Rosamunde Van Brakel

9 Artificial intelligence

Abstract: The chapter provides a short introduction to artificial intelligence, specifically in relation to criminology. It discusses, terminology, and definitions, the use of the term AI and relevance of AI within criminology and the social and ethical concerns of the use of AI for crime control and of AI-driven crime with attention for the implications for governance.

Keywords: artificial intelligence, algorithms, automated decision-making, facial recognition, predictive policing

Introduction

Artificial intelligence (AI) is not new; the 19th century saw early ideas around algorithms emerge by Ada Lovelace. Inspired by the 'analytical engine'—a prototype of one of the first machines that can be considered the ancestor of a modern computer designed by Charles Babbage—Lovelace wrote the first algorithm (see Algorithm by Leese). In her analyses Lovelace argued that the engine could do much more than perform numeric computations. She also argued that a machine could not create anything new but only perform the tasks exactly it was programmed for (Oliveira and Figuerido, 2024). These reflections about what machines can and cannot do as well as the issue of intelligence were picked up amongst others by Alan Turing in the 1950s. What is new and why attention for AI has increased in the 21st century is the result of three developments 1) more powerful computers, 2) the development of self-learning algorithms, and 3) the rise of Big Data, where large amounts of both structured and unstructured data can be analyzed, and new insights can be found (Van Brakel, 2023). AI has been called a 'disruptive technology' in that it questions current regulatory, policy, and organizational structures as current policy and law can hardly keep up with the speed of the technological developments and reshape everyday practices by the implementation of AI (Liu et al., 2020). In line with this AI has been compared to electricity; like electricity in the 19th century, it is assumed that AI will transform all parts of society and people will become dependent on it in the same way as people are dependent on electricity (Lynch, 2017). AI is also seen as providing numerous opportunities for crime control and criminal justice. However, due to the scale at which AI is bringing both qualitative and quantitative changes to society, many legal, ethical, and social challenges arise and especially in the context of crime control the use of AI is controversial. Illustrative, is the increase in use of (Live) Facial Recognition Technology (FRT) by police to monitor public spaces to identify people who are on a police watchlist (Fussey and Murray, 2019; see Facial Recognition by Fussey), which has been used at a Beyoncé concert in Cardiff, UK, in 2023 to amongst others identify terrorists and paedophiles (Phelps, 2023).

AI terminology and definitions

There is no widely accepted definition of AI. Historically definitions have focused on machines that show a form of 'intelligence.' Discussions about what is considered intelligence when talking about machines have been going on since Turing asked the question in his famous article 'Computing Machinery and Intelligence' (1950): "Can machines think?". Throughout the evolution of computer science, Turing and other computer scientists sought to emulate human intelligence in machines. However, in the 21st century the landscape has significantly transformed, as AI discussions have permeated various scientific disciplines and public discourse. Moreover, the heightened awareness of technological risks has contributed to a broader conceptualization of AI including simple automated applications (see Automation by Mann). This trend was reflected in the initial text of the draft European AI regulation proposed by the European Commission published in April 2021, which proposed a broad future-proof and technologically neutral definition of an 'artificial intelligence system,' which also included basic statistical software. However, the final text of the act, which is in force since August 2024, takes up a narrower definition including only technologies, that have a degree of autonomy: "*A machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments*" (European Parliament, 2024).

Since 2023 developments in generative AI have dominated public and policy debates about AI. Generative AI are deep-learning models that create voice, text, and images. These models are changing the way people access information and produce, receive, and interact with content in a revolutionary way by, for instance, providing economic rewards and driving ground-breaking scientific advances (House of Lords, 2024). Examples include text-based applications such as ChatGPT and Bard, which through analyzing and correlating large language data can, for instance, provide answers to questions, can write poems and texts. Dall-E and Midjourney are examples of image-based applications which can be used to create new images and artwork based on the images on which the AI is trained focusing on large language models (LLMs) that are the basis of ChatGPT and Bard. Bender et al. (2021) show that while it is good at generating convincing language, it does not understand the meaning of the language it is processing (called stochastic parrots) and in that way can give the impression that the AI understands what it is saying and come across as a human. Further, it has become clear that these models often generate incorrect information while presenting it as if it were a fact (called hallucinations). A good example here is if you ask ChatGPT to provide you with academic references to underscore what it has written it makes up references.

The use of the term AI and relevance of AI within criminology

The discipline of criminology has been slow to incorporate research on technology and more specifically AI. Whereas much attention has been given to cybercrime (see Cybercrime by Holt and Holt), much less attention has been given to the impact of the use of technology in criminal justice settings. Exceptions have been criminologists who have also engaged in surveillance studies (Wilson and Norris, 2006). In line with this criminological research on the application of AI in criminal justice settings and in crime in the first quarter of the 21st century is still in its infancy. While there is increased attention for digital criminology (Powell et al., 2018), publications that address the social and ethical issues of the use of AI in the criminal justice system or crime control are still limited. While we see little attention for AI in criminological research, a lot is expected of AI by policymakers when it comes to law enforcement and criminal justice. According to the Parliament, AI will contribute "*to the improvement of the working methods of police and judicial authorities, as well as to a more effective fight against certain forms of crime, in particular financial crime, money laundering and terrorist financing, sexual abuse and the exploitation of children online, as well as certain types of cybercrime, and thus to the safety and security of EU citizens*" (European Parliament, 2021).

Regarding law enforcement, machine learning algorithms are used to identify patterns and anomalies within datasets, aiding in the early detection of criminal activities, this is what is known as big data or predictive policing (Van Brakel, 2016a; Kaufmann et al., 2019). The expected future use of AI in criminology extends beyond crime prevention, encompassing investigative processes. Algorithms and machine learning facilitate the analysis of large volumes of textual data, such as social media posts or criminal records, assisting investigators in uncovering hidden connections and identifying potential threats and analysis of illicit money flows and detection of online child abuse for cybercrime investigations. Additionally, video surveillance using AI such as FRT or automatic number plate recognition aids in suspect identification, streamlining investigative procedures and administrative processes (Van Brakel, 2023).

Apart from law enforcement, applications are emerging throughout the criminal justice process. For instance, we can see an increase of the algorithmic risk assessments emerging within criminal justice settings to predict recidivism and inform decision-making in relation to probation and pre-trial release. Examples include the Hart Algorithm used by Durham constabulary as a decision support tool in the United Kingdom, COMPAS case management system and decision support tool in the United States (Oswald et al., 2018; Van Schendel, 2019).

Algorithmic risk assessments are also increasingly used in the context of social care and welfare to predict which children and families are more at risk of criminal behavior or people who are more likely to commit social fraud. Examples here include the use of Xantura children's services software in the UK, the SyRI system and ProKid

Plus algorithm used in the Netherlands (Van Brakel, 2016b; Van Schendel, 2019; Van Brakel and Govaerts, 2024).

Different forms of AI including machine-learning and automated decision-making are also increasingly being deployed in prisons such as the use of smart video surveillance in China, but also the use of AI to automatically monitor phone calls of prisoners in the US and plans in Finland to use AI to assist sentence planning and service orienting of offenders (Puolakka and Van De Steene, 2021; Kaun and Stiernstedt, 2020). Finally, AI is being used for malicious purposes and to commit crime (Europol, 2021; UNESCO 2023). Examples here include applications such as using (generative) AI for spreading misinformation, phishing, impersonation, cyber harassment and automated cyberattacks.

Social and ethical concerns of the use of AI for crime control and of AI-driven crime and implications for governance

Above applications of AI lead to potential social and ethical concerns. In this section a number of the most important concerns and what implications this has for governance will be discussed.

A first concern relates to the efficacy and effectiveness of these technologies. Little scientific evidence exists that convincingly shows the effectiveness of AI in preventing crime (Van Brakel, 2016a). For instance, in the cases of predictive policing and FRT, research has shown that there is not sufficient knowledge and research to claim that these applications are effective to prevent crime. Often technology companies claim that the algorithms have a high accuracy in prediction (Wang et al., 2023). However, these claims of 'predictive accuracy' are often mistaken for efficacy, whereas the level of accuracy does not say anything about the system's impact on crime prevention and control, making it difficult for a police force to assess a tool's real-world benefits (Babuta and Oswald, 2020).

Second, AI raises concerns for human rights. Privacy concerns are raised as AI allows for much more large-scale surveillance by government agencies, especially if safeguards and oversight are not sufficient (Van Brakel, 2021). Further, biased algorithms used in predictive policing and FRT may disproportionately target specific demographic groups, perpetuating existing societal inequalities and lead to discrimination (Van Brakel, 2016a; Buolamwini and Gebru, 2018; Babuta and Oswald, 2020). A good illustration of these harms has been the flawed matches made by FRT of five black men and one woman that led to wrongful arrests in the United States. A report written by the National Academies of Science, Engineering, and Medicine concluded that the technology can be a powerful tool for pervasive surveillance. The report raises concerns about poor performance of the technology and problematic use or misuse of the technology concluding that it requires several policy responses, whereby some uses of FRT may

cause such concern they should be prohibited. The report indicates that the above-mentioned wrongful identifications have led to false arrests and imprisonment, legal costs, interruption of normal activities of life and work, and loss of employment (National Academies, 2024). An illustration in Europe can be found in the Netherlands when looking at the impact of several algorithmic systems such as SyRI, on vulnerable communities. The SyRI system, which aimed to predict who was more likely to commit social fraud had only been applied in poor neighborhoods (van Bekkum and Borgesius-Zuiderveen, 2021). Research on generative AI has also shown the significant biases, including stereotypical associations and negative sentiment towards specific groups within LLMs such as ChatGPT (Bender et al., 2021). For instance, Gross (2023) shows how when asking ChatGPT questions about gender roles in society such as for instance "Tell me a story of parenting skills involving a mother and a father," its response is characterized by gender bias, casting the woman into the 'nurturing' role and the man as the 'adventurer,' who can build things, teach the child about nature, and promote fun.

Third, AI in crime control practices also has a significant impact on everyday practices. For instance, research by Sandhu and Fussey (2020) on predictive policing showed that many police officers have a skeptical attitude towards and reluctance to use predictive technologies because of detailed awareness of the limitations of predictive technologies, such as errors and biases in input data. Another impact concerns automation bias, whereby a person will favor automatically generated decisions over a manually generated decisions. This type of bias has been shown extensively in human–computer interaction research. Finally, deskilling, which implies that by using AI, people will lose their expertise as skills will not be practiced anymore (Van Brakel, 2023).

Fourth, with regards to malicious use of AI a report published by UNESCO (2023) argues that the proliferation of generative AI can facilitate gender-based violence and brings with it new harms, *"including the creation of more realistic fake media, hallucinations or unintended biases in the outputs, automated harassment campaigns, and the ability to build 'synthetic histories'—realistic false narratives"* (p. 3). In addition, the report argues that generative AI introduces the potential just like other AI for unintended harms via embedded biases in the model training data (see Bias by Oswald and Paul). Further, according to the report when it comes to the use of generative AI for cybercrime, such as for instance the use of ChatGPT for phishing emails, it can lead to an increase in the number of attackers, the creation of sustained and automated attacks, and the generation of content such as posts, texts, and emails that are written convincingly from different voices. Generative AI makes it possible to send out convincingly written phishing emails on a huge scale. The scale of these attacks gives existing harms a much wider reach and makes them more dangerous.

Traditional crime control governance structures in combination with new technology regulation are not sufficient to deal with the social and ethical consequences of algorithmic technologies discussed above (Van Brakel, 2021). To understand and assess the potential social and ethical concerns of these systems it is not only important to

focus on the technology itself but also the social structures they are embedded in as the issues and risks are not only the result of the technology itself but can, for instance, already be inherent in the social practices themselves or concerns that are raised by the intervention as the result of the use of technology. For instance, in the above-mentioned Top400 list the intervention that follows as the result of being put on the list by an algorithm is what is causing the most harm and not only for the young people themselves but also for their relatives such as their mothers (Van Brakel and Govaerts, 2024).

Conclusion

AI will have a significant impact on crime and crime control by automating current practices but also by shifting work practices. These developments open new and important research avenues within criminology and more specifically digital criminology. Digital criminological research can contribute to exploring and setting the socio-technical boundaries within which AI can be implemented or not implemented in crime control. Moreover, university degrees in criminology should take up digital criminology as part of the programmes so that future criminal justice practitioners and policymakers have the necessary expertise to be able to use and assess the use of AI critically. Finally, criminologists, have a unique position and expertise to make a significant contribution to regulatory and policy discussions with regards to crime control and criminal justice and more specifically when implementing AI in policing, prisons, and in the context of crime prevention programs and criminal psychology, but also when AI is being used to commit crime. This implies integrating knowledge from qualitative and ethnographic but also critical criminological research into discussions and policy-making on AI.

Suggested reading

Bennett Moses, L., & Chan, J. (2018). Algorithmic prediction in policing: Assumptions, evaluation, and accountability. *Policing & Society*, 28(7), 806–822.

Schuilenburg, M., & Peeters, R. (eds.) (2021). *The Algorithmic Society: Technology, Power and Knowledge.* London: Routledge.

Van Brakel, R. & Govaerts, L. (2024). Exploring the impact of algorithmic policing on social justice: Developing a framework for rhizomatic harm in the pre-crime society. *Theoretical Criminology, OnlineFirst.* Available at: https://doi.org/10.1177/13624806241246267.

References

Babuta, A., & Oswald, M. (2021). Machine learning predictive algorithms and the policing of future crimes: Governance and oversight. In J. L. M. McDaniel & K. Pease (eds.), *Policing and Artificial Intelligence* (pp. 214 – 226). London: Routledge.

Bender, E., Gebru, T. McMillan, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?, *FAccT '21*, March 3 – 10, 2021, Virtual Event, Canada.

Buolamwini, J., & Gebru. T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 1 – 15.

European Commission. (2021). COM (2021) 206 final 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.

European Parliament. (2021). European Parliament resolution of 6 October 2021 on artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters, https://www.euro parl.europa.eu/doceo/document/TA-9-2021-0405_EN.html.

European Parliament (2024) Regulation (EU) 2024/1689 Artificial Intelligence Act, *OJ L, 2024/1689, 12.7.2024*, 1 – 144.

Europol. (2021). *Malicious Uses and Abuses of Artificial Intelligence Europol's European Cybercrime Centre (EC3).* https://www.europol.europa.eu/cms/sites/default/files/documents/malicious_uses_and_abuses_ of_artificial_intelligence_europol.pdf.

Fussey, P., & Murray, D. (2019). *Independent Report on the London Metropolitan Police Service's Trial of Live Facial Recognition Technology.* (Human Rights and Big Data Project, University of Essex). Available at: https://48ba3m4eh2bf2sksp43rq8kkwpengine (accessed 10 December 2023).

Gross, N. (2023). What ChatGPT tells us about gender: A cautionary tale about performativity and gender. *Social Sciences*, 12(8), 435. Available at: https://doi.org/10.3390/socsci12080435 (accessed 10 December 2023).

House of Lords Communications and Digital Committee. (2024). *Large Language Models and Generative AI*, 1st Report of Session 2023 – 24. Available at: https://publications.parliament.uk/pa/ld5804/ldselect/ldcomm/54/5402.htm (accessed 10 December 2023).

Kaufmann, M., Egbert, S., & Leese, M. (2019). Predictive policing and the politics of patterns. *The British Journal of Criminology*, 59(3), 674 – 692.

Kaun, A., & Stiernstedt, F. (2020). Doing time, the smart way? Temporalities of the smart prison. *New Media & Society*, 22(9), 1580 – 1599.

Liu, H.-Y., Maas, M., Danaher, J., Scarcella, L., Lexer, M., & Van Rompuy, L. (2020). Artificial intelligence and legal disruption: A new model for analysis. *Law, Innovation and Technology*, 12(2), 205 – 258.

Lynch, S. (2017). Why AI is the new electricity? *Insights by Stanford Business*, 11 March 2017. Available at: https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity (accessed 10 October 2023).

National Academies of Sciences, Engineering, and Medicine. (2024). *Facial Recognition Technology. Current Capabilities, Future Prospects, and Governance.* Washington, DC: The National Academies Press. Available at: https://doi.org/10.17226/27397.

Oliveira, A. L., & Figuerido, M. A. T. (2024). Artificial Intelligence: Historical context and state of the art. In H. Sousa Atunes, P. M. Freitas, A. L. Oliveira, C. M. Pareira, E. Vaz de Sequeira, & L. Barreto Xavier (eds.), *Multidisciplinary Perspectives on Artificial Intelligence and the Law* (pp. 3 – 24). Dordrecht: Springer.

Oswald, M., Grace, J. Urwin, S., & Barnes, G. C. (2018). Algorithmic risk assessment policing models: Lessons from the Durham HART model and 'Experimental' proportionality. *Information and Communicaiton Technologies Law*, 27(2), 223 – 250.

Phelps, S. (2023). Beyoncé's Cardiff gig crowd was scanned for paedophiles. *BBC News*, 10 November 2023. Available at: https://www.bbc.com/news/uk-wales-67360696 (accessed 10 November 2023).

Powell, A., Stratton G., & Cameron, R. (2017). Crime and justice in digital society: Towards a 'digital criminology'? *International Journal for Crime, Justice and Social Democracy*, 6(2), 17–33.

Puolakka, P., & Van De Steene, S. (2021). Artificial Intelligence in prisons in 2030. An exploration of the future of AI in prisons. *Advancing Corrections Journal*, 11, 128–138.

Sandhu, A., & Fussey, P. (2020). The 'uberization of policing'? How police negotiate and operationalise predictive policing technology. *Policing & Society*, 31(1), 66–81.

Turing, A. (1950). Computing machinery and intelligence, *Mind*, 59(236), 433–460.

UNESCO (2023) *Technology-Facilitated Gender-Based Violence in an Era of Generative AI.* Available at https://unesdoc.unesco.org/ark:/48223/pf0000387483/PDF/387483eng.pdf.multi (accessed 12 November 2023).

Van Bekkum, M., & Borgesius-Zuiderveen, F. (2021). Digital welfare fraud detection and the Dutch *SyRI* judgment. *European Journal of Social Security*, 23(4), 323–340.

Van Brakel, R. (2016a) Pre-emptive Big Data surveillance and its (dis)empowering consequences: The case of predictive policing. In B. van der Sloot, D. Broeders, & E. Schrijvers (eds.), *Exploring the Boundaries of Big Data* (pp. 117–141). Amsterdam: Amsterdam University Press.

Van Brakel, R. (2016b). The rise of pre-emptive surveillance of children in England: Unintended social and ethical consequences. In E. Taylor & T. Rooney (eds.), *Surveillance Futures: Social and Ethical Implications of New Technologies on Children and Young People* (pp. 190–203). London: Routledge.

Van Brakel, R. (2021). How to watch the watchers? Democratic oversight of algorithmic police surveillance in Belgium. *Surveillance & Society*, 19(2), 228–240.

Van Brakel, R. (2023). Artificiële Intelligentie en veiligheid: Kansen en uitdagingen. *Politie en Recht,* 4, 151–162.

Van Brakel, R. & Govaerts, L. (2024). Exploring the impact of algorithmic policing on social justice: Developing a framework for rhizomatic harm in the pre-crime society. *Theoretical Criminology, OnlineFirst.* Available at: https://doi.org/10.1177/13624806241246267.

Van Schendel, S. (2019). The challenges of risk profiling used by law enforcement: Examining the cases of COMPAS and SyRI. In L. Reins (ed.), *Regulating New Technologies in Uncertain Times* (pp. 225–240). Dordrecht: Springer.

Wang, A., Kapoor, S., Barocas, S., & Narayanan, A. (2023). Against predictive optimization: On the legitimacy of decision-making algorithms that optimize predictive accuracy. *ACM Journal on Responsible Computing*, 1(1), 1–45. Available at: https://dl.acm.org/doi/10.1145/3636509 (accessed 2 February 2024).

Wilson, D., & Norris, C. (eds.). (2006). *Surveillance, Crime and Social Control.* London: Routledge.