinion' (2009) 49 Science & Justice 161

Becker S, Heuschkel M, Richter S and Labudde D, 'COMBI: Artificial Intelligence for Computer-Based Forensic Analysis of Persons' (2022) 36 KI-Künstliche Intelligenz 171

[127] See e.g. <https://www.telefi-project.eu/> accessed 28 August 2023.

Biedermann A and Vuille J, 'Bewertung von DNA-Untersuchungsergebnissen aus der Sicht von Gerichten und Sachverständigen: Wie viel von unserer Wahrnehmung können wir "für wahr nehmen"?' (2011) 129 ZStrR 278

Biedermann A and Vuille J, 'Understanding the logic of forensic identification decisions (without numbers)' (2018) sui generis 397

Biedermann A, Bozza S and Taroni F, 'Decision Theoretic Properties of Forensic Identification: Underlying Logic and Argumentative Implications' (2008) 177 Forensic Science International 120

Bredin J, 'Le doute et l'intime conviction' (1996) 23 Droits 21

Buchak L, 'Belief, credence, and norms' (2014) 169 Philosophical Studies 1

Buckleton J and others, 'A Review of Likelihood Ratios in Forensic Science Based on a Critique of Stiffelman "No longer the Gold standard: Probabilistic genotyping is changing the nature of DNA evidence in criminal trials"' (2020) 310 Forensic Science International 110251

Butler J, Iyer H, Press R, Taylor MK, Vallone PM and Willis S, 'DNA Mixture Interpretation: A NIST Scientific Foundation Review (2021) NISTIR 8351–DRAFT <https://doi.org/10.6028/NIST.IR.8351-draft> accessed 28 August 2023

Chessman C, 'A "source" of error: computer code, criminal defendants, and the constitution' (2017) 105 California Law Review 101

Cole SA and Biedermann A, 'How Can a Forensic Result Be a "Decision"? A Critical Analysis of Ongoing Reforms of Forensic Reporting Formats For Federal Examiners' (2020) 57 Houston Law Review 551

Cook R, Evett IW, Jackson G, Jones PJ and Lambert JA, 'A hierarchy of propositions: Deciding which level to address in casework' (1998) 38 Science & Justice 231

Custers B and Stevens L, 'The Use of Data as Evidence in Dutch Criminal Courts' (2021) 29 European Journal of Crime, Criminal Law and Criminal Justice 25

De Finetti B, 'Bayesianism: Its Unifying Role for Both the Foundations and Applications of Statistics' (1974) 42 International Statistical Review 117

Donatsch A, '"DNA-Fingerprinting" zwecks Täteridentifizierung im Strafverfahren' (1991) 109 ZStR 175

Ehrhardt S, 'Autorenerkennung' in Müller E, Schlothauer R and Knauer C, Münchener Anwaltshandbuch Strafverteidigung (3rd edn, C.H.Beck 2022) 2890

ENFSI, 'Guideline for evaluative reporting in forensic science' (2015) <http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf> accessed 28 August 2023

Finkelstein M, Fairley WB, 'A Bayesian Approach to Identification Evidence' (1970) 83 Harvard Law Review 489

Forensic Science Regulator, 'Codes of Practice and Conduct, Development of Evaluative Opinion, FSR-C-118' (2021) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/960051/FSR-C-118_Interpretation_Appendix_Issue_1__002_.pdf> accessed 28 August 2023

Galič M, 'De rechten van de verdediging in de context van omvangrijke datasets en geavanceerde zoekmachines in strafzaken: een suggestie voor uitbreiding' (2021) 2 Boom Strafblad 41

Gill P and others, 'Assessing the Value of Forensic Biological Evidence - Guidelines Highlighting the Importance of Propositions' (2018) 36 Forensic Science International: Genetics 189

Gill P, Benschop C, Buckleton J, Bleka Ø and Taylor D, 'A Review of Probabilistic Genotyping Systems: EuroForMix, DNAStatistX and STRmixTM' (2021) 12 Genes 1559
Gless S and Richter T, Do Exclusionary Rules Ensure a Fair Trial? A Comparative Perspective on Evidentiary Rules (Springer 2019).

Gless S and Weigend T, 'Intelligente Agenten als Zeugen im Strafverfahren?' (2021) 12 Juristenzeitung 612

Gless S, 'AI in the Courtroom: A Comparative Analysis of Machine Evidence in Criminal Trials' (2020) 51 GJIL 195

Gless S, 'Could Robot Judges Believe? Epistemic Ambitions of the Criminal Trial as we Approach the Digital Age. A Comment on Sarah Summers "Epistemic Ambitions of the Criminal Trial: Truth, Proof, and Rights" (2023) 5 International Journal on Evidential Legal Reasoning 1

Gless S, 'Grenzüberschreitende Beweissammlung' (2013) 125 ZStW 573

Gless S, Beweisrechtsgrundsätze einer grenzüberschreitenden Strafverfolgung (Vol. 2, Nomos 2006)

Gless S, Di X and Silverman E, 'Ca(r)veat Emptor: Crowdsourcing Data to Challenge the Testimony of In-Car Technology (2022) 62 Jurimetrics 285

Gless S, Internationales Strafrecht: Grundriss für Studium und Praxis (3rd edn, Helbing Lichtenhahn 2021)

Guo Y, Slay J and Beckett J, 'Validation and verification of computer forensic software tools – Searching function' (2009) 6 Digital Investigation (Supplement) 12

Hall SW, Sakzad A and Choo KKR, 'Explainable artificial intelligence for digital forensics' (2022) 4.2 Wiley Interdisciplinary Reviews: Forensic Science 1434

Ho HL, 'The Legal Concept of Evidence' in Zalta EN (ed) *The Stanford Encyclopedia of Philosophy* (Winter 2021)

Hoyer A, 'Der Konflikt zwischen richterlicher Beweiswürdigungsfreiheit und dem Prinzip „in dubio pro reo"' (1993) 105 ZStW 523

Imwinkelried EJ, 'Computer Source Code: A Source of the Growing Controversy Over the Reliability of Automated Forensic Techniques' (2017) 66 De Paul Law Review 97

Jackson J and Summers S, The Internalisation of Criminal Evidence: Beyond the Common Law and Civil Law Traditions (CUP 2012)

Laschewski G, 'Atemalkoholanalyse und Strafverfahren – unvereinbar? – Eine aktuelle Bestandsaufnahme' (2009) 22 NZV 1

Ministerie van Justitie en Veiligheid, 'Kamerbrief over Gebruik Gezichtsherkenningstechnologie: Waarborgen En Kaders Bij Gebruik Gezichtsherkenningstechnologie' (Ministerie van Justitie en Veiligheid, 20 November 2019)

Muller C and Dignum V, 'Artificial Intelligence Act: ALLAI Analysis & Recommendations' (2021), 13 f. <https://allai.nl/wp-content/uploads/2021/08/EU-Proposal-for-Artificial-Intelligence-Act-Analysis-and-Recommendations.pdf> accessed 28 August 2023

Murphy E, 'The New Forensics: Criminal Justice, False Certainty, and the Second Generation of Scientific Evidence' (2007) 95 California Law Review 721

Neiva L, Granja R and Machado H, 'Big Data applied to criminal investigations: expectations of professionals of police cooperation in the European Union' (2022) Policing and Society 1

Nobles R and Schiff D, 'Misleading Statistics Within Criminal Trials – The Sally Clark Case' (2005) 2 Significance 17

Raponi S, Oligeri G and Ali IM, 'Sound of guns: digital forensics of gun audio samples meets artificial intelligence' (2022) 81.21 Multimedia tools and applications 30387

Ross L, 'The Foundations of Criminal Law Epistemology' (2022) 9 ERGO 58, 66
Roth A, 'Machine Testimony' (2017) 126 Yale Law Journal 1972

Royal Society and Royal Society of Edinburgh, *Forensic DNA Analysis: A Primer for Courts* (Edinburgh 2017), <https://royalsociety.org/-/media/about-us/programmes/science-and-law/royal-society-forensic-dna-analysis-primer-for-courts.pdf> accessed 28 August 2023
Schermer B and Oerlemans JJ, 'AI, Strafrecht En Het Recht Op Een Eerlijk Proces' (2020) 1 Computerrecht 14

Schulhofer SJ, The Fourth Amendment in the Twenty-first Century. More Essential Than Ever (OUP 2012)

Seyyar MB and Geradts ZJMH, 'Privacy Impact Assessment in Large-Scale Digital Forensic Investigations' (2020) 33 Forensic Science International: Digital Investigation 1

Stiffelman B, 'No Longer the Gold Standard: Probabilistic Genotyping is Changing the Nature of DNA Evidence in Criminal Trials' (2019) 24 Berkeley Journal of Criminal Law 110

Stoney DA, 'What Made Us Ever Think We Could Individualize Using Statistics?' (1991) 31 Journal of the Forensic Science Society 197

Stoykova RR, 'Digital evidence: Unaddressed threats to fairness and the presumption of innocence' (2021) 42 Computer Law and Security Review 105575.

Stoykova RR, 'The Presumption of Innocence as a Source for Universal Rules on Digital Evidence' (2021) 22 Computer Law Review International 74.

STRmix, 'STRmix™ Will Interpret DNA Evidence for the St. Louis County Police Department' (posted on 25 January 2023, 9:00 am) <https://strmix.com/news/strmix-will-interpret-dna-evidence-for-the-st-louis-county-police-department/> accessed 28 August 2023

Summers SJ, 'The Epistemic Ambitions of the Criminal Trial: Truth, Proof, and Rights' (2023) 4 Quaestio Facti. Revista Internacional Sobre Razonamiento Probatorio 249

Thaman SC and Brodowski D, 'Exclusion or Non-Use of Illegally Gathered Evidence in the Criminal Process: Focus on Common Law and German Approaches' in Ambos K and others (eds), *Core Concepts in Criminal Law and Criminal Justice* (Vol. 1, CUP 2020)

Turner J and Weigend T, 'Negotiated justice' in Sluiter G and others (eds), *International Criminal Procedure: Principles and Rules* (OUP 2013)

Twining W, Rethinking Evidence: Exploratory Essays (2nd edn, CUP 2006)

Veale M and Borgesius FZ, 'Demystifying the Draft EU Artificial Intelligence Act' (2021) 4 Computer Law Review International 9

Vincent FHR, 'Inquiry into the circumstances that led to the conviction of Mr Farah Abdulkadir Jama' (Victoria Department of Justice 2010) <http://netk.net.au/DNA/Jama.pdf> accessed 28 August 2023

Vuille J and Taroni F, 'Measuring Uncertainty in Forensic Science' (2021) 24 IEEE Instrumentation & Measurement Magazine 5

Walder H, 'Der Indizienbeweis im Strafprozess' (1991) 109 ZStR 299

Watson, 'Ex-FDP-Chef Philipp Müller wegen fahrlässiger schwerer Körperverletzung verurteilt' (watson, 31 October 2016) <https://www.watson.ch/schweiz/gesellschaft%20&%20politik/627658357-ex-fdp-chef-philipp-mueller-wegen-fahrlaessiger-schwerer-koerperverletzung-verurteilt> accessed 28 August 2023

RESOLUTIONS

AIDP RESOLUTIONS

Section 3 – AI and the administration of criminal justice: 'Predictive policing,' 'predictive justice,' and evidence

Buenos Aires, 28–31 March 2023

Preamble

Aware that artificial intelligence (AI) is developing rapidly in contemporary society in various parts of the world. Already ubiquitous in peoples' lives in some countries, it may become part of daily life for a large part of the world's population in the future.

Noting that as a technological innovation, AI pushes consumers to buy new products, thus helping the global economy grow. Therefore, AI plays a non-negligible role in sustaining and even expanding the liberal market economy and the capitalist economic system.

Noting that the companies that create and market AI are frequently based in developed countries in the global North, and they often try to open markets all over the world.

Noting that 'digital divide' widens social inequalities among people. 'AI divide' may be the next phenomenon on the horizon.

***

Considering that AI can be defined as a set of theories and techniques used to create machines capable of simulating human intelligence.[1] As a scientific discipline, it is a blend of statistical and algorithmic mathematics, computer science, and the cognitive sciences. Symbolic AI is based on the rules of logic, whereas connectionist AI uses artificial neural networks.

[1] https://www.larousse.fr/encyclopedie/divers/intelligence_artificielle/187257.

Considering that machine learning is an example of connectionist AI, as is deep learning, which is a subset of machine learning that uses multiple layers of interconnected artificial neurons. As the number of neuronal layers enabling autonomous learning increase, the system's technological complexity increases, making the system more efficient and its calculations less explainable and traceable (deep learning).

Considering that machine learning can make technology extremely powerful, but its decision-making process can be so complex that it resembles a 'black box'.

Noting that many AI systems used in the context of preventing, investigating, detecting, and punishing crime are machine-learning systems. Using self-learning algorithms, they carry out complex probability calculations in nanoseconds. To achieve their assigned goals, they process huge amounts of data and consume a lot of energy. Some of them, such as facial recognition systems, rely on deep learning.

***

Finding that in criminal justice administration, AI systems are used to prevent or detect criminal offenses based on risk assessment ('predictive policing').

Finding that AI systems are also used to help prosecutors and judges make decisions. More specifically, the term 'predictive justice' refers to (i) anticipating someone's behavior, e.g., to assessing the risk of fleeing in the pre-trial procedure or (re-) committing a crime so that decisions concerning them, such as pre-trial detention, sentencing, parole, and probation (actuarial justice, that nowadays may be supported by AI) may be made; and (ii) using AI to perform an ultraquick statistical analysis of prior decisions issued in similar cases and of relevant legal and regulatory provisions (quantitative legal analysis or LegalTech).

Considering that quantitative legal analysis is revolutionary in the sense that a mathematical calculation is meant to support or even to supplant legal reasoning.

Observing that the word 'predictive' used in the phrases 'predictive policing' and 'predictive justice' is confusing because AI systems calculate probabilities, but do not predict the future; these probabilities are based on correlations, not on causations. These calculations nevertheless have a performative effect on people, that is, might induce them to decide in line with their results. General speaking, AI's scientific roots encourage its users to trust and follow the probabilities calculated by the AI system, since 'automation bias' is higher when the system embodies a degree of scientific aura.

Finding that AI systems contribute to innovation in the search for evidence. They can quickly analyze big data and extract information that can be useful to investigators. AI systems can establish correlations between pieces of information that are invisible to the human eye. The crime analysis diagrams they produce can be highly valuable, elaborate information for investigators.

Finding that AI systems can produce information proffered as evidence for use in criminal trials. In particular, AI systems can provide forensic information by comparing biometric traits (e.g., facial images in facial recognition), the sound frequencies of different voices (vocal recognition), and DNA fragments (probabilistic genotyping).

Finding, lastly, that AI-assisted robots and 'smart' objects in various occupational areas and daily life may incidentally produce clues or evidence that may turn out to be useful in establishing facts in a criminal case.

***

Finding that despite significant progress in the past few years, AI systems are not completely reliable. Errors may be due to the poor quality of the data used or to how the algorithm is programmed or to the existence of false positives/negatives in correlations. The probabilities produced by an AI system may therefore be inaccurate.

Finding that the results produced by AI systems are not always entirely neutral. The accuracy of the probabilities calculated by AI systems depends not only on the quality of the collected and processed data, which may reflect bias, but also depends on how the systems have learned (unsupervised vs. supervised learning). Because they reproduce human decisions, self-learning algorithms are influenced by human foibles. One result is xenophobic,[2] racist, misogynist, etc. algorithms.

Finding that AI systems pose transparency problems. So-called 'black box AI' is so opaque that even specialists cannot determine how it arrives at its results. Even scientific experts cannot fully explain a system's reasoning to a court.

Considering that AI systems used in the field of criminal justice may be developed by the private sector. Such systems are products to be sold and must be profitable. The companies that develop them generally invoke trade-secret protection to refuse to reveal their algorithm's source code, without which the system's functioning cannot be properly analyzed.

[2] https://www.amnesty.org/en/documents/eur35/4686/2021/en/

***

Observing that not everything that is technologically possible is socially desirable. In a democracy, the political choices affecting the prevention, detection, investigation, and punishment of criminal offenses must be reflected in a law or a norm of equivalent binding force.

Reiterating that human rights must be fully protected when preventing, detecting, investigating, and punishing offenses, including when technological innovations are used in that context. Whereas AI often raises issues of privacy and personal data-protection law as well as the law of non-discrimination, all laws protecting human beings, in particular their freedom and dignity, as well as all the guarantees of a fair trial, including the presumption of innocence, are potentially threatened by the use of technologies that simulate human intelligence.

Considering that national laws and international and/or regional legal norms can set out the terms on which AI-related technological innovations may be allowed to contribute to the administration of criminal justice.

Reiterating that the ethical standards often referred to by the private sector do not have the same binding force as law.

***

Aware of:

- the *Recommendation of the council on artificial intelligence*, Organization for Economic Co-operation and Development, 22 May 2019, C/MIN(2019)3/FINAL;
- the *Recommendation on the Ethics of Artificial Intelligence*, United Nations Educational, Scientific and Cultural Organization, 23 November 2021, SHS/BIO/PI/2021/1;
- the European ethical Charter on the use of Artificial intelligence in judicial systems and their environment, European Commission for the Efficiency of Justice, Strasbourg, 3-4 December 2018;
- the European Parliament Resolution of 6 October 2021 on artificial intelligence and criminal law and its use by the police and judicial authorities in criminal matters, document 2020, 2016 (INI);
- the resolutions of the XIXth International Congress of Penal Law: *Information society and penal law*, Rio de Janeiro, 2014.

Resolutions

1. Use of AI systems by public authorities for assistance when preventing, detecting, or investigating criminal offences must be authorized in advance by a law or a norm of equivalent binding force.[3]

2. States must ensure that the decisions taken by authorities to focus on preventing, detecting, or investigating a particular type of crime is based on politically and democratically determined criteria rather than on the assumption that using AI technology will make it easier to prevent, detecting or investigating this type of crime.

3. To protect the legitimacy of the public authorities' activities preventing, detecting, and investigating criminal offences, states that wish to use AI systems must choose systems the functioning of which is fully transparent, explainable, and traceable (white box AI). They must ensure that intellectual property objections cannot be raised when seeking transparency, and they should prefer publicly available, open-source systems.

4. Laws or equivalent norms related to using AI systems in the prevention, detection, and investigation of criminal offenses must require that such systems have a high degree of technological reliability. A sufficiently precise regulation requiring appropriate verifications and evaluations, both external to and independent of the AI system's developer and provider, must limit to the greatest possible extent the risk of bias or any form of discrimination in machine learning, coding errors, and other technological malfunctions.

5. Laws or equivalent norms must require that AI systems used to assist in the prevention, detection, and investigation of criminal offenses be fully accessible, verifiable, and auditable by authorities that use them and by authorities that are in charge of verifications and evaluations.

6. Laws or equivalent norms authorizing the use of AI systems to assist in the prevention, detection, and investigation of criminal offenses must require that the training data be of high quality and representativeness.

Concerning data from police or judicial files, laws or equivalent norms must institute a system that ensures that such data are correct and up-to-date and that their use does not infringe the presumption of innocence. The presumption of innocence strictly prohibits the retention and use of data collected in response to the

[3] Below we will shorten "norm of equivalent binding force" to "equivalent norm."

outcome of a predictive assessment when there is no subsequent finding of guilt, except if the data have relevance concerning another suspect.

As regards other data, in particular data accessible on social media, laws or equivalent norms must require compliance with the right to privacy and with personal data protection law when using such data. Appropriate verifications, independent of the police and judicial institutions, must be undertaken.

In general, laws or equivalent norms must be highly demanding with respect to the verification of the reliability of all data used by AI systems in connection with detecting, preventing, and investigating criminal offenses.

7. Laws or equivalent norms must require that before an AI system based on self-learning algorithms may be used in preventing, detecting, or investigating criminal offenses, the algorithms must be developed, trained, tested, and deployed under human supervision (human-in-the-loop machine learning).

These laws and equivalent norms must require a human evaluation before any action is taken to prevent, detect, or investigate criminal offenses based on the probabilities calculated by an AI system.

8. States and law enforcement authorities must ensure that their personnel who use AI to prevent, detect, or investigate criminal offenses receive hands-on training in the proper use of the relevant AI system, as well as training with respect to the risk of error and bias. They must ensure that such personnel have a thorough knowledge of the dangers AI may pose to human rights.

9. International, regional, national, or local authorities must establish independent bodies certifying the quality of AI systems intended to be used in preventing, detecting, or investigating criminal offenses. AI technology that cannot be operated and supervised in a transparent way, due to, inter alia, intellectual property rights, must not be certified.

The private sector should organize or unite to create AI-system quality labels with the goal of creating a virtuous circle for these products so that the authorities working to prevent, detect, or investigate criminal offenses are better able to determine which AI systems meet their needs.

10. All human rights must be protected when AI systems are used in preventing, detecting, or investigating criminal offenses. States and regional and international bodies must ensure that effective, proportionate, and dissuasive sanctions are imposed when such rights are violated.

Laws or equivalent norms must explicitly provide that where the cause of the violation of human rights is the technological malfunction of an AI system, the company that created the system will incur liability for fault or negligence or based on strict liability for defective products. They must also provide that investigations must be carried out to determine the cause of the violation.

11. All present resolutions are also applicable to preventing, detecting, investigating, and sanctioning administrative offenses by the competent authorities.

Resolutions specific to 'predictive policing'

12. States and regional and international human rights bodies must ensure that the use of AI systems in preventing and detecting criminal offenses does not lead to mass surveillance, which would result in a disproportionate reduction of individual freedoms (freedom of movement, freedom of expression, freedom of assembly, freedom of association, and freedom of religion).

In particular, states and local authorities must prohibit the use of AI systems to remotely identify individuals in publicly accessible spaces on the basis of their biometric data, as well as any other uses of AI systems that enable mass surveillance.

States are urged to be more transparent about their use of automated number plate recognition systems in publicly accessible space. When these systems include not only the taking of a picture of the licence plate, but also the taking of a picture of any individual in the vehicle, this option must be explicitly authorized by law. Applying facial recognition technology to the data collected through these pictures must be prohibited for the purposes of 'predictive policing'. It can only happen in the context of a specific investigation if there is a legal framework for it.

13. States must determine or have independent research bodies determine whether using AI systems in preventing criminal offenses helps decrease the number of offenses committed and, if so, in what proportion.

14. States must ensure that the financial cost of AI systems and their maintenance does not deprive the public crime-prevention services working on the *causes* of crime of funds (for psychological support, social support, training, and employment support).

15. Laws and equivalent norms must strictly prohibit the use of data as inculpatory evidence in criminal proceedings where those data were collected by an AI system in connection with crime prevention, that is, where there was no concrete

suspicion that an offense had been committed and therefore the data were collected outside the scope of the legal framework governing criminal investigations.

If data collected by an AI system in the context of crime prevention are used as the basis for investigation ('starting information), in criminal investigation as starting information, the competent judicial authority must be informed of it. The data must be marked as such and the use of AI systems must be documented on the case file.

Resolutions specific to 'predictive justice'

16. Laws and equivalent norms must strictly prohibit the use of AI systems for actuarial justice purposes in sentencing.

Punishing or aggravating the punishment of someone based on the probability that they will commit a criminal offense in the future amounts to applying punishment based in part on a criminal act that has not occurred. That is contrary to human dignity, personal freedom, and fundamental principles of criminal justice.

The use of AI risk-assessment tools must be prohibited when severe security measures, such as detention, come into consideration. When states allow the use of such tools for less severe measures, the law must expressly authorize it, with sufficient procedural safeguard. However, AI probabilities cannot constitute the only basis for a decision.

17. States that wish to use AI to assist prosecutors or/and judges with quantitative legal analysis before taking decisions in criminal cases must limit use of this technology to minor offenses that represent a high volume of cases.

18. Before deciding to use AI to facilitate management of a high volume of cases involving minor offenses, states must assess whether it would be appropriate, in light of the *ultima ratio* principle, to decriminalize the conduct generating such cases.

19. Laws and equivalent norms must prohibit the use of quantitative legal analysis for assisting judges when ruling on guilt.

20. Laws and equivalent norms must prohibit the use of quantitative legal analysis for assisting judges with sentencing. The decision to punish a person and the type of sentence must be made by humans. Otherwise, justice may be dehumanized and people's human dignity may be threatened.

21. Laws and equivalent norms must prohibit the use of quantitative legal analysis for assisting judges with decisions in criminal matters that are issued before judgment and that involve coercive measures.

22. States must ensure that decisions taken with the assistance of quantitative legal analysis do not infringe the right of access to a human judge.

23. Laws and equivalent norms must prohibit the assistance of quantitative legal analysis unless the decision can be appealed by the person concerned. The decision at appeal level shall not be based solely on the quantitative legal analysis.

Resolutions specific to evidence gathered and/or produced by AI

24. Laws and equivalent norms on extracting data for analysis by an AI system must require that before asking a person for the access code of her/his software or hardware from which data may be extracted, the seizing authority must inform the person concerned of their right not to incriminate themselves.

25. Laws and equivalent norms on crime analysis must specify that the crime analysis diagrams produced by AI systems do not have probative value, but may serve as a guide for conducting investigation.

26. Laws and equivalent norms on using AI systems to gather evidence or produce information for criminal justice purposes must clearly indicate that the output of AI systems are only probabilities. They must require that all probability-based judgments indicate not only the probability calculated by the AI system that was used, but also the error rate of that system, as calculated by the certification body that evaluated it.

27. States and judicial authorities must ensure that the use of AI-calculated probabilities does not lower the existing standard of proof in criminal proceedings.

28. Laws and equivalent norms on using AI systems to gather evidence or produce information for criminal justice purposes must prohibit the use, as evidence, of probabilities calculated by AI systems that are not fully explainable (black box AI).

29. Laws and equivalent norms on using AI systems to gather evidence or produce information for criminal justice purposes must require, pursuant to the right to adversarial hearings, that, if data collected or produced by an AI system are used, all parties must be informed of it. The data must be marked as such and the use of AI systems must be documented on the case file.

Laws and equivalent norms must require that a party's production of an AI-calculated probability may be challenged by the other party.

30. Laws and equivalent norms must set forth the principle that the party producing the probability in court must systematically include complete information on how the AI system works and which data it uses.

31. Laws and equivalent norms on using AI systems to gather evidence or produce information for criminal justice purposes must, consistent with defense rights, provide that anyone accused of an offense based on a probability proffered as evidence be able to obtain the AI system's source code and training data so that these may be analyzed by an expert. Trade secret must not be allowed to impinge on defense rights.

32. Due to the high cost of obtaining an expert analysis of an AI system, states must ensure that anyone accused of an offense based on a probability calculated by an AI system have access not only to effective legal aid but also to financial aid for such specific expertise.

Subscriptions & Membership Applications

AIDP/IAPL Membership

Annual Contribution of € 110

Benefactor Member – A member wishing to provide extra financial support to the Association. This type of membership includes subscription to the RIDP as well as online access.

Collective Member – Universities, associations, institutes, etc. This type of membership includes subscription to the RIDP.

National group – AIDP Members have established in numerous countries a National Group, which carries out its own scientific activities. Each National Group, in addition to the fees for individual members, has to pay to the AIDP a membership fee which entitles the national group to participate in the activities of the Association. This type of membership includes subscription to the RIDP.

Annual Contribution of € 85

Individual Member – The AIDP membership includes subscription the RIDP as well as online access to the RIDP archives and the RIDP *libri* series.

Annual Contribution of € 45

Young Penalist – AIDP members under the age of 35 may join the Young Penalist Group, which carries out its own activities and elects representatives to the organs of AIDP. This type of membership includes full access to the electronic archive (incl. RIDP *libri*) but no paper version of the RIDP.

Student or retiree – AIDP membership for a reduced contribution. This type of membership includes full access to the electronic archive (incl. RIDP *libri*) but no paper version of the RIDP.

Reduced-fee countries – If you are residing in a country listed on the reduced country fee list, you will be entitled to membership including a subscription to the RIDP for a limited membership fee. The list can be consulted on the AIDP website under the section 'About Us' – guidelines to establish a national group. http://www.penal.org/en/guidelines-establishment-national-groups . This type of membership includes full access to the electronic archive (incl. RIDP *libri*) but no paper version of the RIDP.

Annual Contribution of € 40

AIDP Individual Membership without RIDP subscription - mere AIDP membership without RIDP subscription and no access to electronic archives.

Membership Application instructions

The membership application form can be downloaded at the AIDP website (http://www.penal.org/en/user/register) and returned by email or mail to the address below:

Email: secretariat@penal.org. Secretariat: AIDP, c/o The Siracusa International Institute, Via Logoteta 27, 96100 Siracusa, Italy

BNP PARIBAS Bordeaux C Rouge N° IBAN : FR76 3000 4003 2000 0104 3882 870

Payment instructions

By check: Join your check to your membership application form and mail it to: AIDP secretariat, c/o The Siracusa International Institute, Via Logoteta 27, 96100 Siracusa, Italy.

Bank transfer: The bank and account details are on the membership application form. Once the bank transfer is done, send your membership application form together with a copy of the bank transfer order by fax or email, or by mail to the secretariat of the Association. The identity of the sender does not appear on the bank statement and if you do not send a copy of the bank transfer separately, we will not be able to credit the transfer to your membership.

Payment by credit card: The cryptogram is the three-digit number on the reverse side of your credit card. It is necessary for payment. Do not forget to sign your application. Please return the form by fax or email, or by mail.

For further information please consult the AIDP website http://www.penal.org/.

Subscription to the RIDP

Single Issue – price indicated for each issue on MAKLU website.

Annual Subscription – For the price of € 85, an annual subscription to the RIDP can be obtained which includes the print and free online access to the RIDP back issues. This subscription does not include AIDP Membership.

For RIDP subscription, please follow the instructions on the MAKLU publisher's website: http://www.maklu-online.eu

Artificial Intelligence systems are used today in several parts of the world to support the administration of criminal justice. The most widespread example concerns "predictive policing", which aims at foretelling crime before it happens and improving its detection. AI allows geospatial as well as person-based policing and is involved in preventing and uncovering economic crimes such as fraud and money laundering. Especially in the context of crime mapping – or hot-spot analysis –, its efficiency has been questioned. As its compliance with human rights is also critically debated, some countries have renounced or ceased to rely on it. Another kind of general surveillance of human activity has however emerged with the performance of machine learning in facial recognition technology.

In contrast, the use of risk assessment tools based on AI by judicial authorities to forecast recidivism has remained limited to a few countries. Nevertheless, a new aspect of so-called "predictive justice" is currently arising, not to foretell the forthcoming behavior of a suspected or condemned person, but surprisingly the decision of judicial bodies themselves, based largely on their former decisions. Legal quantitative analysis is a new achievement, due to AI but raises serious concerns. It may radically change the role of judges and lawyers in the course of criminal justice. Not only does it put several human rights in tension but also does it challenge the very meaning of human intervention in implementing criminal law.

The final intrusion of AI into the administration of criminal justice, addressed here, concerns evidence matters. AI tools help investigation authorities gather and correlate large volumes of data and improve the exploitation of manifold sorts of digital information. It also produces statistical evaluations that may be valuable for forensic purposes, particularly to identify persons based on facial recognition, vocal recognition, and probabilistic genotyping. Whether these results are admissible in courts, and to what conditions – including technical reliability and fair trial issues – they may be proffered as evidence, is an unsolved question for now.

This volume reviews the various uses of AI in the different stages of the criminal process from a country-comparative approach. It addresses the fundamental questions that this new technology raises when confronted with the guarantees of due process, fair trial, and other relevant human rights. It also presents the 32 resolutions that a team of twenty professors of criminal law, representing various legal traditions and parts of the world, have agreed upon to ensure that the use of AI is in line with the essential principles of criminal procedural law and with a fair justice system.

Juliette Lelieur is a Professor of Criminal Law at the University of Strasbourg, France.

www.maklu.be
ISBN 978-90-466-1224-8