Article

NJECL

New Journal of European Criminal Law
2025, Vol. 16(2) 176–196

DOI: 10.1177/20322844251338627
journals.sagepub.com/home/nje
Sage

Exploring algorithmic governance: The AI Act and new realities for criminal justice, and fundamental rights

Melina Anastasopoulou
University of Luxembourg, Luxembourg

Abstract

The article attempts to investigate what remains of our comprehension of criminal justice and the protection of fundamental rights in an algorithmic society, highlighting the new challenges posed by artificial intelligence (AI) predictive systems. The integration of algorithms from the commercial and private sectors into the public sector, particularly in policing and law enforcement, is often justified by claims of enhanced efficiency and security. However, the present manuscript argues that utilising such tools from the public sector to assess and categorise individuals based on their likelihood of engaging in criminal activity or expressing antisocial behaviour undermines the fundamental principles of traditional criminal law. This can be understood since while prediction in other areas can be deemed valuable or even life-changing, in criminal justice, predicting the future on the basis of the past threatens to reify and reproduce existing inequalities of treatment by institutions. The analysis starts by examining the notion of algorithmic governance and justice, providing the necessary conceptual framework of how the domination of algorithms in public administration and governance can have a tremendous impact on the orientation of criminal justice and fundamental rights. A key objective of the article involves the critical examination of the AI Act from a theoretical perspective, in relation to the protection of fundamental rights and the promotion of democracy, but also the scrutiny of specific provisions related to the use of intrusive discriminatory AI systems, such as emotion recognition systems and predictive policing. Finally, the article emphasises that correct political decisions are necessary so that AI systems and techniques do not serve as tools of intrusive social control.

Keywords

algorithmic governance, algorithmic justice, AI Act, predictive policing, emotion recognition

Corresponding author:
Melina Anastasopoulou, University of Luxembourg, Faculty of Law, Economics and Finance (FDEF), 4 Rue Alphonse Weicker, Luxembourg L-2721, Luxembourg.
Email: melina.anastasopoulou@uni.lu

Introduction: Governance and social power of algorithms

Artificial intelligence as a governance tool: Expansion of surveillance and secrecy

The logic of secrecy is increasingly infiltrating our political and legal systems, which are traditionally regarded as spaces of openness and transparency. The executive branches of different nations have increasingly advocated for the authority to establish and enforce 'secret law' and 'emergency law'[1] in their efforts to combat terrorism, drug organisations, recently a global pandemic,[2] or other problems considered to present a high risk to social order. Simultaneously, prominent corporations, financial institutions and governmental bodies employ nondisclosure agreements and 'proprietary methods'[3] to conceal their activities, while the private lives of individuals are becoming progressively transparent. The recording of all online activities has become a pervasive phenomenon, leaving us with question marks regarding the entities that will have access to this data as well as the duration for which it will be retained. It can also be claimed that although the utilisation of anonymising software may provide temporary protection, the vulnerability of citizens to state power remains crucial. Most importantly, it is also uncertain whether attempting to conceal one's identity serves as a conspicuous indicator for vigilant authorities.[4] Surveillance cameras, data brokers, sensor networks, and 'supercookies' are capable of capturing and storing information pertaining to our driving speed, medication consumption, reading preferences and online browsing activities. At the same time, the legal system encounters challenges in achieving optimal equilibrium, raising concerns that it exhibits a disproportionately strong inclination towards lack of transparency and secrecy while displaying a diminishing emphasis on protecting individuals' fundamental rights. The issue pertaining to the absence of transparency and comprehension, which characterises numerous aspects of our existence and has been commonly referred to as the concept of the black box. To be more precise, the term 'black box' may indicate a system characterised by enigmatic operations, wherein individuals are able to perceive its inputs and outputs while being unable to discern the transformation process from one to the other.[5] On a daily basis, individuals encounter this connotation as they become subject to increasingly pervasive monitoring by both private enterprises and governmental entities. However, there also exists a lack of clarity regarding the extent to which this information can be disseminated, the purposes for which it is employed and the resulting implications. Moreover, the significance of emphasising this structural algorithmic shift in our societies lies primarily in the fact that authority is increasingly expressed algorithmically.

1. The term emergency or secret law refers to the legal frameworks and powers activated during crises or during situations deemed to pose a significant threat to social order.
2. Stefan Braum, 'Pandemic Regulation: Virus in the Rule of Law?' in Stefan Braum (ed.), *Experimental Law* (Nomos 2023).
3. Katarina Foss-Solbrekk, 'Searchlights Across the Black Box: Trade Secrecy Versus Access to Information' [2023] Computer Law & Security Review <https://ssrn.com/abstract=4607198> accessed 1 September 2024.
4. Frank A Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2015).
5. Daniel Innerarity, 'Making the Black Box Society Transparent' (2021) 36 AI & Society <https://link.springer.com/article/10.1007/s00146-020-01130-8> accessed 1 September 2024.

In addition, it is important to highlight that the automation of government decision-making is experiencing an expanding phenomenon. Cities employ machine learning algorithms to monitor and analyse instances of gunfire,[6] ascertain optimal deployment of law enforcement personnel[7] and identify underperforming educators.[8] State agencies employ algorithms for the purpose of forecasting potential criminal conduct, whereas courts utilise 'decision-support' tools to assess the likelihood of a suspect posing a risk, determine their suitability for pre-trial release and ascertain the appropriate timing for imposing a more severe sentence.[9] The advent of emerging new technologies has brought about significant transformations in the decision-making processes and exercise of authority within governments, giving rise to a novel paradigm known as algorithmic governance. *Algorithmic governance* is a term that pertains to the utilisation of automated decision-making methodologies by governmental entities with the aim of enhancing the objectivity and reliability of the policy-making and adjudicative processes, thereby increasing their efficiency.[10] It can be, therefore, argued that decision-making becomes a matter of classification rather than a judgement of individual cases, as algorithms might classify situations or even individuals as 'high risk' or 'low risk', based solely on data patterns. As a consequence, new actors or experts are being involved in the process since the discretionary space shifts to the IT professionals who design algorithms, to the data analysts, who will identify the behavioural patterns and, in a certain way, also to the algorithms themselves that can recognise new patterns and adjust their decision-making procedures accordingly, through machine learning.[11,12]

At the same time, secrecy is also becoming widespread in criminal justice systems. One significant aspect of secrecy can be considered the nature of surveillance that government actors and police officers employ to investigate or simply monitor specific individuals who are considered to be

6. Hannah Bloch-Wehba, 'Access to Algorithms' (2020) 88(4) Fordham Law Review <https://ir.lawnet.fordham.edu/flr/vol88/iss4/2> accessed 1 September 2024.
7. Stephen Goldsmith and Chris Bousquet, 'The Right Way to Regulate Algorithms' (CityLab 2018) <https://datasmart.hks.harvard.edu/news/article/right-way-regulate-algorithms> accessed 1 September 2024.
8. Annette Bernhardt, Lisa Kresge and Reem Suleiman, *Data and Algorithms at Work* (LaborCenter 2021) <https://laborcenter.berkeley.edu/wp-content/uploads/2021/11/Data-and-Algorithms-at-Work.pdf> accessed 1 September 2024.
9. Francisco J Castro-Toledo, Fernando Miró-Llinares and Jesús C Aguerri, 'Data-Driven Criminal Justice in the Age of Algorithms: Epistemic Challenges and Practical Implications' (2023) 34 Criminal Law Forum <https://link.springer.com/article/10.1007/s10609-023-09454-y> accessed 1 September 2024.
10. Marta Cantero Gamito and Martin Ebers, 'Algorithmic Governance and Governance of Algorithms: An Introduction' in Martin Ebers and Marta Cantero Gamito (eds), *Algorithmic Governance and Governance of Algorithms* (Springer 2021) <https://doi.org/10.1007/978-3-030-50559-2_1> accessed 1 September 2024.
11. Aneesh Aneesh, 'Technologically Coded Authority: The Post-Industrial Decline in Bureaucratic Hierarchies' (Stanford University 2002) <http://web.stanford.edu/class/sts175/NewFiles/Algocratic%20Governance.pdf> accessed 1 September 2024.
12. Stavros Zouridis, Marlies van Eck and Mark Bovens, 'Automated Discretion' in Peter Hupe and Tony Evens (eds), *Palgrave Handbook on Discretion: The Quest for Controlled Freedom* (Palgrave Macmillanm 2019) <https://ssrn.com/abstract=3453068> accessed 1 December 2024.

dangerous. In addition, as a variety of studies and commentators have shown,[13] the new culture of secrecy was introduced and established after the terrorist attacks of 2001 in the United States and years later in the European Union (EU).[14] Since then, the pre-emptive narrative has indirectly legitimised exceptional legal measures on secrecy to protect public safety and national security. At the same time, the expeditious advancement of big data, machine learning technologies and the 'Internet of Things' demonstrates the recognition that algorithms have assumed a pivotal role in contemporary society, posing significant challenges to criminal justice orientation[15] and thus warranting careful consideration in the formulation of political strategies and problem-solving initiatives.[16] Recasting all complex social institutions, either as defined problems with definite, computable solutions or as transparent and self-evident processes that can be easily optimised, if there is the right algorithm in place, this quest is likely to have unexpected consequences that could eventually cause more damage than the problems they seek to address. This approach oversimplifies the intricate and multifaceted nature of social systems, which are shaped by human behaviour, cultural contexts and unpredictable variables.[17] By reducing these complexities to mere data patterns, we risk overlooking critical aspects that cannot be quantified or predicted by algorithms. The notion of having a technological solution to every social issue has been characterised as the emergence of 'solutionism'[18] within the realm of technology corporations. Furthermore, the existing body of scholarly work on security governance has revealed that private entities actively engage in these interpretation contests.[19] There is, therefore, a tendency to exaggerate certain risks while minimising others[20] or to shape the framing of security issues in a manner that favours technological

13. Jude McCulloch and Sharon Pickering, 'Pre-Crime and Counter Terrorism: Imagining Future Crime in the "War on Terror"' (2009) 49(5) The British Journal of Criminology <https://www.jstor.org/stable/23639183> accessed 1 September 2024.
14. Meghan J Ryan, 'Criminal Justice Secrets' [2022] Georgetown Law <https://www.law.georgetown.edu/american-criminal-law-review/wp-content/uploads/sites/15/2022/05/59-4_Ryan-Criminal-Justice-Secrets.pdf> accessed 1 September 2024.
15. Aleš Završnik, 'Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings' (2021) 18 European Journal of Criminology 623 <https://doi.org/10.1177/1477370819876762> accessed 1 September 2024.
16. Mihai and others, 'Artificial Intelligence-Based Decision-Making Algorithms, Internet of Things Sensing Networks, and Deep Learning-Assisted Smart Process Management in Cyber-Physical Production Systems' (2021) 10 Electronics 2497 <https://doi.org/10.3390/electronics1020249> accessed 1 September 2024.
17. Linda C. Theron and Linda Liebenberg, Understanding Cultural Contexts and Their Relationship to Resilience' in *Resilience and Sustainability in Relation to Natural Disasters: A Challenge for Future Cities* (Springer 2014) <https://link.springer.com/chapter/10.1007/978-94-017-9415-2_2> accessed 1 September 2024.
18. Evgeny Morozov, *To Save Everything, Click Here: The Folly of Technological Solutionism* (PublicAffairs 2013).
19. Hans-Georg Ehrhart, Hendrik Hegemann and Martin Kahl, 'Putting Security Governance to the Test: Conceptual, Empirical, and Normative Challenges' (2014) 23(2) European Security <https://www.tandfonline.com/doi/pdf/10.1080/09662839.2013.851676> accessed 1 December 2024.
20. Lennart Maschmeyer, Ronald J Deibert and Jon R Lindsay, 'A Tale of Two Cybers – How Threat Reporting by Cybersecurity Firms Systematically Underrepresents Threats to Civil Society' (2021) 18(1) Journal of Information Technology & Politics <https://www.tandfonline.com/doi/full/10.1080/19331681.2020.1776658> accessed 1 September 2024.

resolutions.[21] Briefly, it can also be claimed that technological companies can exercise 'ideational power', which refers to their ability to influence other actors' normative and cognitive beliefs by employing ideational elements.[22]

In addition, it can be argued that the Snowden revelations brought mass surveillance into public and academic focus, intensified awareness of the potential misuse and abuse of information, particularly on social media and brought the dangers squarely into the public frame. However, while the attention of people has indeed been intensified, the utilisation of machine learning, specifically in contexts involving sensitive matters like law enforcement or criminal justice, has prompted inquiries regarding racial profiling and bias within algorithms or the underlying data they depend on. At the same time, blind reliance on software and expert systems, which are considered less complex.[23] compared to machine learning, may subtly alter governmental processes in ways that invisibly erode the protections encapsulated in the rule of law. Despite occasional errors,[24] expert systems are currently being utilised by certain police forces, government agencies and courts, as decision-making aids.[25,26] Moreover, it is crucial to note that these decision-making and enforcement systems are not readily subject to direct scrutiny or questioning by the individuals impacted. This is primarily due to the fact that they are facilitated or even executed by complex algorithms, which are not publicly accessible, not easily comprehensible by those without expertise and not even accessible to the developers and decision-makers themselves.[27]

Yet, except for the notion of opacity as a technical feature of machine learning algorithms, the 'intentional' or corporate opacity in the form of trade secrecy also presents significant challenges.[28] A common lament concerns the refusal by developers of artificial intelligence (AI) systems to

21. Anita Lavorgna and Pamela Ugwudike, 'The Datafication Revolution in Criminal Justice: An Empirical Exploration of Frames Portraying Data-Driven Technologies for Crime Prevention and Control' (2021) 8(2) Big Data and Society <https://journals.sagepub.com/doi/full/10.1177/20539517211049670> accessed 1 September 2024.
22. Martin B Carstensen and Vivien A Schmidt, 'Power Through, Over and in Ideas: Conceptualizing Ideational Power in Discursive Institutionalism' (2016) 23(3) Journal of European Public Policy <https://www.tandfonline.com/doi/full/10.1080/13501763.2015.1115534> accessed 1 September 2024.
23. Richard Eric Susskind, 'Expert Systems in Law: A Jurisprudential Approach to Artificial Intelligence and Legal Reasoning' (1986) 49 Modern Law Review <https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1468-2230.1986.tb01683.x> accessed 1 January 2025.
24. Michael Z Bell, 'Why Expert Systems Fail' (1985) 36(7) Journal of the Operational Research Society 613, 619 <https://doi.org/10.2307/2582480> accessed 1 January 2025.
25. Ana-Maria Cornelia and others, 'Expert Systems with Applications in the Legal Domain' (2015) 19 Procedia Technology 1123, 1129 <https://doi.org/10.1016/j.protcy.2015.02.160> accessed 1 January 2025.
26. Jarek Gryz and Rojszczak Marcin, 'Black Box Algorithms and the Rights of Individuals: No Easy Solution to the "Explainability" Problem' (2021) 10(2) Internet Policy Review <https://policyreview.info/articles/analysis/black-box-algorithms-and-rights-individuals-no-easy-solution-explainability> accessed 1 September 2024.
27. Aleksandre Asatiani and others, 'Challenges of Explaining the Behavior of Black-Box AI Systems' (2020) 19(4) MIS Quarterly Executive Article 7 <https://aisel.aisnet.org/misqe/vol19/iss4/7> accessed 1 September 2024.
28. Katarina Foss-Solbrekk, 'Three Routes to Protecting AI Systems and Their Algorithms Under IP Law: The Good, the Bad and the Ugly' [2021] Journal of Intellectual Property Law & Practice <https://ssrn.com/abstract=3813797> accessed 1 January 2025.

share and indicate information about those systems, even with public authorities that wish to use them, or with those whose lives are affected by these systems.[29] While trade secrets serve the critical purpose of preventing unauthorised use of protected information by competitors, in a broader societal context, the greatest concern surrounding the excessive use of trade secrets is what some authors have described as 'confidentiality creep'.[30] The overuse of trade secrets in an information society context, or algorithmic opacity more generally, compromises essential societal values such as equality, privacy and safety. Simultaneously, it has been noted that protecting AI technologies with trade secrets contradicts the expectation that these technologies should be transparent and explainable.[31]

To address this type of opacity, the related literature suggests that the focus should primarily shift towards achieving an optimal equilibrium between the protection of private interest to keep the information confidential and the public interest of demanding an explanation.[32] Many commentators and authors have suggested that alternative explanation methods, such as subject-centric exogenous explanations using counterfactuals, can help mitigate this type of opacity.[33] In such cases, experts could be consulted to recommend the most suitable technique to address the specific inquiry while minimising the potential interference with trade secret protection. Nonetheless, although the latter might be satisfactory in some cases, in domains such as criminal justice, it is imperative that decisions regarding the public disclosure of information are not left to private initiatives and the private sector in general. The aforementioned is particularly crucial when it involves algorithms that significantly affect the lives of individuals, and individuals' rights and freedoms. For that reason, scholars have also, correctly, called for such algorithms to be made publicly accessible so they can be scrutinised by independent experts, including NGOs and academics in the field.[34] Some have suggested that any algorithms used by public authorities should be made available for public scrutiny.[35] In addition, they should be validated by independent bodies and all information made publicly available, since preventing public access to data models undermines democratic processes. When developers of proprietary algorithms decline to disclose the methods

29. Ryan Abbott (ed.), *Research Handbook on Intellectual Property and Artificial Intelligence* (Edward Elgar Publishing Limited 2022) <http://ebookcentral.proquest.com/lib/unilu-ebooks/detail.action?docID=7158652> accessed 1 January 2025.
30. David S Levine, 'Confidentiality Creep and Opportunistic Privacy' (2017) 20 Tulane Journal of Technology & Intellectual Property 11.
31. W Nicholson Price II and Arti K Rai, 'Clearing Opacity Through Machine Learning' (2021) 106 Iowa Law Review <https://ilr.law.uiowa.edu/print/volume-106-issue-2/clearing-opacity-through-machine-learning> accessed 1 January 2025.
32. Rita Matulionyte and Tatiana Aranovich, 'Trade Secrets versus the AI Explainability Principle' in Ryan Abbott (ed.), *Research Handbook on Intellectual Property and Artificial Intelligence* (Edward Elgar Publishing 2022).
33. ibid.
34. European Parliament, *A Governance Framework for Algorithmic Accountability and Transparency*, Study by the Panel for the Future of Science and Technology (2019) 48 <https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf> accessed 1 December 2024.
35. Alyssa M Carlson, 'The Need for Transparency in the Age of Predictive Sentencing Algorithms' (2017) 103 Iowa Law Review 303 <https://ilr.law.uiowa.edu/print/volume-103-issue-1/the-need-for-transparency-in-the-age-of-predictive-sentencing-algorithms> accessed 1 January 2025.

and logic underlying their models it leaves these tools 'shrouded in secrecy'.[36] Ultimately, when private companies benefit from providing public service, they should be held to the same transparency standards as public agencies.

The present manuscript will examine the emergence of algorithmic governance and its relevant implications for the orientation of criminal justice, with a specific focus on the protection of fundamental rights in an algorithmic reality. The growing utilisation and integration of AI systems in the field of crime control and criminal justice can be considered an exogenous shock for traditional criminal justice systems. This is primarily because these AI systems aim to pre-empt and anticipate potential incidents of future dangerous or criminal behaviour before the actual commission or preparation of any such acts. In addition, the critical analysis and scrutiny of the current legal policy developments in the regulatory framework of artificial intelligence – such as the EU's AI Act provide the necessary foundations to fully comprehend the related potential challenges for criminal justice and fundamental rights. Significant parts of the article constitute the critical examination of certain specific provisions of the AI Act and the dangers of establishing a form of predictive justice and a society in which people who have been algorithmically assessed as dangerous will be isolated and monitored. The manuscript aims to contribute to the doctrinal discourse on the predominance of algorithms in public governance and justice by scrutinising the ramifications of algorithmic justice and governance within the framework of the AI Act, alongside the associated political discourse.

Setting the scene: The shift to algorithmic governance

To better comprehend algorithmic governance and its impact on the orientation of criminal justice and fundamental rights, the general outlines of algorithmic governance should be analysed and explained. For that purpose, a brief historical overview of various governance models is essential in understanding the evolution that has culminated in the predominance of algorithms within public administration and governance,[37] and ultimately the consequences for their integration and influence in criminal justice.

The term 'governance' is commonly employed to denote the advent of a novel framework for economic management, characterised by a departure from the territorial, hierarchical and controlling structure prevalent in the 1930s and 1940s.[38] This paradigm shift entails a more global and pluralistic approach, albeit with reduced interventionism.[39] The development of algorithmic governance is contextualised and has its roots in the ideology of 'new public management' (NPM),

36. Danielle Keats Citron and Frank Pasquale, 'The Scored Society: Due Process for Automated Predictions' (2014) 89 *Washington Law Review* 1, 4. <https://digitalcommons.law.uw.edu/wlr/vol89/iss1/2/> accessed 1 January 2025.
37. John Danaher and others, 'Algorithmic Governance: Developing a Research Agenda through the Power of Collective Intelligence' (2017) 4(2) Big Data & Society <https://doi.org/10.1177/2053951717726554> accessed 1 September 2024.
38. Hal Kempley Colebatch, 'Making Sense of Governance' (2014) 33 Policy and Society 307, 316 <https://doi.org/10.1016/j.polsoc.2014.10.001> accessed 1 September 2024.
39. Orly Lobel, 'The Renew Deal: The Fall of Regulation and the Rise of Governance in Contemporary Legal Thought' (2004) 63 Minnesota Law Review <https://core.ac.uk/download/pdf/217211732.pdf> accessed 1 September 2024.

which gained prominence during the 1980s and 1990s, profoundly transforming the practice of public administration. NPM has been characterised as a collection of ideas that have as their primary focus the importation of private sector tools, such as efficiency, privatisation, outsourcing and performance indicators, into the public service.[40] Nevertheless, NPM has been significantly criticised for two primary reasons: the creation of opportunities for corruption through privatisation and the generation of unrealistic expectations, which destabilised existing structures.[41] Moreover, the advent of substantial technological advancements has also catalysed the transition to a new paradigm known as Digital-Era Governance. In terms of understanding the term 'digital era governance', a vital element is the use of modern information and communication technologies, especially internet web technology, by a public organisation to support and/or redefine the existing or future relations with 'stakeholders' in the internal and external environment, to create added value[42].

However, since the 1950s, when digital computers were first available as commercial products, to the present day, there has been an enormous reliance and focus on Information Communication Technology as a support to either directly or indirectly implement policy.[43] Vast quantities of data are collected, processed and presented to the decision-makers or immediately used to arrive at conclusions. The elements and the characteristics of algorithmic governance are present throughout our societies, including filtering internet search results to exclude certain illegal materials, automatically detecting credit card fraud and allocating police officers to areas and neighbourhoods where crime seems more likely to occur.[44] Although the critical large-scale deployment of those techniques and practices can be found in 'smart cities', algorithms' role and salience in public and private life is profound almost everywhere, also in cities that are not considered to be 'smart'. In most instances, the latter can be observed in the ever-increasing importance of collecting and processing data regarding individuals and the environment in which they live by digital technology as a critical element in the oversight, management and regulation of specific behaviours.[45] Therefore, it is reasonable to argue that this process encompasses more than

40. Victor Lapuente and Steven Van de Walle, 'The Effects of New Public Management on the Quality of Public Services' (2020) 33(3) Governance <https://onlinelibrary.wiley.com/doi/full/10.1111/gove.12502> accessed 1 September 2024.
41. Kennedy Rónán, 'The Rule of Law and Algorithmic Governance' in Woodrow Barfield (ed.), *The Cambridge Handbook of the Law of Algorithms* (Cambridge Law Handbooks, Cambridge University Press 2020) 209, 232
42. Rana Tassabehji, Ray Hackney and Aleš Popovič, 'Emergent Digital Era Governance: Enacting the Role of the "Institutional Entrepreneur" in Transformational Change' (2016) 33(2) Government Information Quarterly <https://www.sciencedirect.com/science/article/abs/pii/S0740624X16300338> accessed 1 September 2024.
43. James W Cortada, 'How New Technologies Spread: Lessons from Computing Technologies' (2013) 54(2) Technology and Culture <https://www.jstor.org/stable/24468014> accessed 1 September 2024.
44. Daria Gritsenko and others, 'Algorithms, Contexts, Governance: An Introduction to the Special Issue' (2022) 24(4) New Media and Society <https://journals.sagepub.com/doi/10.1177/14614448221079037> accessed 1 September 2024.
45. Amin Ullah and others, 'Smart Cities: The Role of Internet of Things and Machine Learning in Realizing a Data-Centric Smart Environment' (2024) 10 Complex & Intelligent Systems 1607 <https://link.springer.com/article/10.1007/s40747-023-01175-4#citeas> accessed 1 September 2024.

calculation processes. More accurately, it involves large-scale systems for gathering and working with information. It is, therefore, also necessary and valuable to unpack some various elements and types of notions of algorithms to comprehend the problematic issues when used and influence the criminal justice system.

Even though some of the discourse surrounding algorithms in the media approaches deification, it can be claimed that the general concept is, at heart, relatively uncomplicated. A rule-based algorithm can be regarded as a set of instructions or tasks undertaken to solve a mathematical problem.[46] The usefulness and effectiveness of the algorithm will depend very much on the correctness of the initial specification, the data which is provided to it and the final implementation: in simple words, the algorithm may not be able to solve the problem posed, or it may also produce the wrong answer because it relies on inappropriate data or even faulty, biased logic. It can also be argued that ensuring that all of these issues do not arise in practice might be a challenge for any problem, regardless of its complexity.[47] However, it is even more complicated when the rules to be applied are themselves imprecise and rely on the human capacity to fill in the gaps in a regulatory scheme. In addition, in certain areas where data management is challenging, the algorithm can be formulated to acquire knowledge through experimentation with various approaches or parameters. Consequently, it is not necessary to predefine or design all potential behaviours in advance. Furthermore, despite that this notion leads certain individuals to believe that we can aspire to attain an 'ideal type' of bureaucracy, it is essential to acknowledge that prevailing social biases and prejudices frequently contaminate both the foundational data and the computer system utilised in various contexts.[48]

Nevertheless, while some algorithms are relatively straightforward and pose no significant threat to governance or justice, others exhibit a higher degree of complexity and opacity, and this opacity can obscure their inner workings and decision-making processes. For this article, the term 'algorithms' specifically refers to machine learning algorithms, which are often characterised by their intricate nature and the difficulty in fully explaining their operations. Machine learning algorithms, learn patterns and relationships from data without being explicitly programmed.[49] More importantly, machine learning usually provides systems with the ability to learn and enhance from experience automatically without being specifically programmed and is generally referred to as the most popular latest technologies in the fourth industrial revolution.[50]

46. Noson Yanofsky, 'Towards a Definition of an Algorithm' (2006) 21(2) Journal of Logic and Computation <https://www.researchgate.net/publication/2125636_Towards_a_Definition_of_an_Algorithm#fullTextFileContent> accessed 1 September 2024.
47. Deven R Desai and Joshua A Kroll, 'Trust But Verify: A Guide to Algorithms and the Law' (2017) Harvard Journal of Law & Technology, Paper No. 17-19 <https://ssrn.com/abstract=2959472> accessed 1 December 2024.
48. Solon Barocas and Andrew D Selbst, 'Big Data's Disparate Impact' (2016) 104 California Law Review <https://ssrn.com/abstract=2477899> accessed 1 December 2024.
49. Rene Y Choi and others, 'Introduction to Machine Learning, Neural Networks, and Deep Learning' (2020) 9 Translational Vision Science & Technology 14 <https://tvst.arvojournals.org/article.aspx?articleid=2762344> accessed 1 September 2024.
50. Iqbal H Sarker, 'Machine Learning: Algorithms, Real-World Applications and Research Directions' (2021) 2 SN Computer Science <https://link.springer.com/article/10.1007/s42979-021-00592-x#citeas> accessed 1 September 2024.

Although there is some controversy regarding the degree of their understandability,[51] rule-based approaches of early AI are believed to offer comprehensible decision-making processes.[52] Nevertheless, the same comprehensibility cannot necessarily be applied to machine learning algorithms, leading to challenges in understanding the rationale behind their predictions or decisions. Finally, in the end, the system's results must be useful to individuals. The crucial issue at this point is the fact that the choices that are made here will significantly structure the understandings, perceptions and procedures applied by the individuals involved. Various AI tools, particularly those utilising machine learning, have been developed with the purpose of analysing extensive datasets, identifying concealed patterns and providing solutions in various domains, such as the legal sector. The issue at hand pertains to the lack of knowledge regarding the process and rationale behind the algorithm's generated solution, which can be attributed to a multitude of factors. Consequently, the crucial question lies in whether individuals would be satisfied with accepting such a judgement unconditionally or if they would instead demand a fundamental understanding of the algorithm's decision-making process.[53] This dilemma highlights the critical need for transparency and explainability in AI systems, especially in high-stakes domains such as criminal justice. The consequences of algorithmic decisions in these areas can profoundly affect individuals' lives and the public's trust in legal institutions.

Algorithmic justice in the profiling era: Governing the 'others'

The fast-growing use of algorithms in the fields of justice, policing and public welfare could end in biased and erroneous decisions, boosting inequality, discrimination and unfair consequences and undermining constitutional rights, such as privacy, freedom of expression and equality.[54] It is important to highlight that this use raises considerable concerns not only for the specific policy area in which they are operated but also for our society as a whole, since there is an increasing perception that humans do not have complete control over the algorithmic state decision-making processes.[55] Despite their predictive outperformance over analogue tools, algorithmic decisions are difficult, if not impossible, to understand and explain. More importantly, this lack of transparency and the diminished ability to understand the operation of systems used in criminal justice and, in

51. Andrew Bell and others, 'It's Just Not That Simple: An Empirical Study of the Accuracy-Explainability Trade-off in Machine Learning for Public Policy' (2022 ACM Conference On Fairness, Accountability, and Transparency (FAccT '22), Seoul, Republic of Korea, 21–24 June 2022) <https://dl.acm.org/doi/pdf/10.1145/3531146.3533090> accessed 1 December 2024.
52. Grzegorz J Nalepa, 'Diversity of Rule-based Approaches: Classic Systems and Recent Applications' (2016) 7(2) AVANT: Trends in Interdisciplinary Studies <https://avant.edu.pl/wp-content/uploads/Nalepa-Diversity_of_Rule-based_Approaches.pdf?> accessed 1 December 2024
53. Gryz and Rojszczak, 'Black Box Algorithms and the Rights of Individuals' (n 26).
54. Oreste Pollicino and Giovanni De Gregorio, 'Constitutional Law in the Algorithmic Society' in Hans-W Micklitz and others (eds), *Constitutional Challenges in the Algorithmic Society* (Cambridge University Press 2021) 3, 24.
55. Claudia Aradau and Tobias Blanke, 'Governing others: Anomaly and the Algorithmic Subject of Security' (2017) 3(1) European Journal of International Security <https://www.cambridge.org/core/journals/european-journal-of-international-security/article/governing-others-anomaly-and-the-algorithmic-subject-of-security/784AB53AF6E2D81A9E5928CC204B913E> accessed 1 September 2024.

general, by the structures of governance is significantly challenging and eroding traditional notions underpinning the rule of law.[56]

In addition, it constitutes a reality that the advanced technology used by decision-making systems surpasses the cognitive abilities of humans and thus presents challenges for legal frameworks to ensure complete transparency. A direct consequence of the latter is that the requirements of a criminal justice system and society that adhere to the rule of law, such as understanding, openness, impartiality and comprehensibility, are exceedingly difficult to fulfil.[57] Simultaneously, incorporating specific predictive tools from the public sector to assess and categorise individuals based on their likelihood of engaging in criminal activity or expressing antisocial behaviour undermines the fundamental principles of the traditional criminal justice systems. This novel form of pre-emptive algorithmic justice has the potential to result in the continuous surveillance of targeted individuals, extensive monitoring and the establishment of a 'control' society.[58] Within this 'control' society, actions that are perceived as posing a threat to social order may be classified as hazardous and subjected to measures of correction.

The fundamental basis for employing an actuarial approach in decision-making is the notion that historical patterns offer valuable insights into future activities and events. But in criminal justice, predicting the future on the basis of the past threatens to reify and reproduce existing inequalities of treatment by institutions.[59] As data collection engines are socially organised, people's visibility of them is influenced. This leads to digital coding becoming a new way for race, gender and class structures to operate. Hence, the excessive law enforcement in communities predominantly inhabited by people of colour, along with the systematic implementation of racial profiling, results in crime data sets that exhibit a significant over-representation of these specific groups.[60] Furthermore, areas that are anticipated to have high levels of criminal activity or individuals who are deemed to be at greater risk will receive increased police focus, resulting in a higher number of arrests and subsequently, a heightened level of scrutiny.[61] This repetitive process of examination has been described as the ratcheting effect. At the same time, disproportionate entanglement in the criminal justice system record can have social costs, including barriers to employment, education or housing, all of which can potentially result in further harm. As a result, the inherent nature of predictive analytics results can be described as a self-fulfilling prophecy, perpetuating historical

56. Stanley Greenstein, 'Preserving the Rule of Law in the Era of Artificial Intelligence' (2021) 30 Artificial Intelligence and Law <https://link.springer.com/article/10.1007/s10506-021-09294-4> accessed 1 September 2024.
57. Aziz Z Huq, 'Artificial Intelligence and the Rule of Law' (2021) Public Law and Legal Theory Working Paper Series 764 <https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=2194&context=public_law_and_legal_theory> accessed 1 September 2024.
58. Deleuze Gilles, 'Postscript on the Societies of Control' (1992) 59 October <https://www.jstor.org/stable/778828> accessed 1 December 2024.
59. Fernando Ávila, Kelly Hannah-Moffat and Paula Maurutto, 'The Seductiveness of Fairness: Is Machine Learning the Answer? – Algorithmic Fairness in Criminal Justice Systems' in Marc Schuilenburg and Rik Peeters (eds), *The Algorithmic Society* (Routledge 2020).
60. Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (St. Martin's Press 2018).
61. Cathy O' Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Penguin Books 2016).

disadvantages and structural inequities, even without any intentional discrimination.[62] In addition, the perception of fairness arises from the fact that these measurements are based on individual behaviour rather than group behaviour. Due to the mathematical equations, these systems are often regarded as more objective or progressive compared to discriminatory systems from the past. Mathematical equation is a formal expression that defines the relationship between different variables within a model and a fundamental role in AI, since they are essential for representing and solving problems, enabling AI systems to learn from data, make predictions and optimise outcomes.[63] Nevertheless, the purported objectivity of machine learning systems does not resolve the intrinsic issue of inexplicability inherent in their black-box nature. Consequently, they may be more difficult to identify and contest. Yet, it can be argued that simplifying justice (and punishment) to a computational model is also inherently problematic. Forecasting future criminal behaviour based on the analysis of data and prediction, excessively influences decision-making, prioritising it over factors such as suspicion, deterrence, intent and the extent of harm caused. This amounts to a deconstruction of our traditional understanding of criminal justice. What is also vital to emphasise is that 'the other', 'an enemy' or even 'the risky abnormal' is algorithmically produced. Anomaly detection refers to the promise of big data to compare data at a large scale and patterns and correlations that could reveal the 'needle in the haystack'.[64] The latter statement appears as another mode of anticipatory and pre-emptive justice and security, oriented in targeting, monitoring and isolating certain groups of people for the sake of 'public good' and stability.[65] In conclusion, the uncritical reliance on predictive analytics in the criminal justice system, not only can perpetuate historical disadvantages and structural inequalities but also risks distorting the very essence of justice itself, ultimately leading to the algorithmic creation of societal 'others' and their potential pre-emptive isolation and monitoring in the name of public safety.

The Artificial Intelligence Act (AIA): Uncovering loopholes and gaps

As the development and use of AI continue to expand, policymakers worldwide are increasingly confronted with the challenge of regulating the use of this technology. Although, initially, national authorities took the lead in regulating AI, recent years have witnessed the emergence of various regulatory initiatives at regional and global levels.[66] The most far-reaching international effort to

62. Kevin Bauer and Andrej Gill, 'Mirror, Mirror on the Wall: Algorithmic Assessments, Transparency, and Self-Fulfilling Prophecies' [2023] Information Systems Research <https://pubsonline.informs.org/doi/full/10.1287/isre.2023.1217> accessed 1 September 2024.
63. Masahito Ohue, Kotoyu Sasayama and Masami Takata, 'Mathematical Modeling and Problem Solving: From Fundamentals to Applications' (2024) 80 Journal of Supercomputing <https://link.springer.com/article/10.1007/s11227-024-06007-x#citeas> accessed 1 September 2024.
64. Sreenivasulu Thudumu and others, 'A Comprehensive Survey of Anomaly Detection Techniques for High Dimensional Big Data' (2020) 7 Journal of Big Data 42 <https://doi.org/10.1186/s40537-020-00320-x> accessed 1 December 2024.
65. Mike Zajko, 'Artificial Intelligence, Algorithms, and Social Inequality: Sociological Contributions to Contemporary Debates' (2022) 16(3) Sociology Compass <https://compass.onlinelibrary.wiley.com/doi/10.1111/soc4.12962> accessed 1 September 2024.
66. Giusella Finocchiaro, 'The Regulation of Artificial Intelligence' (2024) 39 AI & Society 1961, 1968 <https://doi.org/10.1007/s00146-023-01650-z> accessed 1 September 2024.

regulate the development and use of AI technology can be considered the EU AI Act, proposed by the European Commission in 2021 and entered into force in August 2024[67]. The principal objective of this Regulation, as reflected in Recital 1, is to improve the functioning of the internal market by laying down a uniform legal framework in particular for the development, the placing on the market and the use of AI systems in the Union, in accordance with the Union's value, while ensuring high-level protection for health, safety and fundamental rights.[68]

Moreover, the EU AI Act introduces a common European regulatory framework encompassing all sectors and all types of AI technology while following a risk-based methodology.[69] The risk management model is based on the classification of AI systems into three categories, depending upon the risks they entail: systems that create an unacceptable risk and therefore should be prohibited, systems that create a high risk and systems that create a low or minimal risk. In addition, the risk-based approach aims to balance innovation with fundamental rights protection by classifying systems following the risk they present, and by setting requirements according to that risk.[70] In accordance with this approach, the regulation imposes specific requirements on high-risk AI systems while leaving low or minimal-risk AI systems largely unencumbered.

The second section of the AI Act includes rules concerning the categorisation of high-risk systems and the standards they must meet. According to Annex III, high-risk AI systems are those used in sectors such as law enforcement, critical infrastructure, education, employment, migration, access to essential private services and essential public services, as well as the administration of justice.[71] For systems classified as high risks, certain requirements must be respected such as the implementation of a risk management system, data governance, transparency and human oversight.[72] As stated in Recital 27, transparency means that AI systems are developed and used in a way that allows for appropriate traceability and explainability. It also means duly informing deployers of the capabilities and limitations of that AI system and affected persons about their rights. However, achieving this level of transparency can be challenging due to the inherent opacity of AI systems. Despite efforts to make AI processes understandable, the complexity and proprietary nature[73] of many AI algorithms can obscure their inner workings.[74] This opacity poses a challenge to transparency, as it becomes difficult, or even impossible, to fully explain how decisions are made within these systems. Therefore, while transparency aims to build trust and

67. Nathalie A Smuha and Karen Yeung, 'The European Union's AI Act: Beyond Motherhood and Apple Pie?' 2024 <https://ssrn.com/abstract=4874852> accessed 1 September 2024.
68. Regulation (EU) 2024/1689 on Artificial Intelligence, (Artificial Intelligence Act), Recital 1.
69. European Commission, 'Explanatory Memorandum to the Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)' COM (2021) 206 final <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206> accessed 1 January 2025.
70. AI Act, Recital 64.
71. AI Act, Annex III.
72. AI Act, Articles 8–15.
73. Ulla-Maija Mylly, 'Transparent AI? Navigating Between Rules on Trade Secrets and Access to Information' (2023) 54 International Review of Intellectual Property and Competition Law (IIC) 1013 <https://doi.org/10.1007/s40319-023-01328-5> accessed 1 January 2025.
74. Madalina Busuioc, Deirdre Curtin and Marco Almada, 'Reclaiming Transparency: Contesting the Logics of Secrecy Within the AI Act' (2023) 2(1) European Law Open 79, 105, <https://doi.org/10.1017/elo.2022.4> accessed 1 September 2024.

accountability, the opaque nature of AI can undermine these objectives, creating a tension between the need for understandable and trustworthy AI operations and the often inscrutable mechanisms behind them. In addition, other requirements and obligations indicate that high-risk AI systems deployed by public authorities or entities acting on their behalf must be registered in a public EU database.[75] The same obligation, however, does not apply to systems used in the context of law enforcement or migration, since the latter will have to be registered in a non-public part of the database that will be only accessible to relevant supervisory authorities.[76]

In the AI framework, standards are essential components for the regulation of AI systems, similar to other product regulations. The role of standards and conformity assessments is crucial in ensuring the reliability, safety and trustworthiness of high-risk AI systems. Furthermore, Section 2 and Section 3 encompass harmonised standards that need to be applied and implemented when assessing the risk of the system.[77] More specifically, Article 40 paragraph 1, introduces the presumption of conformity, meaning that high-risk systems shall be presumed to conform to the requirements mentioned in Sections 2 and 3.[78] Finally, and most importantly, standards would also play a pivotal role in the protection of fundamental rights, as compliance with the standardised rules for high-risk systems, as outlined in Article 40 and Sections 2 and 3, is closely linked to ensuring the protection of fundamental rights. In essence, the effective implementation of standards is considered vital for safeguarding fundamental rights.[79] It is also believed that by establishing guidelines for compliance, standards help reduce the risks associated with AI technologies[80].

Standards can be understood as crucial 'information rules' for complex digital technologies, with legacies that are historically and institutionally determined.[81] The increasing importance of standardisation in the current economic and geopolitical landscape indicates that standard-setting is solidifying its position as a non-conventional power. At the same time, its political and regulatory influence can no longer be ignored.[82] The public interest aspect of standardisation reinforces its role and power as a 'transnational hybrid authority' that involves a combination of public and private actors.[83] While governments still play a central role in policy definition, standardisation broadly and standard-developing organisations, in particular, are influential private policy

75. AI Act, Article 26.
76. AI Act, Article 49 para 4.
77. AI Act, Sections 2 and 3.
78. AI Act, Article 40 para 1.
79. Marta Cantero Gamito and Christopher T Marsden, 'Artificial Intelligence Co-regulation? The Role of Standards in the EU AI Act' (2024) 32(1) International Journal of Law and Information Technology <https://doi.org/10.1093/ijlit/eaae011> accessed 1 September 2024.
80. Peter Kim, 'Auditing of AI: Legal, Ethical and Technical Approaches' (2023) 10(3) AI & Society 74 <https://link.springer.com/article/10.1007/s44206-023-00074-y> accessed 1 September 2024.
81. Paul David and Shane Greenstein, 'The Economics of Compatibility of Standards: A Survey' (1990) 1(1–2) Economics of Innovation and New Technology 3 <http://neconomides.stern.nyu.edu/networks/David-Greenstein_The_Economics_of_Compatibility_Standards.pdf> accessed 1 September 2024.
82. Tim Büthe and Walter Mattli, *The New Global Rulers: The Privatization of Regulation in the World Economy* (Princeton University Press 2011).
83. Niels ten Oever and Stefania Milan, 'The Making of International Communication Standards: Towards a Theory of Power in Standardization' (2022) 1 Journal of Standardisation <https://journals.open.tudelft.nl/jos/article/view/6205> accessed 1 September 2024.

venues.[84] As specified in Article 40 of the AI Act, the European Standardisation Organisations (ESOs) are responsible for developing standards to ensure consistent and transparent application of technical regulations across the EU.[85] Consequently, standards will set clear boundaries for the development and deployment of AI systems. Nevertheless, the extent to which this standardisation procedure can be demonstrated to be significant in safeguarding fundamental rights remains uncertain. This uncertainty is particularly pronounced given that ESOs are private, bureaucratic entities heavily influenced by industry stakeholders.[86] These stakeholders include numerous multinational corporations whose interests may not entirely align with the values of the EU. This situation raises critical concerns about the alignment of these standards with the fundamental rights and values upheld by the EU. The influence of powerful industry players could potentially skew the standards to favour commercial interests over public welfare. Consequently, the effectiveness of these standards in genuinely protecting fundamental rights remains a contentious issue, warranting ongoing debate and rigorous scrutiny.[87]

Furthermore, the AI Act can be considered as a safety product legislation, following the New Legislative Framework of the EU, since the proposal from the commission was based on the NLF architecture.[88] Similar regulatory initiatives were implemented in the EU when regulating toys, washing machines and dishwashers. This type of legislation outlines broad principles that are translated into harmonised standards, which in turn give rise to a presumption of conformity.[89] However, the explanatory memorandum that accompanied the proposal highlighted that there is strong evidence that certain AI systems can have a significant impact on all fundamental rights enshrined and recognised in the Charter.[90] Therefore, in contrast to other product legislation, the AI Act regulates products, namely AI systems, that can have a significant impact on fundamental rights, unlike a toy or a washing machine, and as a result creates a conceptual mismatch by the embedding of fundamental rights into an instrument bedded in product safety law.[91]

In addition, although many references to fundamental rights exist throughout the regulation, a sufficient mechanism or methodology for examining what a risk to fundamental rights is, and if or when a violation/interference can be justifiable, is lacking. Although a general definition of the notion of risk exists in the regulation, as the combination of the probability of an occurrence of harm and the severity of that harm,[92] in the EU fundamental rights law, rights are safeguarded as

84. Paul Moritz Wiegmann, Henk J de Vries and Knut Blind, 'Multi-Mode Standardisation: A Critical Review and a Research Agenda' (2017) 46(8) Research Policy 1370, 1386 <https://www.sciencedirect.com/science/article/pii/S0048733317301002> accessed 1 September 2024.
85. AI Act, Article 40.
86. European Digital Rights (EDRi), 'The Role of Standards and Standardisation Processes in the EU's Artificial Intelligence (AI) Act' (EDRi 2022) <https://edri.org/wp-content/uploads/2022/05/> accessed 1 September 2024.
87. Michael Veale, 'Value-Laden Areas for Standardization in the AI Act' (2022) <https://michae.lv/value-laden-areas-in-the-ai-act/> accessed 1 September 2024.
88. European Commission, 'Explanatory Memorandum' (n 69).
89. Hadrien Pouget, 'Institutional Context' Future of Life Institute (2022) <https://artificialintelligenceact.eu/context/> accessed 1 January 2025.
90. European Commission, 'Explanatory Memorandum' (n 69).
91. Michèle Finck, 'Fundamental Rights and the AI Act' (*Computers, Privacy, and Data Protection (CPDP) Conference on AI Governance, 2024*, Brussels, 3 June 2024).
92. AI Act, Article 3(2).

inherent rights, implying that any infringement on them will constitute a violation in and of itself, irrespective of whether tangible harm materialises.[93] The attempt to materialise the violation of fundamental rights raises significant questions on the conceptualisation of fundamental rights protection.[94] At the same time, the variability of the impact of AI on fundamental rights cannot be adequately captured in assessment standards. There is, therefore, a need to implement a methodology for the assessment of high-risk systems and go beyond standardisation.

Furthermore, despite the common belief that AI systems can pose significant threats to society as a whole, during the negotiations surrounding the final text of the AI Act, and in the first draft proposal, there was no reference or mention of any provision related specifically to the protection of fundamental rights.[95] Eventually, emerging from extensive negotiations and driven by strong advocacy from civil society organisations,[96] the European Parliament proposed to incorporate a Fundamental Rights Impact Assessment (FRIA) into the AI Act with the aim to address the potential impact of some AI systems on fundamental rights. More specifically, Article 27 of the Regulation clarifies that the provision for conducting a FRIA concerns high-risk AI systems.[97] This process aims to ensure that high-risk AI systems are deployed responsibly, with a clear understanding of their potential impact on fundamental rights. Nevertheless, these standards should be regarded as broad guidelines to consider when assessing the risk to fundamental rights. They do not, however, provide precise methodological criteria for accurately evaluating such risks. This lack of specificity raises concerns about their effectiveness in safeguarding fundamental rights, highlighting the need for more detailed and robust assessment frameworks. In addition, another problematic aspect in the context of criminal law can be considered that certain AI systems are not subjected to a FRIA. According to Annex III paragraph 6 of the AI Act, AI systems utilised specifically in sentencing or probation are not explicitly listed, implying that they will not be subject to a FRIA despite their significant impact on individuals' lives. More precisely, the high-risk classification includes AI systems used by law enforcement authorities for evaluating the reliability of evidence in the course of investigations or prosecutions, but it does not explicitly mention AI systems used specifically for sentencing and probation decisions.

Finally, the fifth paragraph of Article 27 indicates that the AI Office, established and created within the European Commission, is mandated to create a questionnaire, including through an automated tool, to facilitate deployers in complying with their obligations under this Article in a simplified manner.[98] However, although we would only be able to have more information about the

93. Mireille Hildebrandt, 'Law as Computation in the Era of Artificial Legal Intelligence' (2018) 68(1) University of Toronto Law Journal <https://www.jstor.org/stable/90019651> accessed 1 September 2024.
94. Alessandro Mantelero, 'The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, Legal Obligations and Key Elements for a Model Template' (2024) 54 Computer Law & Security Review <https://www.sciencedirect.com/science/article/pii/S0267364924000864> accessed 1 September 2024.
95. Alessandro Mantelero, 'Fundamental Rights Impact Assessments: Comparative Approach Between EU and Brasilian Legislation' (*Computers, Privacy, and Data Protection(CPDP) Conference on AI Governance, 2024*, Brussels, 6 June 2024).
96. Joint Statement, 'EU Artificial Intelligence Act for Fundamental Rights A Civil Society Statement' (2021) <https://www.fairtrials.org/app/uploads/2022/05/Civil-society-reacts-to-EP-AI-Act-draft-report_FINAL.pdf> accessed 1 September 2024.
97. AI Act, Article 27(1).
98. AI Act, Article 27(5).

implementation of this provision, as soon as the templates are created by the AI Office, it is believed that a solely questionnaire-based approach, especially if it is automated, might be insufficient to fully encapsulate the contextual nature of FRIA.

Notwithstanding that the AI Act does not constitute criminal law legislation, being primarily based on Articles 114 and 16 of the Treaty on the Functioning of the EU, its significance for the orientation of criminal justice and the protection of rule of law values is substantial. This importance arises from its provisions concerning the prohibition and the use of certain intrusive AI applications by the public sector, including policing and law enforcement agencies.[99] In addition, while certain uses are read as prohibitions, their prohibition is only partial and subject to many dangerous exceptions.

More specifically, despite the commitment of the European regulator to promote the protection of fundamental rights, and the rule of law,[100] the final text of the regulation was seen with great disappointment by human rights experts, and civil societies, stating that the EU had missed the opportunity to protect civil liberties.[101,102] Significant gaps and loopholes that can turn prohibitions into empty declarations exist, while the use of certain harmful discriminatory applications of AI, such as risk assessment or predictive policing tools not fully prohibited. A prominent example of a prohibition that, in practice, may be interpreted as permissible under specific circumstances and criteria is the use of real-time biometric identification systems in publicly available spaces. In Recital 32, the Regulation acknowledges that the use of this kind of AI application is particularly intrusive to the rights of individuals, to the extent that it may affect the private life of a larger population, evoking a feeling of constant surveillance.[103] For these reasons, Recital 33 indicates that the use of those systems for the purpose of law enforcement should therefore be prohibited, except in exhaustively listed cases, where its use is necessary to achieve a substantial public interest.[104] The notion of 'real time remote biometric identification system' referred to in this Regulation is defined as an AI system intended to identify natural persons without their active involvement, typically at a distance, through the comparison of a person's biometric data with the biometric data contained in a reference database, irrespectively of the particular technology, processes or types of biometric data used.[105] In addition, real-time systems involve the use of live or near-live materials. More precisely, two critical points can be raised here. First, this prohibition is not absolute and has certain limitations since the use of such technology can be considered permissible under certain circumstances. From the exhaustive list of circumstances, the one that can be considered controversial is the permission of such technology when its use is strictly designed and is necessary to achieve a substantial public interest. The vague scope of the latter term presents significant risks since issues

99. AI Act, Articles 5 and 6.
100. AI Act, Recital 1.
101. European Center for Non-Profit Law, 'Packed with Loopholes: Why the AI Act Fails to Protect Civic Space and the Rule of Law' (2024) <https://ecnl.org/news/packed-loopholes-why-ai-act-fails-protect-civic-space-and-rule-law?> accessed 1 September 2024.
102. European Disability Forum, 'EU's AI Act Fails to Set Gold Standard for Human Rights' (2024) <https://www.edf-feph.org/publications/eus-ai-act-fails-to-set-gold-standard-for-human-rights/> accessed 1 September 2024.
103. AI Act, Recital 32.
104. AI Act, Recital 33.
105. AI Act, Article 3.

of 'public interest' can be politically constructed. A second criticism of this provision is the fact that post-time remote biometric identification continues to be allowed. From a human rights perspective, it remains unclear why the rights and freedoms of data subjects are not equally implicated in the context of 'post' use of remote biometric identification systems, especially since the sole functional distinction between these two systems lies in the presence of a 'significant delay' between data capture and identification. This observation aligns with expert feedback on the draft AIA. In a joint opinion[106] issued in May 2021, the European Data Protection Supervisor and the European Data Protection Board emphasised that the intrusiveness of 'post' use of remote biometric identification systems can be just as pronounced as that of 'real-time' use. Importantly, this intrusiveness does not necessarily hinge on the temporal span during which biometric data is processed.

Furthermore, the Regulation stipulates that, in accordance with the presumption of innocence, individuals within the Union should always be evaluated based on their actual behaviour.[107] It is also acknowledged that it is essential that natural persons are not assessed solely[108] on AI-predicted behaviour, which relies on profiling, personality traits or characteristics. Instead, any judgement should be grounded in reasonable suspicion of the person's involvement in criminal activity, supported by objective and verifiable facts and subject to human assessment. While regulators and stakeholders have welcomed this prohibition as a significant development,[109] it can be argued that certain unseen harms remain unaddressed. To begin with, the word 'solely' presents significant challenges since it excludes situations where the assessment of whether a natural person constitutes a risk or not is based on 'largely', 'significantly' or 'partially'. This limitation is problematic because it permits assessments that are not entirely, but still substantially, influenced by AI predictions. Consequently, individuals could still be judged based on AI-generated profiles, which may incorporate biases and inaccuracies, thereby undermining traditional principles of criminal law and fundamental rights, such as the presumption of innocence.

Moreover, the relevant provision does not ban geographic crime prediction systems, which are widely used by police forces across Europe, despite evidence that these systems reinforce existing

106. Andrea Jelinek and Wojciech Rafał Wiewiorowski, 'Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act 2021) <https://www.edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf> accessed 1 September 2024.

107. AI Act, Recital 42.

108. In the case C-634/21 (Schufa) the ECJ elucidated that the Article 22(1) of the GDPR which grants individuals the right not to be subject to decisions based *solely* on automated processing, including profiling, that produce legal effects concerning him or her or similarly significant impacts, constitutes a general prohibition **General Data Protection Regulation**. The Court clarified that the word 'solely' also applies in scenarios where there is human participation, but without the possibility of influencing the causal link between the automated processing and the final decision (para 44). Although this ECJ ruling specifically pertains to the banking sector, it is anticipated that the same rationale will be followed and extended to other areas, where data subjects may encounter equally or even more severe repercussions from automated decision-making processes, including areas such as law enforcement and policing.

109. Lexie White, 'How Stakeholders Are Welcoming EU AI Act', International Association of Privacy Professionals (2024) <https://iapp.org/news/a/reactions-to-the-newly-passed-eu-ai-act> accessed 1 September 2024.

racism and discrimination in policing.[110] Instead, the Regulation merely prohibits risk assessment tools and predictive policing tools that rely on personal characteristics or traits. Consequently, it remains permissible to develop predictive policing systems that utilise postal codes as risk factors. When postal codes are used as risk factors, the bias inherent in socioeconomic conditions becomes embedded in the algorithm. For example, neighbourhoods with predominantly minority populations or lower socioeconomic status may be flagged as high-risk areas, leading to increased police presence and surveillance. This over-policing further perpetuates stereotypes and disproportionately impacts marginalised communities. It is imperative to recognise that structural injustices are exacerbated by predictive policing systems.

As a result, the partial ban on predictive policing and risk assessment tools, under which not all applications of AI systems in policing and 'crime prediction' are classified as unacceptable risks, can lead to a significant erosion of the presumption of innocence that constitutes a foundation of democratic liberal criminal law. Effective safeguarding of fundamental rights cannot be attained without a comprehensive and complete prohibition of such technologies. Any AI or automated system deployed by law enforcement and criminal justice authorities to predict behaviour in individuals or groups, with the aim of identifying high-risk areas and potential offenders based on historical data or geographical information (such as postal codes), can be considered contradictory to the values and democratic principles of criminal justice and will inevitably perpetuate and amplify existing discriminatory practices. In essence, relying on such predictive systems is akin to navigating a ship with a faulty compass; it may seem to guide us, but it ultimately leads us astray, reinforcing the very biases and injustices we seek to eliminate.

Lastly, the AIA contains provisions addressing the deployment of emotion recognition systems. These systems are designed to identify or infer the emotions or intentions of natural persons based on their biometric data.[111] However, it is crucial to recognise that substantial gaps exist within the Act concerning the protection of fundamental rights, leading us to serious concerns about the potential future use of these technologies. More precisely, the only explicitly prohibited practice that pertains to the use of emotion recognition is in the workplace and educational institutions,[112] while no specific analysis addresses their prohibition in law enforcement and policing contexts. Emotional facial recognition can be regarded as an emerging new form of surveillance in public urban spaces, namely what has been termed as 'emotiveillence', where people's emotional states may be subject to surveillance and their intentions may be inferred, 'usually for the purposes of influencing and managing people'.[113] A prominent study by researchers in the field of psychology concluded[114] that despite '[t]echnology companies [. . .] investing tremendous resources to figure

110. Griff Ferris, Bruno Min and Misha Nayak-Oliver, 'Automating Injustice' (Fair Trials Organisation Report 2021) <https://www.fairtrials.org/app/uploads/2021/11/Automating_Injustice.pdf> accessed 1 September 2024.
111. AI Act, Recital 18.
112. AI Act, Recital 44 and Article 5, para 1f.
113. Karen Lumsden and Aaron Doyle, 'We Have to Talk About Emotional AI and Crime' (2022) 37(4) AI & Society <https://link.springer.com/article/10.1007/s00146-022-01435-w> accessed 1 September 2024.
114. Lisa Feldman Barrett and others, 'Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements' (2019) 20(1) Psychological Science in the Public Interest <https://journals.sagepub.com/doi/10.1177/1529100619832930> accessed 1 September 2024.

out how to objectively "read" emotions in people by detecting their presumed facial expressions [. . .] the science of emotion is ill-equipped to support any of these initiatives'. These systems, therefore, rely on simplistic, inaccurate inferences and cannot reliably perform their intended function, and fail to capture diverse expressions of emotion across different cultures. The group of experts analysed more than a thousand scientific papers examining the relationship between facial expressions and inner emotional states and concluded that there is no reliable correlation between the two. Nevertheless, what is important to emphasise is that even if these technologies are proven accurate in recognising emotions and intentions, they still raise the significant question as to whether they will ever be legitimate in liberal democracies that place an emphasis on privacy and freedom of personal thoughts, feelings and emotions. Further to this, the danger in the use of such invasive surveillance for the purpose of policing and crime prevention is that it potentially leads to a highly regulated and control-oriented society. Unlike facial recognition, emotion recognition takes the notion a step further by not only undertaking surveillance of existing situations but also making inferences and probabilistic predictions about future events, as well as emotions and intentions. Finally, emotion recognition strikes at the heart of individual rights: human dignity.[115] It classifies people into arbitrary categories bearing on the most intimate aspects of their inner lives. It necessitates constant surveillance to make intrusive and arbitrary judgements about individuals. The technology's assumptions about human beings and their character endanger our rights to privacy, freedom of expression and the right against self-incrimination.

The EU has embarked on a mission to promote sustainable and trustworthy AI. However, the extent to which government regulation can ensure trust in AI remains a subject of debate. Achieving a consensus on what constitutes a truly trustworthy AI system is complex and must be integrated into the ongoing discourse. It has also been elaborated that the general discussion around trustworthiness had become 'a land of plenty',[116] with too many meanings lacking agreement on what counts for whom, which ultimately erodes ethics efficacy as a regulatory agent.[117] At the same time, related literature suggests that the conflation of trustworthiness with the acceptability of risks adopts a simplistic conseptualisation of the notion of trust.[118] Conversely, certain artificial systems can be deemed untrustworthy, raising concerns about their legitimacy within democratic criminal justice systems. Therefore, it is imperative to critically examine and address the potential loopholes in the AIA. If fundamental rights and the underlying basic values are at risk, the political system and criminal justice are also at risk. Specifically, the imminent challenge lies in preserving the

115. Anne De Hingh, 'Some Reflections on Dignity as an Alternative Legal Concept in Data Protection Regulation' (2018) 19(5) German Law Journal <https://www.cambridge.org/core/journals/german-law-journal/article/some-reflections-on-dignity-as-an-alternative-legal-concept-in-data-protection-regulation/D57FF3D530CC96663C90ADB85538E270> accessed 1 September 2024.

116. Konrad Reinhardt, 'Trust and Trustworthiness in AI Ethics' (2022) 3 AI and Ethics 735 <https://doi.org/10.1007/s43681-022-00200-5> accessed 21 December 2024.

117. Elisa Stamboliev and Tom Christiaens, 'How Empty is Trustworthy AI? A Discourse Analysis of the Ethics Guidelines of Trustworthy AI' [2024] Critical Policy Studies 1 <https://doi.org/10.1080/19460171.2024.2315431> accessed 21 December 2024.

118. Johann Laux, Sandra Wachter and Brent Mittelstadt, 'Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and Acceptability of Risk' (2024) 18 Regulation & Governance 3 <https://onlinelibrary.wiley.com/doi/full/10.1111/rego.12512> accessed 1 September 2024.

orientation and values of criminal justice since, traditionally, liberal criminal law is centred on the individual, limiting the state's power and ensuring fair trial rights. AI has the potential to disrupt these traditional values and principles by significantly undermining individual participation and comprehension of the criminal process.

Conclusion

The steady establishment of algorithmic governance and the turn towards algorithmic justice can constitute an attack on traditional normative values of criminal justice. As a direct result of the latter, the transition from narrative and database to automated algorithmic justice and algorithmic policing poses a threat to civil liberties and the democratic division of power. Safeguarding against the uncritical utilisation of algorithmic systems requires a concerted effort to uphold principles of transparency, accountability and democratic oversight when deploying these systems. The erosion of trust in artificial intelligence systems, coupled with the unseen harms and challenges of AI regulation, can significantly undermine people's confidence in the fairness of the justice system and disrupt the orientation of criminal justice. In conclusion, the increased focus on efficiency and security based on mathematical calculations should also remind us that security issues are always political decisions, while it is also essential to highlight that with the wrong decisions, big data insights can threaten democratic processes and liberal criminal justice by substituting democracy with autocracy.

Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Melina Anastasopoulou https://orcid.org/0009-0007-1690-1101