

Artificial Intelligence as Evidence in Criminal Trial

Eftychia Bampasika, LL.M.

Doctoral Researcher, Member of the Otto Hahn Research Group on Alternative Criminal Justice
Max Planck Institute for the Study of Crime, Security and Law
Günterstalstraße 73, 79100, Freiburg i. Br.
e.bampasika@csl.mpg.de

ABSTRACT

This paper touches upon the intertwining of AI technology and criminal justice systems and assesses especially the issue of using AI as an evidence-generating mechanism in criminal trials. The paper revolves, in particular, around three focal points. Firstly, it sets the context for the following analysis and gives a short definition of AI. Secondly, it examines some thorny parameters of the evidentiary proceedings and focuses on the most important AI weaknesses that could jeopardise the smooth incorporation of AI in the criminal justice systems. Thirdly, it presents the ways in which AI could affect basic procedural rights of the defendant and concludes with some safety requirements and suggestions that could facilitate the transition to an AI-criminal-justice-era.

KEYWORDS

Artificial Intelligence; Criminal Justice; Evidence; Procedural Rights

1. INTRODUCTION

It is trite to say that Artificial Intelligence (AI) will reshape and is indeed already reshaping many aspects of our reality, and yet it is true. The digital transformation of the global society due to AI does not leave the justice systems around the world unaffected, bringing at this very moment the first challenges for crime control and criminal justice to the surface.

As crime becomes more and more complex, sophisticated and opaque, it is extremely difficult for the law enforcement agencies

to detect certain criminal behaviours and find their operational patterns. This fact has a negative impact on the social credibility of traditional justice systems.¹ In this context, the use of AI in the criminal justice system may prove to be of strategic importance and a game changer for prevention, investigation, fact finding and procedural economy. This in turn calls for a deeper understanding of AI's functions and operation processes within the scope of criminal justice. Given that software programmes of predictive policing,² predictive analytics and face recognition are already being used in police departments in the U.S.,³ Europe⁴ and China,⁵ and criminal justice systems have begun to use machine learning to assist in investigations for fraud and other white-collar crimes, it is imperative that legal scholars begin to dig in this uncharted and unlegislated territory. This paper focuses on the use of AI as evidence in the context of the traditional criminal trial. After briefly dealing with the definition of AI, the paper outlines the use of AI as an evidence-generating mechanism, examines the procedural rights of the defendant in view of the problems inherent to the AI technology and concludes the proposed solutions and guidelines.

2. DEFINITION OF ARTIFICIAL INTELLIGENCE

In the field of criminal law, there is a long-standing, close bond between criminal justice and technology. Over the past 150 years, criminal courts have deployed the so called 'machine evidence', in order to form and support their verdict, and the 'silent testimony of instruments' has supplemented the testimony of humans.⁶ One could just think of toxicology, ballistics, anthropometry,

¹ On this matter, see U Sieber, V Mitsilegas, C Mylonopoulos, E Billis and N Knust (Eds), *Alternative Systems of Crime Control: National, Transnational, and International Dimensions* (Berlin, Duncker & Humblot, 2018).

² For an overview on the use of predictive policing software around the world, see: <https://privacyinternational.org/examples/predictive-policing>.

³ See B Kartheuser, 'Kontrolle Ist Gut, Überwachung Ist Besser' *Der Spiegel* (27 January 2018), available at: <https://www.spiegel.de/panorama/justiz/predictive-policing-in-los-angeles-kontrolle-ist-gut-ueberwachung-ist-besser-a-1188578.html>; K Hao, 'Police Across the US Are Training Crime-Predicting AIs on Falsified Data' *The MIT Technology Review* (13 February 2019), available at: <https://www.technologyreview.com/2019/02/13/137444/predictive-policing-algorithms-ai-crime-dirty-data/>; ES Levine, J Tisch, A Tasso and M Joy, 'The New York City Police Department's Domain Awareness System' (2017) 47(1) *INOFRMS Journal on Applied Analytics* 1-15.

⁴ Meeting Report: PHRP Expert Meeting on Predictive Policing (20 September 2019), available at: https://ec.europa.eu/knowledge4policy/publication/meeting-report-phrp-expert-meeting-predictive-policing_en; F Jansen, Report on 'Data Driven Policing in the Context of Europe' (7 May 2018), available at:

<https://datajusticeproject.net/wp-content/uploads/sites/30/2019/05/Report-Data-Driven-Policing-EU.pdf>; Bundeskriminalamt, 'Das Programm "Polizei 2020"', available at: https://www.bka.de/DE/UnsereAufgaben/Ermittlungsunterstuetzung/ElektronischeFahndungsInformationssysteme/Polizei2020/Polizei2020_node.html; S Egbert, 'Siegeszug der Algorithmen? Predictive Policing im Deutschsprachigen Raum' *Bundeszentrale für Politische Bildung* (4 August 2017), available at: <http://www.bpb.de/apuz/253603/siegeszug-der-algorithmen-predictive-policing-im-deutschsprachigen-raum?p=all>.

⁵ eg, B Schmidt, 'Hong Kong Police Already Have AI Tech That Can Recognize Faces', *Bloomberg* (22 October 2019), available at: <https://www.bloomberg.com/news/articles/2019-10-22/hong-kong-police-already-have-ai-tech-that-can-recognize-faces>; P Mozur, 'One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority', *The New York Times* (14 April 2019), available at: <https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html>.

⁶ MR Damaška, *Evidence Law Adrift* (New Haven, Yale University Press, 1997) 143.

fingerprints, uhlenhuth test, maturation, forensic graphology and DNA-test.

The striking difference, though, between the AI—in the form of face, voice or video recognition, machine learning for the detection of fraud or other crimes, etc.—and the previous forensic methods of past decades is that machines, then, whenever put into use, operated according to rules that humans painstakingly programmed by hand.⁷ By contrast, an offshoot of AI, that is, machine learning, refers to a programme's ability to extract patterns from raw data.⁸ The machine has now the ability to keep improving its performance without humans having to explain exactly how to accomplish a task.⁹ Indeed, many times the programmer herself cannot account for how the machine came to a particular result, even if the result is accurate.¹⁰

For the purposes of this paper, we adopt the High-Level Expert Group on Artificial Intelligence (AI HLEG) general definition of AI, specifically that of its AI Ethics Guidelines:

'Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans 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. As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems).'¹¹

Core to the concept of AI, as stated above, is the notion of an agent capable of taking relatively autonomous decisions, depending on its perception and cognition of its environment. The emphasis on agency implies that we are not dealing with a rigid execution of rules but with systems capable of learning how to improve their performance on the basis of feedback.¹² When it comes down to the use of such programmes as evidence in a criminal trial, and

potentially as basis for the subsequent acquittal or conviction of the defendant, the focus must be, thus, on the compatibility of this particular characteristic of AI with the traditional purposes and guarantees of the criminal trial and the evidentiary process.

3. CRIMINAL TRIAL AND EVIDENTIARY PROCESS

Adjudicative fact-finding as such is bound to be conducted in conditions of uncertainty. The evidentiary process of the trial rests upon probabilities, not certainties, and hence involves a risk of error. That is why it is impossible to eliminate erroneous convictions and acquittals.¹³ The courts may aspire to ascertain the truth, but at the end of the day they must come to a decision.¹⁴ For this reason, the criminal courts often turn to science, in order to help them reach a verdict that is, as much as possible, objective and facts-based. As already mentioned, the use of machines, scientific evidence and expert witnesses in the evidentiary proceedings is not a novelty for criminal justice. Equally old is also the fact that scientific evidence might get it wrong sometimes. By their conduct, courts have expressed, over the years, a tolerance for some level of both ignorance and risk in machine evidence: ignorance in how these processes work, and risk that they might not 'get it right' every time.¹⁵ The purpose of the finality of trial trumps the purpose of finding of truth, if the latter ever was possible. The beyond-all-reasonable-doubt standard itself recognizes this inevitability of sporadically convicting innocent people.¹⁶

Further, evidence is not necessarily produced during trial. It is instead the outcome of the process of appraising what is produced at trial by the fact finder, who in doing so invokes a large storehouse of 'evidence' that is summed up in her beliefs.¹⁷ This makes it almost impossible for the fact finder to avoid heuristics and cognitive bias. Increasing reliance on machines in litigation could consequently, for some scholars, help minimize the 'whim and caprice' of the bench or jury. All in all, no one could refuse that the generation of unpredictable, idiosyncratic decisions is the antithesis to the rule of law.¹⁸ Therefore, the AI could serve as an auxiliary mechanism, assisting the court in the fact-finding process, by reducing judicial arbitrariness, systematizing the proof process and improving trial efficiency. Some proponents of the deployment of AI in the judicial system even invoke phenomena of judicial corruption, to advocate in favour of the use of the AI in the judicial field.¹⁹ This could be the case, though, if

⁷ V Fomin, 'The Shift from Traditional Computing Systems to Artificial intelligence and the Implications for Bias' in JS Gordon (ed), *Smart Technologies and Fundamental Rights* (Brill | Rodopi, 2020, to be published).

⁸ I Goodfellow, Y Bengio and A Courville, *Deep Learning*, 9th edn (Cambridge, MA, The MIT Press, 2016) 2-3; H Surden, 'Machine Learning and Law' (2014) 89 *Washington Law Review* 87-115, 88: 'Machine learning systems are computer algorithms that have the ability to learn or improve in performance over time on some task'.

⁹ On AI, in general, see the reference book of S Russell and P Norvig, *Artificial Intelligence: A Modern Approach*, 4th edn (New Jersey, Pearson, 2020).

¹⁰ A Holzinger, C Biemann, CS Pattichis and DB Kell, 'What Do We Need to Build Explainable AI Systems for the Medical Domain?': 'Often the best-performing methods are the least transparent', 2, available at: <https://arxiv.org/pdf/1712.09923.pdf>.

¹¹ Available at: <https://ec.europa.eu/futurium/en/ai-alliance-consultation>, 36.

¹² M Hildebrandt, 'Criminal Law and Technology in a Data-driven Society' in MD Dubber and T Hörnle, *The Oxford Handbook of Criminal Law* (Oxford, Oxford University Press, 2014) 175-96, 188.

¹³ A Stein, *Foundations of Evidence* (Oxford, Oxford University Press, 2005) 2.

¹⁴ A Keane, P McKeown, *The Modern Law of Evidence*, 12th edn (Oxford, Oxford University Press, 2018) 2-3. On the search for the truth as purpose of the criminal trial, see E Billis, *Die Rolle des Richters im Adversatorischen und im Inquisitorischen Beweisverfahren* (Berlin, Ducker & Humblot, 2015) 93-120.

¹⁵ PW Nutter, 'Machine Learning Evidence: Admissibility and Weight' (2019) 21(3) *Journal of Constitutional Law* 919-58, 925.

¹⁶ *ibid* 173.

¹⁷ RJ Allen, 'Artificial Intelligence and the Evidentiary Process: The Challenges of Formalism and Computation' (2001) 9 *Artificial Intelligence and Law* 99-114, 103.

¹⁸ *ibid* 101.

¹⁹ Y Cui, *Artificial Intelligence and Judicial Modernization* (Singapore, Springer, 2020) 22.

the AI deployed in the criminal justice field, could promise a high percentage of objectivity and accuracy. Research so far has shown, however, that AI is susceptible to biases and, as a result, its outcome accuracy cannot be fully trusted.²⁰ Nevertheless, since AI applications are already being used in many criminal jurisdictions around the world, it is imperative to properly examine their weak spots.

4. THE PROBLEMS OF AI AS MEANS OF EVIDENCE

As artificial intelligence does not equal artificial perfection, the weaknesses of AI as an evidence-generating mechanism must be put under a magnifying glass, in order to figure out satisfactory safety requirements conditioning its use in the realm of the criminal justice. The most problematic characteristics of AI could be summed up around the following terms: 1. inexplicability, 2. discrimination and bias and 3. lack of accountability.

4.1. Inexplicability

AI is revolutionary in its applications and capabilities, though, with respect to its potential uses in criminal justice, it is functionally similar to traditional software: data go in and conclusions come out. In between, there is a 'black box' of calculations that not only is occasionally inaccessible to the experts themselves but also few in the courtroom would understand.²¹ Here lies the danger of the AI being improperly afforded a presumption of reliability, objectivity and certainty, due to its mechanical appearance and apparently simple output.²² In order for the bench or jury to make an informed decision on the guilt of the defendant, light must somehow be shed on this black box. Moreover, given that the AI output is often inexplicable, the question of how the defendant will be able to defend herself and contest the evidence produced by it inevitably arises.

4.2. Discrimination and Bias

At the same time, decisions taken by algorithms could result from data that is incomplete and therefore not reliable: data may be tampered with by cyber-attackers, biased or simply mistaken. Applying the technology as it develops without due consideration would, therefore, lead to problematic outcomes as well as reluctance by citizens to accept its use by the courts, since the risk of malfunction always remains a distinct possibility. Thus, one of the toughest challenges for a successful incorporation of AI in criminal justice is the elimination of all kinds of biases that AI is susceptible to. Indeed, such biases may subsequently lead to poor and unjust judicial decision making, when factored in by the bench or jury. In truth, all these processes have hidden

subjectivities and errors that often go unrecognized and unchecked, thus potentially 'facilitat[ing] the masking of illegitimate or illegal discrimination behind layers upon layers of mirrors and proxies.'²³

4.3. Lack of accountability

Furthermore, when data are first gathered or generated, basic human error in collection or interpretation is common.²⁴ Human errors could occur in the training phase of the data or even later in the further development of the programme. Nevertheless, in order to establish accountability, one needs to demonstrate the person behind the programme, who did something wrong. In machine learning systems, where computer scientists are often unable to determine how or why a machine learning system has made a particular decision, this is very difficult to achieve.²⁵ Furthermore, one of the typical supportive arguments from the side of AI experts and AI companies is that AI systems and especially the machine learning ones evolve in unforeseen ways, due to their autonomous and self-learning nature. As a result, no programmer could be held liable for their evolution.

5. PROCEDURAL JUSTICE IN THE AI-ERA

In view of the characteristics of AI outlined above, legal scholars need to come up with new, effective safeguards in criminal procedure, or reinterpret those already existing. Since the use of AI will be a state privilege, at least in this phase of the digital judicial transformation, the defendants need to be equipped with the procedural rights that will preserve the equality of arms between them and the state, and the fairness of the trial. The defendant must be able, in this new criminal procedural framework, to defend herself against the all-mighty AI and contest the evidence produced by the latter.

Legal scholars must also consider changes in the law of evidence. Rules regarding the admissibility of AI generated evidence and methods to determine the reliability of its outcomes, exclusionary rules, and rules on risk-allocation are some of the problems which lie at the heart of this issue. The principle of judicial discretion must, likewise, find its place in this new environment, since the danger of over-evaluating the importance of AI generated evidence could lead to total reliance upon science and to 'abdication of responsibility by law'.²⁶ Human-computer interaction research on the biases involved in all algorithmic decision-making systems has shown that it is extremely difficult for a human decision-maker to refute a 'recommendation' made by a high-tech tool.²⁷

Furthermore, since we already witness a 'dissolution of the procedural infrastructure within the criminal justice system'²⁸

²⁰ See the ProPublica research on the COMPAS recidivism algorithm (23 May 2016), available at: <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>.

²¹ Nutter (n 15) 922.

²² A Roth, 'Trial by Machine' (2016) 104(5) *Georgetown Law Journal* 1245-306, 1269-70.

²³ O Tene and J Polonetsky, 'Judged by the Tin Man: Individual Rights in the Age of Big Data' (2013) 11 *Journal on Telecommunications and High Technology Law* 351-68, 358.

²⁴ WA Logan and AG Ferguson, 'Policing Criminal Justice Data' (2016) 101(2) *Minnesota Law Review* 549-615, 559.

²⁵ J Buyers, *Artificial Intelligence—The Practical Legal Issues* (Minehead, Somerset, Law Brief Publishing, 2018) 22.

²⁶ P Alldrige, 'Do C&IT Facilitate the Wrong Things?' (2000) 14(2) *International Review of Law, Computers & Technology* 143-54, 144.

²⁷ A Završnik, 'Criminal Justice, Artificial Intelligence Systems, and Human Rights', *ERA Forum* 20 (2020) 567-83, 574.

²⁸ A Marks, B Bowling and C Keenan, 'Automatic Justice?: Technology, Crime, and Social Control' in R Brownsword, E Scotford and K Yeung, *The Oxford Handbook of Law, Regulation and Technology* (Oxford, Oxford University Press, 2018) 705-30, 714.

because of the profiling, risk assessment and predictive analytics techniques, a new conceptualization of the fundamental procedural rights of the defendant is more than necessary. Procedural fairness is the ultimate prerequisite, if we want this new architectural scheme to work and gain social acceptance.

5.1. The presumption of innocence

The presumption of innocence was traditionally connected with a temporal distance between the criminal charge and the conviction or the acquittal.²⁹ The new AI environment challenges ‘the linear sense of time’.³⁰ In other words, it challenges the delay inherent in procedural safeguards embodying protection against hasty judgments, as we are confronted with a series of real-time decisions taken by automated decision systems based on machine learning techniques.³¹ Data-driven surveillance challenges the very foundations of the presumption of innocence by suggesting precognition of criminal intent³² and thus ‘creating a de facto presumption of guilt’.³³ Hence, some scholars go as far as advocating the construction of a ‘presumption of innocence by design’³⁴ and the interjection of ‘explanation systems’ into AI solutions, since the inculpatory evidence must have some kind of discernible logic, explanation, ability to be examined or challenged. In the context of law enforcement and intelligence, default settings of the computational technologies should prevent the reversal of the presumption of innocence by the automation of suspicion,³⁵ especially where data-mining is used to flag behaviours.

5.2. The right to confrontation

One of the oldest rights that belongs to the core of the defendant’s procedural arsenal in the Western legal systems is the right to confrontation.³⁶ The accused enjoys the right to be confronted with the witnesses against him, cross examine them and contest the incriminating evidence. Normally, the defendant would be given full access to the evidence against him, in order for him to exercise this very right. Since the inner workings of these tools are trade secrets of the companies that developed them, one wonders how the defendant would be able to effectively defend herself in lack of access to the very algorithmic assessment tool, that brought her to the stand.

5.3. The right to privacy

With the advance of big data, privacy is the right that has suffered the most. Individuals are investigated, judged and sometimes punished ‘en masse’ and ‘at a distance’,³⁷ blurring the clear distinctions between citizen, suspect, defendant, convict and

acquitted. Even at the level of the European Union where the General Data Protection Regulation³⁸ and the accompanying Law Enforcement Directive³⁹ establish strict data protection standards in the area of criminal offences and penalties, fully automated decision-making remains possible, albeit rarely. The Member States still have the possibility of providing for a decision based solely on automated processing, which produces an adverse legal effect concerning the data subject. Sole prerequisite is the authorisation by Union or Member State law, as long as it provides appropriate safeguards for the rights and freedoms of the data subject.

5.4. The principle of equality of arms

The above-mentioned impact of AI on the procedural rights disturbs the fair balance between the parties. Procedural equality of arms is designed to ‘treat the accused as a thinking and feeling human being worthy of respect, who is entitled to be given the opportunity to play an active part in procedures with a direct and possibly catastrophic impact on their life, rather than as object of state control to be manipulated for the greater good’.⁴⁰ Therefore, it is imperative for the defendant to be afforded a reasonable opportunity to present his case and take actively part in the criminal trial including the evidentiary proceedings.

6. POSSIBLE SOLUTIONS

AI has entered the premises of criminal justice systems in the aspiration to improve the procedural justice and economy, their effectiveness and efficiency. In order to fulfil those aspirations, we need greater social acceptance of AI. Trustworthy AI has three components according to the High-Level Expert Group on Artificial Intelligence (AI HLEG): (1) it should be lawful, ensuring compliance with all applicable laws and regulations, (2) it should be ethical, ensuring adherence to ethical principles and values and (3) it should be robust, both from a technical and social perspective since to ensure that, even with good intentions, AI systems do not cause any unintentional harm. The current state of the art does not provide for systems to self-report their decisions, but there is a widely held view in the relevant scientific community that regulators will force developers to interject “explanation systems” into their AI solutions when they are deployed in environments

²⁹ Hildebrandt (n 12) 181.

³⁰ Hildebrandt (n 12) 182.

³¹ Hildebrandt (n 12) 182.

³² Hildebrandt (n 12) 194–95.

³³ Hildebrandt (n 12) 184.

³⁴ Hildebrandt (n 12) 174, 195.

³⁵ Hildebrandt (n 12) 183.

³⁶ For the U.S., see P Marcus, DK Duncan, T Miller and J Moreno, *The Rights of the Accused under the Sixth Amendment*, 2nd edn (Chicago, American Bar Association, 2016). For Europe, see S Maffei, *The Right to Confrontation in Europe* (Groningen, Europa Law Publishing, 2012).

³⁷ Marks, Bowling and Keenan (n 28) 708.

³⁸ 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, available at: <https://gdpr.eu/tag/gdpr/>.

³⁹ 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, and repealing Council Framework Decision 2008/977/JHA, Article 11: 1. Member States shall provide for 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, to be prohibited unless authorised by Union or Member State law to which the controller is subject and which provides appropriate safeguards for the rights and freedoms of the data subject, at least the right to obtain human intervention on the part of the controller. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32016L0680>.

⁴⁰ P Roberts and A Zuckerman, *Criminal Evidence*, 2nd edn. (Oxford, Oxford University Press, 2010) 21.

where their outputs (or decisions) are likely to have a significant regulatory or human impact.⁴¹

According to the Ethics Guidelines for Trustworthy Artificial Intelligence prepared by the same expert group (AI HLEG), the key requirements that any AI system must fulfil in order to be accepted are: a. the human agency and oversight, b. technical robustness and safety, c. privacy and data governance, d. transparency, e. diversity, non-discrimination and fairness, f. societal and environmental well-being and g. accountability.⁴² Overall, the trust in AI control mechanisms poses many challenging regulatory questions, given the fact that trust must not be an elusive and muddled idea, but a reflection and a result of crystal-clear regulation. In order for these requirements to gain true meaning instead of remaining empty vessels, legal scholars must ally with AI experts, in order to come up with solutions that comply with the actual practice of the law as fairly and as efficiently as possible.

7. CONCLUSION

This paper served the purpose of highlighting the interplay between the criminal justice systems and the AI technology in connection with AI being employed as evidence tool, the sticking points of this risky venture, and some brief deliberations about possible solutions. The inducement that AI offers to criminal justice systems is big. The challenge will be, as *Hildebrandt* puts it, for the law to engage with AI ‘without either sacrificing or petrifying its identity’.⁴³ It is imperative, therefore, for legal scholars to cross disciplinary boundaries and work together with AI experts, in order for them to demystify and understand in depth the workings of AI. Only then they can produce relevant and applicable laws that will effectively incorporate AI in our legal reality and justice systems and regulate its possible dangers.

⁴¹ Buyers (n 25) 22–3.

⁴² Available at: <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top>. Especially on the matter of creating credibility-testing mechanisms, see A Roth, ‘Machine Testimony’ (2017) 126 *The Yale Law Journal* 1972–2051, 2022–38.

⁴³ Hildebrandt (n 12) 187.