THE USE OF AI IN CRIMINAL JUSTICE: UNPACKING THE EU'S HUMAN-CENTRIC AI STRATEGY

MUSTAFA T KARAYIGIT* & DENIZ ÇELIKKAYA†

In order to mitigate concerns over potential disruptive impacts of the integration of artificial intelligence in the criminal justice system on criminal justice, this article explores the European Union's human-centric approach towards that integration, emphasising the balance to be struck between technological advancement and fundamental values and rights on the basis of legal and ethical principles. While existing literature explores AI's role in the criminal justice systems, there is a gap in examining how the EU's human-centric strategy directly shapes legal, ethical and regulatory frameworks. Based on the EU AI strategy with the aim of moderately filling this gap, this article discusses how the framework addresses ethical concerns in order to keep human's place central with safeguarded fundamental rights and values in the application of AI systems within the criminal justice system. To attain that objective, the analysis highlights the mitigation of bias and enhancement of fairness, the protection of privacy and data, the significance of human oversight, encouraging multi-stakeholder engagement and the non-substitution of human judges by automated decision-making within the framework of the EU's commitment to developing AI technologies that all serve the public good while respecting fundamental rights and values. The article contributes to the ongoing discourse on responsible AI integration into criminal justice by synthesising insights from legal, ethical and AI governance frameworks.

1 INTRODUCTION

Artificial Intelligence (AI) has become a pivotal instrument in different sectors in recent years, including criminal justice. The reason behind the incorporation of AI in the criminal justice system lies in its capability to process and analyse large volumes of data, identify patterns that may escape human perception, generate predictions based upon those patterns and offer recommendations grounded in data.[1] From predictive policing algorithms that forecast crime hotspots and facial recognition technologies that assist in suspect identification to case-law analysis, enabling a more efficient legal research process and decision drafting, the scope of AI application is massive in criminal justice.[2]

In the criminal justice systems, AI is generally used for crime prevention, crime prediction, crime analysis and recidivism risk assessment, and technologies designed by

* Professor of EU Law; Marmara University, the Institute of European Studies.
† MA in EU Law; Solicitor of the Senior Courts of England & Wales; Partner at ARC Law Firm.
[1] European Commission, 'Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on the EU Security Union Strategy' COM(2020) 605 final, 12.
[2] Fair Trials, 'Automating Injustice: The Use of Artificial Intelligence & Automated Decision-making Systems in Criminal Justice in Europe' (2021) <https://www.fairtrials.org/app/uploads/2021/11/Automating_Injustice.pdf> accessed 25 January 2025.

private companies are used especially for law enforcement.[3] Additionally, public authorities have begun to integrate surveillance data into their own systems by collaborating with private companies.[4] With the development of technology, the use of AI systems in the criminal justice field is expanding[5] and its use carries the potential to transform several aspects of the criminal justice domain, including analysing data, processing files, validating evidence, predicting criminal activity, identifying patterns and making legal decisions, and reshape the criminal judicial processes and the landscape of law enforcement. The AI-driven risk assessment tools that are now being used through complex digital evidence for unveiling insights mark a significant shift towards data-driven judicial processes.

The journey towards this AI-driven future in the criminal justice system nevertheless presents numerous ethical, legal and societal dilemmas. The adoption of AI technologies especially in the forms of machine and deep learning in criminal justice, as a sensitive field, necessitates a careful consideration of its ethical, social and legal implications and requires a precautionary approach towards their use in the criminal justice system. The European Union (EU) has been leading the effort to address these implications with its progressive policies on digital technology and fundamental rights. Its rights-driven regulatory model sets the European human-centric approach apart from market-driven United States and state-driven Chinese models.[6] Having defined its leadership in AI as 'the development and use of AI that is relevant and useful to all',[7] the EU in that respect puts human beings at the centre of AI development and regards AI primarily as a tool to maximise human well-being and prosperity. It is committed to using its resources, authority and political backing to collaborate and compete globally in the field of AI with the purpose of its development and utilisation that benefits all.[8] Its goal is to ensure that AI being created aligns with the EU founding values, in particular respect for human dignity and human rights, democracy and the rule of law, by prioritising the advantages of society and people as a whole.

Ultimately, according to the EU human-centric approach, the integration of AI into the criminal justice system must be guided by a commitment to enhance fundamental rights while protecting against potential harms. A human-centric approach provides a guidance to achieving this balance, ensuring that AI serves as a tool for justice that is equitable, just and reflective of the EU founding values. The utilisation of AI system in criminal justice could therefore be accompanied by legal safeguards and ethical values to reduce possible risks

---

[3] Asma Idder, Stephane Coulaux, 'Artificial Intelligence in Criminal Justice: invasion or revolution?' (*International Bar Association*, 13 December 2021) <https://www.ibanet.org/dec-21-ai-criminal-justice> accessed 1 June 2024.

[4] Alfred Ng, 'Amazon's helping police build a surveillance network with Ring doorbells' (*CNET*, 5 June 2019) <https://www.cnet.com/features/amazons-helping-police-build-a-surveillance-network-with-ring-doorbells> accessed 13 July 2024.

[5] Aleš Završnik, 'Criminal justice, artificial intelligence systems, and human rights' (2020) 20 ERA Forum 567.

[6] Anu Bradford, *Digital Empires - The Global Battle to Regulate Technology* (Oxford University Press 2023) 131 and 145; Sümeyye Elif Biber, 'Between Humans and Machines: Judicial Interpretation of the Automated Decision-Making Practices in the EU' (2023) University of Luxembourg Law Research Paper Series 2023-19.

[7] European Parliament, 'EU guidelines on ethics in artificial intelligence: Context and implementation', (European Parliamentary Research Service, 2019), 3 <https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf> accessed 11 February 2024.

[8] European Commission, 'Artificial Intelligence for Europe, Communication From the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions' COM(2018) 237 final; European Commission, 'Building Trust in Human-centred Artificial Intelligence' COM(2019) 168 final.

associated with utilisation of AI system in criminal justice. The role of human rights, in this sense, nevertheless serves as a protective safeguard against the misuse of AI technologies in the criminal justice domain rather than a framework for conceptualising and developing AI in alignment with human values.[9] This approach placing humans at the heart of AI and prioritising human needs and wellbeing accordingly sets the EU AI strategy apart from those of other countries with the capacity to offer valuable global lessons in terms of the use of AI in the criminal justice system.

Within that comprehension, the European Commission's High-Level Expert Group on AI states in *Ethics Guidelines for Trustworthy AI* that a trustworthy AI system must be legally, ethically and technically sound and robust.[10] The guide reiterates the core principle that the EU needs to develop a human-centric AI in accordance with its own rules and values. The EU AI strategy is therefore founded on the human-centric principles and serves to balance the benefits of AI with societal values and individual rights.[11] Moreover the EU AI Act,[12] drafted with a risk-based approach, aims to reduce errors and biases, as part of a broad initiative to develop AI in a human-centred, safe and reliable way. In that regard, it sets important requirements regarding the quality of data sets used in the development of AI systems with a focus on minimising the risks of algorithmic discrimination. It also requires certain AI systems to operate under human control in order to reduce risks in critical fields such as health, security and fundamental rights.

The integration of AI in the criminal justice system concisely creates ethical, legal and societal concerns about the disruptive impacts of AI on criminal justice arising mostly from idiosyncrasies of AI. As a sensitive field, use of algorithm in criminal justice might lead in all its phases to unjust condemnation of persons on the basis of (potentially inaccurate) crime risk assessments or even the punishment of innocent persons. In order to mitigate potential disruptive impacts of deployment of AI on criminal justice and to be able to attain a fair criminal justice on the basis of legal and ethical principles, this article argues in the footsteps of the European human-centric approach that this integration must be guided by a commitment to enhance fundamental rights and values by putting the human at the centre of the AI development/deployment for the sake of human dignity and the common well-being of humans while protecting against potential harms. In order to extract key insights from the EU AI strategy, the article accordingly aims to unpack the EU's human-centric AI strategy with its specific legal and ethical implications and influence for/on the development and application of AI in the criminal justice system. For that purpose, it adopts a qualitative legal research approach relying upon a normative legal research in order to explore legal rules and ethical principles for addressing the legal issue at stake.

---

[9] David Restrepo Amariles and Pablo Marcello Baquero, 'Promises and limits of law for a human-centric artificial intelligence' (2023) 48 Computer Law & Security Review, Article 105795.

[10] High-Level Expert Group on Artificial Intelligence (AI-HLEG), 'Ethics Guidelines for Trustworthy AI', (2019), 4 <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines.1.html> accessed 21 July 2024.

[11] Access Now, 'Mapping Regulatory Proposals for AI in Europe' (2018) <https://www.accessnow.org/wp-content/uploads/2018/11/mapping_regulatory_proposals_for_AI_in_EU.pdf> accessed 4 August 2023.

[12] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) [2024] OJ L2024/1689.

The structure of the article is as follows. It initially analyses the notion of human-centric AI and then examines under five subtitles the issues and concerns arising from the incorporation of AI in criminal justice in the light of the EU's AI strategy in order to explore implications of that strategy for criminal justice. As the reflection of main concerns to be taken into consideration in the development and deployment of AI in the criminal justice systems, implications of the human-centric approach are therefore analysed from the points of: reducing bias and enhancing fairness; ensuring transparency and accountability; safeguarding privacy and data protection; encouraging multi-stakeholder engagement; and the choice of the degree of integration of AI as a tool of assisting or replacing the human judge. It ends with general remarks.

2 THE NOTION OF HUMAN-CENTRIC AI

The concept of human-centric/centred AI has emerged as a key goal in policy papers aimed at establishing public governance of AI.[13] According to *Ethics Guidelines for Trustworthy AI*, AI systems 'need to be human-centric, resting on a commitment to their use in the service of humanity and the common good, with the goal of improving human welfare and freedom'.[14] Human-centric AI is defined in the Ethics Guidelines as an approach that 'strives to ensure that human values are central to the way in which AI systems are developed, deployed, used and monitored, by ensuring respect for fundamental rights'.[15] The cornerstone of the EU human-centred approach is the belief that AI should be developed and deployed in a manner that respects fundamental rights and human values – ultimately the EU's fundamental values enshrined in Article 2 of the Treaty on European Union (TEU) – by putting the human at the centre of the AI development and so integrating them into the lifecycle of AI development.[16] This perspective is particularly important in the criminal justice field, where the potential for AI to impact human lives is significant and where maintaining public trust and accountability and ethical values such as respect for fundamental rights, equality, transparency and accountability are paramount.[17] Ethical concerns regarding privacy and the potential de-humanisation of justice are also at the forefront of this approach, emphasising the need to balance technological innovation with respect for fundamental rights.

As stated by the High-Level Expert Group, the strategy aims to ensure that human values are at the core of the way that AI systems are to be developed, deployed, used and monitored, by respecting fundamental rights and values as well as the natural environment and other living beings as part of the human ecosystem and so by serving the public good.[18] The common foundation that unites the EU fundamental rights can be comprehended as rooted in respect for human dignity and thereby reflecting a human-centric approach enabling the human being to enjoy a unique and inalienable moral status of primacy in the

[13] Anton Sigfrids et al, 'Human-centricity in AI governance: A systemic approach' (2023) 6 Frontiers in Artificial Intelligence 2 <https://www.frontiersin.org/articles/10.3389/frai.2023.976887/full> accessed 11 February 2024.
[14] AI-HLEG (n 10) 4.
[15] ibid 37.
[16] Anna Pirozzoli, 'The Human-centric Perspective in the Regulation of Artificial Intelligence' (2024) 9 European Papers 105.
[17] AI-HLEG (n 10) 37.
[18] ibid.

all civil, political, economic and social fields.[19] Briefly, the EU human-centric approach highlights the importance of human values, rights and dignity in the development and use of AI technologies and that humans should be repositioned at the centre of AI lifecycle.[20]

On the other hand, technology does not come without a cost. The EU human-centric approach acknowledges the potential of AI to preserve or even exacerbate existing biases and introduce new forms of discrimination if not carefully designed and regulated. Transparency and accountability are also central tenets of the EU human-centric approach, addressing the complex nature of many AI systems. By prioritising fairness, transparency and accountability, a human-centric approach seeks to mitigate the risks of algorithmic bias by maintaining human oversight and control over AI systems and ensuring that AI systems do not reinforce discrimination or target vulnerable groups. In that regard, the EU AI Act emphasises accuracy, reliability, transparency, accountability, fairness and equity in developing and utilising AI applications.[21] Moreover, the EU places a high value on privacy and personal data protection, especially in the sensitive context of the criminal justice system. Additionally, while AI can help streamline certain processes, how it is used must be carefully watched and analysed so that the justice system always works effectively and in line with human values. With concerns about the de-humanisation of justice and the allocation of liability, the EU's human-centric approach suggests within the comprehension of the human-in-the-loop approach that AI should only be a tool to complement and enhance human decision-making in ways that ensure fairness and impartiality and not to be used to replace human judgment in justice systems. Overall, the EU AI strategy guides how to create a more just and effective criminal justice system by prioritising human values in technological advancements. These guiding principles are rooted in the EU foundational values such as the protection of fundamental rights, ensuring human control and supervision, maintaining technical integrity and safety, ensuring equality and fairness and promoting societal and environmental welfare.

3 IMPLICATIONS OF THE EU HUMAN-CENTRIC AI STRATEGY

It should be expressed at the outset that implementing the human-centric AI framework involves a multi-faceted approach, including legislative measures, research and innovation funding, education and training and international collaboration. The EU AI Act primarily aims to ensure the use of AI systems in the EU in accordance with EU values and promote the uptake of human centric and trustworthy AI by creating a legal framework for trustworthy AI with strict standards of transparency, security and bias mitigation. Operationalising these principles however presents significant challenges. For instance, ensuring transparency and explainability in complex AI systems is a technical challenge that requires ongoing research and innovation. These complex systems are called 'black box', referring to the difficulty of providing clear explanations of their outputs. Whilst the

[19] AI-HLEG (n 10) 10.
[20] Ozlem Ozmen Garibay et al, 'Six human-centered artificial intelligence grand challenges' (2023) 39(3) International Journal of Human-Computer Interaction 391.
[21] European Parliament Resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL)) OJ C 404/63; Recitals 27 and 59 of the EU AI Act.

technology progresses and we see more and more explainable AI models, the technical difficulty of making these systems fully explainable without sacrificing their effectiveness still remains. Additionally, there currently seems to be inverse proportion between performance and explainability in the AI systems, since the highest performing methods are the least explainable, whereas the most explainable methods are the least accurate.[22] Balancing innovation with regulation to maintain the EU's competitiveness on the global stage while safeguarding ethical standards also arises as an ongoing policy challenge. Similarly, preventing bias in AI systems necessitates continuous vigilance, diverse data sets, inclusive design processes and cross sector collaboration between ethicists, computer engineers and legal workers.

Supporters of the integration of AI systems into criminal justice argue that these systems offer a faster, fairer, more consistent and cost-effective solution to human errors, such as biased decisions, lack of up-to-date information and inconsistent reasoning, and reduction of courts' workloads.[23] However, these technologies also have possible negative effects, which require careful evaluation. For example, crime forecasting algorithms (predictive policing systems) are found to disproportionately target minority neighbourhoods, which leads to over-policing. In that respect, drawn from the EU human-centric AI strategy in the realm of criminal justice on the basis of substantial issues, the following key implications thus emerge.

3.1 REDUCING BIAS AND ENHANCING FAIRNESS

Algorithmic objectivity seems to be illusory. Discriminatory outcomes might arise from algorithms on the basis of endogenous and exogenous factors. The use of AI in criminal justice can be complicated by the fact that the data used in predictive profiling processes in particular has the potential to reflect historical biases and socio-economic inequalities. Data sets used by AI systems, which reflect the value judgments of their designers and operate essentially on the basis of generalisation, may therefore reflect societal biases and so may contain misleading information by perpetuating or even amplifying them. During the development of AI systems, the biases of human developers, regardless of malicious intent, can also produce biased results. In other words, despite the good intentions of their designers, algorithms may take an unpredictable path in reaching their goals through choices, connections, correlations, inferences and interpretations made.[24] Moreover, in terms of overall accuracy of algorithms, they naturally optimise better for the majority, at the expense of vulnerable minorities or marginalised communities.[25] Algorithms may even produce biased decisions and lead to direct or indirect discrimination not only because of replication,

[22] David Gunning et al, 'XAI - Explainable Artificial Intelligence' (2019) 37(4) Science Robotics aay7120.
[23] Wojciech Wiewiórowski and Michal Fila, 'AI and Data Protection in Judicial Cooperation in Criminal Matters' (*Eurojust*, 2022) <https://www.eurojust.europa.eu/20-years-of-eurojust/ai-and-data-protection-judicial-cooperation-criminal-matters> accessed 21 July 2024.
[24] Aleš Završnik, 'Algorithmic justice: Algorithms and big data in criminal justice settings' (2021) 18(5) European Journal of Criminology 623.
[25] Michael Kearns and Aaron Roth, *The Ethical Algorithm – The Science Of Socially Aware Algorithm Design* (Oxford University Press 2019) 78.

perpetuating or reinforcing of incorporated certain social values and existing societal biases, but also because of the reproduction of biases from input data.[26]

These biases in data can cause algorithms to produce biased results against certain demographic groups, increasing false positives or false negatives and so lead to direct or indirect discrimination due to biases (intentional or not) both in the training and operational phases. Within the context of criminal justice, AI tools such as predictive policing algorithms and decision-making aids for judges thus can inadvertently perpetuate or even increase existing biases if not carefully designed and monitored.[27] Algorithm biases thus may consolidate discrimination and impair the neutrality of judgments and the legitimacy of their use in the criminal justice system. This could lead to individuals and communities being unfairly targeted and discriminated with the consequence of hindering the equal and fair administration of justice. This situation would be exacerbated by proneness of judges to fall into judicial conformism by aligning themselves with the outcomes and recommendations generated by the algorithms.[28] Judges may also use AI technology selectively by relying more on extra-legal factors in criminal cases.[29]

Hacking and designing or reverse-engineering the decision-making processes in AI systems with the malicious intent by programmers, software engineers or information technology companies[30] with the purpose of manipulation of judgments present additional threats of the algorithmic systems to fair trial in criminal justice.

Furthermore, 'the risk assessment method yields probabilities, not certainties, and measures correlations, not causations'.[31] Machine learning provides statistical results deriving from the establishment of mere correlations and so not relying on causality as legal reasoning does.[32] Purely statistical-mathematical correlations would therefore remain unsatisfactory in meeting the standards of a reasoned decision, especially in criminal matters.[33] In that regard, AI generally operates to apply rules to the treatment of people through the use of statistical

---

[26] Kathrin Hartmann and Georg Wenzelburger, 'Uncertainty, risk and the use of algorithms in policy decisions: a case study on criminal justice in the USA' (2021) 54 Policy Sciences 269; Raphaële Xenidis and Linda Senden, 'EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination' in Ulf Bernitz et al (eds), *General Principles of EU law and the EU Digital Order* (Kluwer Law International 2020) 151-182.

[27] Anastasia Siapka, 'The Ethical and Legal Challenges of Artificial Intelligence: The EU response to biased and discriminatory AI' (Thesis, Panteion University of Athens, 2018) 14.

[28] Florence G'sell, 'AI Judges' in Larry A DiMatteo, Cristina Poncibò, and Michal Cannarsa (eds), *The Cambridge Handbook of Artificial Intelligence, Global Perspectives on Law and Ethics* (Cambridge University Press 2022) 347-363.

[29] Dovilė Barysė and Roee Sarel, 'Algorithms in the court: does it matter which part of the judicial decision-making is automated?' (2024) 32 Artificial Intelligence and Law 117.

[30] Changqing Shi, Tania Sourdin, and Bin Li, 'The Smart Court – A New Pathway to Justice in China?' (2021) 12(1) International Journal for Court Administration 4; David Freeman Engstrom, Daniel E Ho, Catherine M Sharkey, and Mariano-Florentino Cuéllar, 'Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies', Report Submitted to the Administrative Conference of the United States, February, 2020 <https://law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf> accessed 9 August 2024.

[31] Md Abdul Malek, 'Criminal courts' artificial intelligence: the way it reinforces bias and discrimination' (2022) 2 AI and Ethics 233.

[32] Juliette Lelieur et al, 'General Report' in Juliette Lelieur (ed), *Artificial Intelligence and Administration of Criminal Justice* (International Colloquium, Buenos Aires, Argentina, 28-31 March 2023) 94 Revue Internationale de Droit Pénal 11, 49.

[33] Jasper Ulenaers, 'The Impact of Artificial Intelligence on the Right to a Fair Trial: Towards a Robot Judge?' (2020) 11(2) Asian Journal of Law and Economics 1.

generalisations and so de-individualises decisions rather than assessing each individual on their own merits with the unavoidable outcome of the product of a generalisation and de-individualised assessment of the case at stake.[34] De-individualised assessment based on statistical generalisations may thus undermine the fair administration of justice by sacrificing individual justice for the sake of consistency. AI use in criminal justice may also infringe certain principles such as presumption of innocence enshrined in Article 6 of the European Convention on Human Rights (the ECHR) in the case of use of AI system for the purpose of risk assessment in the pre-trial phase.

Since the algorithm is based upon the inputs, inadequate, incomplete, inaccurate, misclassified, outdated, undiversified and biased data distort it and lead to poor performance.[35] Even though removing biased data from these systems can be thought of as a solution, it might nonetheless be challenging to determine whether the discriminatory output was caused by the data or the AI system itself.[36] For instance, if we propose that the training data should be inclusive,[37] we might be adding more variables that can lead to discrimination. On the other hand, removing too many variables that can be considered leading to discrimination can make the AI system non-functional.[38] Moreover, the call for diverse data sets in training AI models is not just about variety but also about depth and representativeness to ensure that the AI's 'learning' reflects the complexity and diversity of real-world scenarios. This is particularly crucial in criminal justice, where decisions can significantly affect not only individuals' lives, but also broader societal perceptions of fairness and justice. Continuous monitoring for biased outcomes represents an acknowledgement that AI systems are not static, but evolve and adapt over time. As such, their impacts can shift and so ongoing vigilance is necessitated to ensure that biases do not creep in or worsen as the system learns from new data. In that regard, a delicate balance as to data sets should be struck.

That is why the implementation of predictive profiling systems requires careful ethical and regulatory consideration throughout their development and use cycle with the EU's human-centred AI principle in mind. Training, validation and testing of data sets should therefore be subject to comprehensive data management and governance practices. Data sets should be evaluated for possible biases, omissions and improvements and should be representative, error-free and complete to avoid discriminatory outcomes. These data sets must lawfully represent the target audience of the AI system, including gender, ethnicity and other grounds of discrimination. Since not only would technology have legitimacy in

[34] Kate Jones, 'AI governance and human rights – Resetting the relationship' (January 2023) Chatham House Research Paper, International Law Programme, <https://www.chathamhouse.org/sites/default/files/2023-01/2023-01-10-AI-governance-human-rights-jones.pdf> accessed 10 April 2024; Laura Notaro, 'Predictive Algorithms and Criminal Justice: A Synthetic Overview from An Italian and European Perspective' (2020) 2 Roma Tre Law Review 49.

[35] Brandon L Garrett and Cynthia Rudin, 'The Right to A Glass Box: Rethinking the Use of Artificial Intelligence in Criminal Justice' (*SSRN*, 22 November 2022) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4275661#> accessed 25 August 2024.

[36] Fair Trials (n 2) 30.

[37] Lana Bubalo's lecture about Legal protection against discrimination by AI on GDHRNet Training school 'Human Rights and Artificial Intelligence', held in Kuressaare, Saaremaa, remotely on 7 July 2023.

[38] ibid.

correlation with the degree of scientific progress and objectiveness,[39] but also be truly human-centred in accordance with the principles of social justice, AI governance must look beyond the technical aspects of AI technology, respond to the pre-existing societal structures breeding algorithmic biases and remedy them.[40]

The EU AI Act in that regard includes a multifaceted approach aimed at reducing the risk of inaccurate or biased decisions made by AI in critical areas such as criminal justice. The Act imposes obligations to minimise algorithmic discrimination by focusing on the quality of the data sets used during the development of AI systems. This approach will be applied throughout the entire lifecycle of AI systems, namely testing, risk management, documentation and human oversight. Moreover, the Act introduces comprehensive regulation for the use of 'real-time' biometric recognition systems in public spaces. Rather than a blanket ban, these systems are allowed to be used under certain situations and conditions, for instance to identify certain victims of crime, prevent certain threats or find specific criminals. Such uses must comply with the legal framework, be approved in advance by a judicial or administrative authority and comply with detailed guidelines in the legislation of the Member States.[41] Lastly, according to Recital 42 of the EU AI Act crime risk assessments based solely on profiling natural persons or on assessing their personality traits and characteristics should be prohibited and so

> [n]atural persons should never be judged on AI-predicted behaviour based solely on their profiling, personality traits or characteristics, such as nationality, place of birth, place of residence, number of children, level of debt or type of car, without a reasonable suspicion of that person being involved in a criminal activity based on objective verifiable facts and without human assessment thereof.

Ultimately, mitigating bias is strongly correlated with other pillars of the human-centric AI model. Change in laws and regulations could force algorithms to be more transparent, accountable and effective tools subject to human oversight for identifying and preventing bias.[42] This strategy not only advocates for mechanisms that ensure transparent and accountable AI systems, but also emphasises the importance of human values and ethical considerations embedded at every stage of AI development and deployment. By integrating human oversight with efforts to minimise bias, the EU is charting a path toward AI application in criminal justice that are not only technologically advanced but also deeply aligned with societal values and fundamental rights. This holistic approach accordingly serves as a guiding principle for leveraging AI to enhance justice and equity, while vigilantly guarding against the perpetuation of existing disparities.

One of the critical implications arising from the EU AI strategy is thus the emphasis on reducing bias and enhancing fairness in AI systems. This emphasis is pivotal especially

---

[39] Stanley Greenstein, 'Preserving the rule of law in the era of artificial intelligence (AI)' (2022) 30 Artificial Intelligence and Law 291.

[40] Karine Gentelet and Sarit K Mizrahi, 'A Human-Centered Approach to AI Governance: Operationalizing Human Rights through Citizen Participation' in Catherine Régis et al (eds), *Human-Centered AIA Multidisciplinary Perspective for Policy-Makers, Auditors, and Users* (CRC Press 2024).

[41] Article 5 of the EU AI Act.

[42] Bruno Lepri, Nuria Oliver, and Alex Pentland 'Ethical machines: The human-centric use of artificial intelligence' (2021) 24(3) iScience, Article 102249.

when considering the profound impact AI systems can have within the criminal justice sector. The potential for AI to either uphold or undermine justice is based on its design and application which necessitates a rigorous framework for its ethical use for positive results. The EU approach auspiciously goes beyond mere technical adjustments by advocating for a systemic integration of ethical principles throughout the AI development lifecycle. The EU advocates for the development of AI systems that are transparent and include mechanisms to identify and mitigate biases. This strategy involves diverse data sets for training AI models, continuous monitoring for biased outcomes and the inclusion of human oversight in AI-assisted decisions. A system of AI vigilance could accordingly be constructed to entail the systematic flaws in the system operations in terms of the protection of fundamental rights to be monitored and reported by stakeholders and so to trigger an obligation on the system designer to review, reassess and modify the design and operation of the system.[43]

The EU's stance on the use of AI in criminal justice, rooted in reducing bias and enhancing fairness, reflects a broader commitment to ensuring that technological advancements contribute positively to society. Incorporating human oversight into AI-assisted decisions in criminal justice serves multiple purposes. It not only acts as a safeguard against the uncritical acceptance of AI recommendations but also ensures that the nuanced and context-specific judgments that are often required in legal settings are preserved. Article 8a of Annex III of the EU AI Act appropriately qualifies in the administration of justice 'AI systems intended to be used by a judicial authority or on their behalf to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts, or to be used in a similar way in alternative dispute resolution' as high-risk AI systems. However, the EU has missed an important step here. Human rights impact assessments carried out on high-risk systems as an obligation for deployers under Articles 26 and 27 of the Act are restricted to certain areas such as AI use in public organisations and credit scoring, but do not cover all high-risk systems. The EU may nonetheless monitor the gradual implementation of the Act and expand its scope of application. However, some AI systems that we cannot fit into certain categories will be excluded from this human rights impact assessment, which may cause some AI solutions to slip under the radar. Although there are some missteps, such as not forcing all developers and deployers to implement human rights impact assessments, the EU approach generally highlights the necessity of a multidisciplinary approach to AI development, involving legal experts, ethicists, technologists and the wider community to create a criminal justice system that is not only technologically advanced but also socially responsible and just.

3.2 ENSURING TRANSPARENCY AND ACCOUNTABILITY

In the criminal justice system, where decisions can profoundly affect fundamental rights and freedoms, it is crucial that AI-assisted processes are transparent and those responsible for these systems are held accountable.[44] Non-transparent AI systems impede the detection of

[43] Karen Yeung, Andrew Howes, and Ganna Pogrebna, 'AI Governance by Human Rights-Centred Design, Deliberation and Oversight: An End to Ethics Washing' in Markus D Dubber, Frank Pasquale, and Sunit Das (eds), *The Oxford Handbook of AI Ethics* (Oxford University Press 2020) 76-106.
[44] AI-HLEG (n 10).

discrimination, the fact of which also prevents accountability.[45] Transparency and accountability arise as pillars of the EU human-centric AI approach. The EU AI strategy encourages the use of explainable AI, where the decision-making processes of AI systems can be understood and scrutinised by humans and the responsibility behind the decision made or supported by algorithms can be clarified.[46] The strategy also calls for transparency in data handling practices, ensuring that individuals are informed about how their data is used, stored and protected.[47] This transparency is crucial for maintaining public trust, especially in high-stakes domains like criminal justice, where the implications of data misuse can be profound. Transparency is therefore vital for building trust in AI systems and ensuring that they are used ethically and responsibly.

The emphasis on transparency and accountability in the EU AI framework is a recognition of the need for clarity in how AI systems make decisions, especially in the critical context of criminal justice. Opacity of AI system makes detection of shortcomings in the system and understanding the legal reasons underlying judicial decisions difficult. Explainable AI ensures that the rationale behind AI-driven decisions can be examined the fact of which accordingly may offer insights into the factors and data that influence outcomes. This level of transparency is essential for fostering an environment where AI's contributions to justice are not only recognised but also critically evaluated for fairness and integrity.[48] In essence, the EU's focus on transparency and accountability in AI applications within criminal justice is about ensuring that these powerful tools are developed and used in a manner that respects human dignity, human rights and democratic values. It is about creating a foundation of trust and ethical assurance, where AI's benefits are maximised and whose challenges are addressed with vigilance and a commitment to justice and equity.

The obligation to lay down the foundations behind the decision-making, especially when it comes to judicial decisions, is a principle that is established by the courts in many countries. The constitutional duty to provide reasons for judicial decisions taken place in the constitutional traditions of the Member States is also enshrined in Article 36 of Protocol (No 3) on the Statute of the Court of Justice of the EU (the CJEU), according to which '[j]udgments shall state the reasons on which they are based'. This Article has been upheld by the CJEU on various occasions as obliging that judgments shall give reasons upon which they are based. For instance the obligation laid down in Article 296 of the Treaty on the Functioning of the EU (the TFEU) and Article 36 of the Protocol and incumbent upon the General Court to state reasons for its judgments, as an essential procedural requirement, enables the persons concerned to understand the grounds of its judgment and provides the CJEU with sufficient information to exercise its powers of review on appeal.[49] Moreover according to the European Court of Human Rights (ECtHR), the general principles

[45] Lepri, Oliver, and Pentland (n 42).
[46] Ibid.
[47] CEPEJ, 'European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment' (Ethical Charter, Council of Europe, 2018), 25 <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c> accessed 01 March 2025.
[48] David Leslie, 'Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector' (The Alan Turing Institute, 2019), 39-40 <https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf> accessed 10 July 2024.
[49] Case C-486/15 P *European Commission* v *French Republic* EU:C:2016:912 paras 79-80; Case C-54/20 P *European Commission* v *Stefano Missir Mamachi di Lusignano* EU:C:2022:349 paras. 69-70.

concerning the right to a reasoned judgment and the corollary duty to give reasons oblige the courts and tribunals to provide for their judgments adequately stating the reasons on which they are based and presuppose that parties to judicial proceedings can expect to receive a specific and explicit reply and explanation to their arguments which are decisive for the outcome of those proceedings.[50]

Accountability extends beyond the technical aspects of AI systems to encompass the ethical responsibilities of those who design, deploy and manage these technologies. Transparency enables not only explainability, but also auditing. Third-party auditing thus may help to enhance trust in algorithms.[51] In that regard, the EU AI Act requires human oversight, especially in high-risk AI systems. It is therefore aimed to minimise risks in certain areas and ensure that the operations of the systems are sufficiently transparent so that users understand the system outputs and use them correctly. These requirements aim to contribute to respect for fundamental rights by ensuring transparency and traceability of the entire path to outcomes throughout the lifecycle of AI systems. It involves establishing clear lines of responsibility for AI's actions and decisions with the aim of ensuring that there are mechanisms in place for redress when AI systems cause harm to fundamental rights or operate contrary to ethical or legal standards. This aspect of the EU's AI strategy therefore aims to cultivate a culture of responsibility among AI practitioners that reinforces the principle that innovation should not come at the expense of ethical conduct or societal values. However, not facilitating the protection it expected to set, the Act places an additional burden on the citizens stating that if an individual wants to challenge the deployment of an AI system, he/she needs to prove individual harm.[52] This burden on individuals has the potential to restrict the public oversight of the societal impact of AI systems and so accountability.

Furthermore, the call for transparency and accountability aligns with broader efforts to demystify AI technologies by making them more accessible and understandable to the public and stakeholders within the criminal justice system. This democratisation of AI knowledge is pivotal for inclusive dialogue on AI's role in society, encourages diverse perspectives and fosters collaborative efforts to harness AI's potential while mitigating its risks. That is especially significant, since black box systems constantly underperform and conceal errors.[53] It is a fact that machine learning algorithms may rely upon assumptions about relationships of various categories of data which might remain hidden even to the designers of those AI systems.[54] In other words, AI, using especially machine learning, is too complex and inscrutable to fully understand even for the engineers who create it.[55] The black box nature of algorithms due to its complexity, lack of expertise by the system users/stakeholders or legal constructions associated with intellectual property rights[56] (business secret protection), which does not allow revelation of the algorithm even to

[50] *Zayidov v Azerbaijan* (No. 2) App no 5386/10 (ECtHR, 24 March 2022) para 91; *Çetinkaya v Türkiye* App no 76619/11 (ECtHR, 16 January 2024) para 18.
[51] Završnik, 'Algorithmic justice' (n 24).
[52] Leslie (n 48) 39-40.
[53] Garrett and Rudin (n 35).
[54] Kia Rahnama, 'Science and Ethics of Algorithms in the Courtroom' (2019) 1 Journal of Law, Technology & Policy 169.
[55] Jumpei Komoda, 'Designing AI for Courts' (2023) 29(3) Richmond Journal of Law & Technology 145.
[56] Greenstein (n 39).

prosecutors and judges,[57] and the lack of transparency make extremely difficult to discern whether the judicial decision is fair and unbiased and even to appeal decisions made by AI systems or with their assistance.[58] This poses also the risk of privatisation of justice because of the fact that AI systems designed by private companies endanger the role of lawmakers in criminal law.[59] The possibility of disclosure of the algorithm contrarily carries a risk that the algorithmic system could be manipulated and reverse-engineered by adversaries for the purpose of opposite outcomes.[60] For the human-centric AI system, prevalence of the rights of defendants should nevertheless be provided over the protection of interests of private companies in the preclusion of disclosure of their trade secrets.[61] As a consequence, not only would users and operators generally not be exactly aware of how the algorithm works and reaches its decision, but also the legal reasoning and justification behind a judicial decision may not always be transparent, which accordingly would lead to the deprivation of the capability of defendants to question a decision's accuracy and legality with the consequence of upsetting the very logic of adversarial proceedings and the undue influence on justice.[62] In order to establish this superiority and so strike a balance in favour of data subjects against intellectual property rights of programmers, the General Data Protection Regulation (GDPR) obliges that, though the right to explanation should not adversely affect trade secrets or intellectual property, the result nonetheless should not be a refusal to provide all information to data subjects.[63]

Lastly, regarding uncertainties around the EU AI Act and its application, there are no standards yet concerning compliance with the Act. The European Commission asked CEN/CENELEC to create European standards for compliance with the Act, which the providers of the high-risk systems will have to insert a CE marking showing their compliance according to Articles 43 and 48 of the EU AI Act. Although there is still time before the Act is implemented, some organisations are eager to start their compliance, as there are uncertainties with how the Act will be implemented. In that respect, there are some international standards which could be a starting point for some organisations trying to determine their risks when it comes to AI. For example, ISO/IEC 42001:2013

> is an international standard that specifies requirements for establishing, implementing, maintaining, and continually improving an Artificial Intelligence Management System (AIMS) within organizations. It is designed for entities

[57] Komoda (n 55).
[58] Taylor Brodsky, 'Artificial Intelligence in the Criminal Justice System: The Ethical Implications of Lawyers Using AI' (2023) Hofstra Law Student Works 25.
[59] Lelieur et al (n 32) 49-50.
[60] Komoda (n 55).
[61] Mirko Bagaric et al, 'The Solution to the Pervasive Bias and Discrimination in the Criminal Justice System: Transparent and Fair Artificial Intelligence' (2022) 59(1) American Criminal Law Review 95.
[62] Sergio Carrera, Valsamis Mitsilegas, and Marco Stefan, 'Criminal Justice, Fundamental Rights and the Rule of law in the Digital Age – Report of CEPS and QMUL Task Force' (Centre for European Policy Studies (CEPS) Brussels, May 2021).
[63] Recital 63 of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L 119/1.

providing or utilizing AI-based products or services, ensuring responsible development and use of AI systems.[64]

For the current compliance practices, as it was stressed by the EU AI Office, current ISO standards lack very important aspects of the Act.[65] Current ISO standards, especially ISO 42001, ISO 31000 and ISO 23894, are not sufficient for regulatory compliance with the risk management approach under the Act. Given that they focus more on company policies and documentation, the requirements of the Act and the human-centred approach to transparency, human oversight, accountability, bias mitigation and continuous and comprehensive post-market monitoring frameworks are missing in those standards. This means that, until standards are published by CEN/CENELEC, organisations which are implementing ISO standards need to supplement them with additional controls and practices that address the EU AI Act's specific requirements, especially when it comes to transparency and accountability, in order to align themselves with the Act.

3.3 SAFEGUARDING PRIVACY AND DATA PROTECTION

The collection, processing, analysing and retention of biometric data from a variety of sources through AI systems such as predictive policing, facial recognition or probabilistic genotyping DNA, the security of stored data and duration of data storage all might create deep concerns about the right to privacy and data protection. In particular, while aiming to ensure public security, use of surveillance technologies such as public surveillance cameras, license plate recognition systems or social media platforms for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including data-driven predictive policing/justice in law enforcement, might pose risks to fundamental rights, in particular the right to privacy.

The integration of AI into the criminal justice system, with its inherent reliance on vast amounts of data, therefore makes the safeguarding of privacy and data protection a critical concern as well. In criminal justice, where sensitive personal data is often involved, safeguarding privacy is paramount. 'AI […] has an impact on the entire fabric of society'.[66] Given that modern justice universally tends to be a data-oriented justice,[67] the significance of respect for privacy and data protection escalates especially with the development of the technology-driven and network society and digitalisation.

The right to privacy is enshrined in Article 8 of the ECHR and Article 7 of the Charter of the Fundamental Rights of the EU (Charter), while the right to the protection of personal data is enshrined in Article 16(1) of the TFEU and Article 8 of the Charter. Privacy is interrelated to physical, psychological or moral integrity, personal identity, development,

---

[64] ISO/IEC 42001:2023 - Information technology — Artificial intelligence — Management system <https://www.iso.org/standard/81230.html> accessed 17 July 2024.

[65] The European AI Office, 'Webinar on the risk management logic of the Act and related standards' (30 May 2024) <https://digital-strategy.ec.europa.eu/en/events/1st-european-ai-office-webinar-risk-management-logic-ai-act-and-related-standards> accessed 17 July 2024.

[66] Catelijne Muller, 'The Impact of Artificial Intelligence on Human Rights, Democracy and the Rule of Law' (Ad Hoc Committee on Artificial Intelligence, Council of Europe, Strasbourg, 24 June 2020) CAHAI(2020)06-fin.

[67] Pilar Martín Ríos, 'Predictive algorithms and criminal justice: expectations, challenges and a particular view of the Spanish VioGén system' (2024) 2/2024 Rivista italiana di informatica e diritto 547.

autonomy, the right to be forgotten, the right not to be the subject of solely automated decision-making and, as being its origin, to human dignity.[68]

The EU's strong stance on data protection and privacy, as also evidenced by the GDPR, extends to its AI strategy. The protection of personal data for the purposes of criminal matters is the subject of a specific Union legal act, namely Law Enforcement Directive (LED).[69] Article 6 and Recital 31 of the LED make a clear distinction between personal data of different categories of data subjects such as suspects, persons convicted of a criminal offence, victims, witnesses, persons possessing relevant information or contacts, associates of suspects and convicted criminals.

The EU's framework emphasises the importance of secure and ethical data handling practices, ensuring that the use of AI respects individuals' privacy rights and complies with data protection laws.[70] Under Article 10 of the GDPR, processing of personal data relating to criminal convictions and offences or related security measures shall be carried out only under the control of official authority or when the processing is authorised by Union or national law providing for appropriate safeguards for the rights/freedoms of data subjects. According to Recital 27 of the EU AI Act, AI systems shall be 'developed and used in accordance with privacy and data protection rules, while processing data that meets high standards in terms of quality and integrity'. According to Recital 59 of the EU AI Act high-risk AI systems should include AI systems intended to be used by or on behalf of law enforcement authorities or in support of law enforcement authorities for assessing the risk of natural persons to become a victim of criminal offences, for the evaluation of the reliability of evidence in the course of investigation or prosecution of criminal offences and for crime risk assessing not solely on the basis of the profiling of natural persons or the assessment of personality traits and characteristics or their past criminal behaviour for profiling in the course of detection, investigation or prosecution of criminal offences.

According to Recital 69 of the EU AI Act, those rights shall be guaranteed throughout the entire lifecycle of the AI system and so the principles of data minimisation and data protection by design and by default are applicable when personal data are processed and not only are measures of anonymisation and encryption taken, but also the use of technology is carried out without the transmission between parties or copying of data. The EU AI strategy underlines the need for robust encryption and anonymisation techniques to protect data integrity and confidentiality. Recital 53 of the LED emphasises the use of pseudonymisation as a tool that could facilitate also the free flow of personal data within the area of freedom, security and justice. This is particularly vital in criminal justice applications, where data breaches could have severe repercussions for individuals' privacy and the broader integrity

---

[68] Özgür Heval Çınar, 'The current case law of the European Court of Human Rights on privacy: challenges in the digital age' (2021) 25(1) The International Journal of Human Rights 26; Andrej Krištofík, 'The Role of Privacy in the Establishment of the Right Not to Be Subject to Automated Decision-Making' (2024) 2/2024 TLQ 236.

[69] Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA [2016] OJ L119/89.

[70] MSI-NET, 'Algorithms and Human Rights: Study on the human rights dimension of automated data processing techniques and possible regulatory implications' (2017), 12 <https://rm.coe.int/%20algorithms-and-human-rights-en-rev/16807956b5> accessed 06 July 2023.

of the justice system.[71] As declared by the Commission, a significant part of investigations against crime and terrorism involve encrypted information. Encryption, which is essential to the digital world, on the one hand secures digital systems and transactions and protects certain fundamental rights, in particular privacy and data protection, and on the other hand, if used for criminal purposes, may mask the identity of criminals and hide the content of their communications. In that regard, while combating crime and terrorism, balanced technical, operational and legal solutions to those challenges to maintain the effectiveness of encryption in protecting privacy and security of communications should be provided.[72]

According to Recital 94 of the EU AI Act, any processing of personal (biometric) data needs to respect the principles of data minimisation, purpose limitation, accuracy and storage limitation. Under Article 4(1) of the LED, personal data shall be processed lawfully and fairly, collected for specified, explicit and legitimate purposes not to be processed in an incompatible manner with those purposes, adequate, relevant and not excessive in relation to those purposes, accurate and kept up to date, ensured that inaccurate personal data are erased or rectified without delay, kept in a form which permits identification of data subjects for no longer than is necessary for those purposes and processed in a manner that ensures appropriate security of the personal data with protection against unauthorised or unlawful processing and against accidental loss, destruction or damage. According to Recital 47 of the LED, natural persons should have the right to have their inaccurate personal data rectified and the right to erasure where the processing of such data infringes the LED. Under Article 5 of the LED appropriate time limits are to be established for the erasure of personal data or for a periodic review of the need for the storage of personal data whose observation is to be ensured through procedural measures.

Furthermore, cases of AI systems wrongfully flagging individuals based on biased training data demonstrate the need for enhanced oversight and transparency. The EU emphasises the need for accountability mechanisms in data processing within AI systems to ensure that entities handling data can demonstrate compliance with privacy and data protection standards. Implementing legal accountability mechanisms is crucial for addressing any misuse of personal data. Deployment of AI technologies must be subject to regular audits, data protection impact assessments and transparent reporting to ensure compliance with the principles of privacy and data protection to ensure that the human-centric AI strategy aims to actively implement mechanisms to counteract AI-driven privacy and data protection infringements.

In essence, the EU's emphasis on privacy and data protection within its AI strategy reflects a comprehensive approach to ensuring that the deployment of AI in criminal justice not only enhances efficiency and effectiveness, but also rigorously protects individuals' rights and maintains the ethical integrity of the justice system.

3.4 ENCOURAGING MULTI-STAKEHOLDER ENGAGEMENT

The development and deployment of AI in criminal justice, according to the EU approach, should not be left solely to technologists or law enforcement agencies. It requires

[71] CEPEJ (n 47) 25.
[72] Commission, 'Communication on the EU Security Union Strategy' (n 1).

a multi-stakeholder engagement, including lawyers, legal academics, bar associations, legal ethicists, civil society organisations and the general public.[73] A lack of inclusive dialogue could lead to biased AI frameworks, democratic deficits and reduced public trust in AI-driven criminal justice systems. This inclusive approach thus helps to ensure that AI tools are developed with a broad perspective, considering various ethical, social, democratic and legal implications. By involving a wide array of stakeholders, the strategy aims to capture the complexity of ethical, legal, democratic and social dimensions that AI technologies intersect with, especially in sensitive areas such as criminal justice. Devising and ensuring that the principles of transparency, explainability and accountability are respected along the entire algorithmic design chain also requires a holistic multidisciplinary approach in the criminal justice system in which all stakeholders such as computer scientists, lawyers and social scientists, psychologists, sociologists, philosophers, etc. will have to join forces.[74] Engagement should include active stakeholder participation in AI system evaluations, policy development and ongoing monitoring to ensure that the AI systems operate within ethical and legal constraints. The EU's emphasis on multi-stakeholder engagement within the context of use of AI in criminal justice is therefore grounded in the understanding that diverse perspectives enrich the development process and lead to more equitable and effective solutions.[75]

This collaborative approach also facilitates a more transparent AI development process, where decisions are made openly and with the consideration of public interest. It encourages the co-creation of AI solutions, where stakeholders can contribute their expertise and insights, which would lead to more robust, fair and socially beneficial AI systems. Furthermore, multi-stakeholder engagement in AI development helps in identifying and addressing potential risks and unintended consequences early in the process. It ensures that safeguards and corrective measures are integrated into AI systems from the outset rather than as afterthoughts. Multi-stakeholder engagement alone is not however sufficient. AI decision-making in criminal justice must also address power imbalances between stakeholders. Law enforcement and private tech companies often hold disproportionate influence over AI policy development, which may lead to bias in or influence on regulatory decisions. To counteract this, civil society organisations must be granted greater access to AI evaluation processes, impact assessments and regulatory discussions. Diverse stakeholder representation, balanced stakeholder engagement and multi-stakeholder collaboration not only feed regulatory frameworks and public trust and foster greater transparency and ethical/legal oversight, but also enhance accountability in AI development and responsibility.[76] Consultation and collaboration with stakeholders may accordingly enhance in the end the legitimacy of use of AI in the criminal justice system.

Wide range stakeholder involvement, as being an essential aspect of the EU human-centric approach, in designing, deploying and developing (trustworthy and robust) AI systems in criminal justice, accordingly provides for meaningful input and deliberation from various components of the criminal law society and so ensures reflection of human

[73] Leslie (n 48) 3.
[74] Xenidis and Senden (n 26).
[75] ibid.
[76] Dimitrios Sargiotis, 'Fostering Ethical and Inclusive AI: A Human-Centric Paradigm for Social Impact' (2025) 6(1) International Journal of Research Publication and Reviews 3754.

element in its social context, balancing of interests and concerns of divergent components of the society, keeping the notion of criminal justice along with the evolving society and its values, mitigating concerns, promoting public awareness, building public trust, positive contribution of the integration of AI system in criminal justice to the society and in the end consolidating the legitimacy of the criminal justice system.

Briefly, the EU's call for multi-stakeholder engagement in the development and deployment of AI in criminal justice reflects a commitment to democratic, inclusive and responsible innovation. This approach not only enhances the legitimacy and effectiveness of AI applications in criminal justice but also aligns with broader societal values and the principle of good governance.

3.5 IS AI ASSISTING OR DECENTRING/REPLACING THE HUMAN JUDGE?

It is crucial to determine which tasks and to what extent they could be delegated to AI in the administration of justice, in particular to automated decision-making. In that respect it is significant under the primary question of whether certain judicial decisions should be made subject to automated decision-making or whether algorithms should merely support the decision-making process in criminal justice. While AI applications can streamline legal research, automate case law analysis and provide risk assessments and automated recommendations/decisions in helpful way to human judges, let alone fully automated judicial decision-making, even in the form of AI integration in assistance to human judge decision-making there arise significant concerns. In that respect accuracy/reliability of AI-generated evidence, overreliance on automation, inappropriate trust in AI outputs/recommendations affecting discretion of human judges, automation bias (discrimination) especially in recidivism risk assessments, de-individualisation/standardisation[77] and dehumanisation[78] of (criminal) justice, opacity preventing defence and then appeal, openness of the system to malicious reverse engineering and manipulation,[79] the certain loss of human control/oversight and the erosion of judicial independence and impartiality come to forefront.

The following three factors contributing to automation bias should be taken into consideration when determining the appropriate degree of delegation of decision-making to any AI system in the form of AI integration in assistance to human judges:

1) Under the cognitive miser hypothesis, there is a tendency of humans to choose the path of the least cognitive effort and so adhere to what the algorithm decides by relying on automated decisions, even when they suspect malfunction, and by following directives or suggestions of automated decision-making systems as a strong decision-making heuristic;

---

[77] Giulia Gentile, 'Artificial Intelligence and the Crises of Judicial Power: (Not) Cutting the Gordian Knot?' (*SSRN*, 22 February 2024) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4731231> accessed 10 July 2024.

[78] Jiahui Shi, 'Artificial Intelligence, Algorithms and Sentencing in Chinese Criminal Justice: Problems and Solutions' (2022) 33 Criminal Law Forum 121.

[79] Engstrom, Ho, Sharkey, and Cuéllar (n 30).

2) There may arise humans' perceived trust of automated decision-making systems as with superior and outperforming analytical capabilities by overestimating their performance and ascribing them greater capability and authority than humans;
3) When sharing decision-making tasks with machines, humans may feel less responsible for the outcome as a result of diffusion of responsibility and may reduce their own effort in analysing and monitoring the data available.[80]

As regards fundamental rights, there are certain risks regarding the integration of AI in criminal justice. As mentioned above, there is a risk of alterations of the system or intrusions/interventions on the algorithm/data aimed at manipulating the system and influencing the judicial decision-making process.[81] Moreover, automated judicial decision-making would also amount to turning criminal law and criminal justice over to technocrats and experts by making it less sensitive to popular emotion and more sensitive to expertise and would thus transform 'criminal law from the public re-enactment of a society's moral habitus into the coldly calculating work of minimising net social harm'.[82] Given that data is in fact contextual and spatio-temporal and that the meaning of data is dependent upon the context in which it is used and variable according to the situation with the course of time, bias can creep into data through context to lead to unfair outcomes where contextual data or algorithmic systems being developed for one context are used in another.[83]

When it comes to risks arising from solely automated judicial decision-making in the criminal justice system, empathetic human judges equipped with emotional rationality to understand human beings having motivations, intentions and goals by relying upon their intuitive experiences[84] should thus be preferred to executory cold-blooded algorithmic machines.[85] Without human involvement, AI would be unable to replicate contextual notions of fairness.[86] The removal of humans may also remove human virtues, such as human discretion and judgment, empathy, conscience and intuition, from the criminal justice system.[87] This is because current algorithms either screen out value issues or interpret them as factual issues and are unable to accommodate value judgments. Thus, they may produce justice only on a formal level without dealing with the substantive legal questions.[88]

Secondly, automated decision-making may also risk de-humanising the court experience with the consequence of standardised justice under the auspices of computational law.[89] Given that the human judge constitutes an integral part of judicial decision-making, de-humanised justice might arise in cases where a human might delegate responsibility to an

[80] Willem H Gravett, 'Judicial Decision-Making in the Age of Artificial Intelligence' in Henrique Sousa Antunes et al (eds), *Multidisciplinary Perspectives on Artificial Intelligence and the Law* (Springer 2024) 291.
[81] Notaro (n 34).
[82] Vincent Chiao, 'Fairness, accountability and transparency: notes on algorithmic decision-making in criminal justice' (2019) 15(2) International Journal of Law in Context 126.
[83] Greenstein (n 39).
[84] Nina Peršak, 'Automated Justice and its Limits: Irreplaceable Human(E) Dimensions of Criminal Justice' in Gert Vermeulen, Nina Peršak, and Nicola Recchia (eds), *Artificial Intelligence, Big Data and Automated Decision-Making in Criminal Justice* (2021) 92/1 Revue Internationale de Droit Pénal 13.
[85] Završnik, 'Algorithmic justice' (n 24).
[86] Jones (n 34).
[87] Karen Yeung, 'Why worry about decision-making by machine?' in Karen Yeung and Martin Lodge (eds), *Algorithmic Regulation* (Oxford University Press 2019) 29.
[88] Shi (n 78).
[89] Gentile (n 77).

AI decision-support system or where AI system is designed not to have any human involvement in decision-making.[90] Automated decision-making offers an aura of objectivity or de-subjectivation, replaces subjectivity and the case-specific narrative and curtails the discretion of the practitioners.[91] Algorithms, which are not completely free of biases/prejudices, might draw inappropriate or offensive inferences[92] and thus lead not only to indirect discrimination, as generally regarded, but also to direct discrimination.[93] Due to liability and responsibility concerns, decision-making processes should not be automated, and decisions should be taken by persons capable of carrying responsibility and liability which are strongly related to the exercise of discretion in reaching those decisions.

Thirdly, as regards processes of case law analysis, legal research and decision drafting, quantitative legal analysis operates by identifying the most probable outcome out of past decisions and so makes tentative moves in operation toward the common law tradition, albeit on the strict basis of *stare decisis*, by linking future case law to past case law rather than the civil law tradition.[94] What happens in situations where no identical or similar precedent exists? AI systems, which are not currently able to go beyond the reproduction of precedence, remain unable to adapt to social changes. There is accordingly another risk of standardisation of decisions based on the prevalent case law and so the ossification of that case law.[95] Mechanical jurisprudence may thus stagnate the evolution of the law and lead to petrification of the legal system, which will be unable to adapt to contemporary legal and social challenges with different perspectives.[96] Probable risks arising from unprecedented situations should be taken into consideration for the sake of the development and adaptation of law to maintain its vivid characteristics.

Fourthly, lack of legal reasoning in decisions undermines the effectiveness of the justice system. On inscrutable integral aspects of AI, regarding utilisation of AI algorithms in judicial decisions Volokh expresses that consider the output, not the method[97] by advising to focus on the outcomes of such utilisation rather than to comprehend the decision-making process. Legal reasoning has, however, various functions, such as teaching/training legal minds, convincing fairness of the judgment and to provide legitimacy of justice, enables the right to contest/appeal. Full replacement would however make meaningless not only the defences made by human lawyers, but also the very existence of the appellate system. This is because of the deterministic nature of automated judicial decisions, since they, with the ultimate decision-making quality, would not be subject to any further interpretation, thus entailing that machines would influence or even create laws, which may lead to the invasion

[90] The Law Society Commission on the Use of Algorithms in the Justice System, 'Algorithms in the Criminal Justice System' (04 June 2019) Report, The Law Society of England and Wales <https://www.lawsociety.org.uk/topics/research/algorithm-use-in-the-criminal-justice-system-report> accessed 20 October 2024.
[91] Završnik, 'Algorithmic justice' (n 24).
[92] Joshua P Davis, 'Of Robolawyers and Robojudges' (2022) 73(5) Hastings Law Journal 1173.
[93] Jeremias Adams-Prassl, Reuben Binns, and Aislinn Kelly-Lyth, 'Directly Discriminatory Algorithms' (2023) 86(1) MLR 144.
[94] Lelieur et al (n 32) 49.
[95] Notaro (n 34).
[96] Federico Galli and Giovanni Sartor, 'AI Approaches to Predictive Justice: A Critical Assessment' (2023) 5 Humanities and Rights, Global Network Journal 165.
[97] Eugene Volokh, 'Chief Justice Robots' (2019) 68(6) Duke Law Journal 1135.

of automation of decisions beyond the courtroom and into the legislative process.[98] The paralysis of the appeal system thus arises if the software used at first instance and on appeal become identical, the fact of which would render the right to appeal illusory.[99] In this regard, how to devise the criteria for appellate court machines' decision-making is challenging.[100] Such an automated decision-making encoded with an ultimate paradigmatic conception would also hamper the right to lawful judge. Automated decision-making has the potential to affect also the preliminary ruling procedure.

The application of automated decision-making in the criminal justice system should therefore be examined from the perspective of certain criminal law principles, such as the right to lawful judge, the right to a fair trial, the right to defence and equality of arms in adversarial proceedings.[101] For instance, on the one hand, while law enforcement authorities have access to data possessed by private companies constructing AI systems, defence lawyers may not, on the other hand, while private parties can afford AI tools, due to budgetary restrictions, prosecutors and judges might not.[102] The right to access to court, the right to fair trial under Article 6(1) of the ECHR and the principle of effective judicial protection enshrined in Article 47 of the Charter would also be infringed. The right of access to court under Article 6(1) of the ECHR requires judicial review by a domestic court of full jurisdiction to examine all questions of fact and law relevant to the dispute before it, the factual background of the case, the relevant evidence and the application of the relevant law to the facts of the case.[103] Article 6(1) of the ECHR requires effective access to court to obtain such a review, being deprived of access to an appellate jurisdiction satisfying the requirements of Article 6(1) would in that regard constitute infringement of the right to access to justice.[104] In terms of the right to a fair trial, the asymmetries in information between the parties, especially within the context of the black-box problem and inequality of arms further carry the potential to infringe both the ECHR and EU fundamental rights law. De-humanised, de-subjectivated, non-individualised and legal reasoning absent justice based upon automated decision-making would therefore undermine those rights.

On those grounds, certain instances of decision-making in criminal justice should remain a domain reserved to human judges.[105] Judicial decision-making that is especially subject to the exercise of discretion should be kept as a unique human faculty. Law has been a human activity and must remain as such, as merely supported by the technology of AI but

[98] Galli and Sartor (n 96).
[99] Lelieur et al (n 32) 46.
[100] Žarko Dimitrijević, 'Smart Algorithms as a Prerequisite for the Use of Artificial Intelligence in Judicial Decision-Making' (2023) Year XII Issue 2023, Harmonious Journal of Legal and Social Studies in South East Europe 78.
[101] *Sigurður Einarsson and Others v Iceland* App no 39757/15 (ECtHR, 4 June 2019) paras 66, 85-89.
[102] Lelieur et al (n 32) 46.
[103] *Capital Bank Ad v Bulgaria* App no 49429/99 (ECtHR, 24 February 2006) para 98; *Project-Trade D.O.O. v Croatia* App no 1920/14 (ECtHR, 20 December 2013) paras 50 and 67.
[104] *Credit and Industrial Bank v The Czech Republic* App no 29010/95 (ECtHR, 21 October 2003) paras 72-73; *Capital Bank Ad v Bulgaria* (n 103) para 117.
[105] Johanna Sprenger and Dominik Brodowski, "Predictive policing', 'Predictive Justice', and the use of 'Artificial Intelligence' in the Administration of Criminal Justice in Germany' (2023) A-02 Association internationale de droit penal 5.

never replaced by or subordinated to it.[106] Otherwise, judges and legal professionals may delegate their tasks to machines with the result of relegating humans to a subordinate position to algorithms.[107] Although there is lack of proper ethical criteria for a comparative assessment between the performance of algorithms and humans in criminal justice and of the theoretical resources to determine which is ethically preferable,[108] there should be categorical objection to the substitution or full replacement of human judges. Substitution of AI for human judgment would otherwise undermine judicial independence. There should not therefore arise concerns whether algorithms will bring the future with a rule of law or a rule of algorithm.[109] Categorical objection to such substitution is not only a matter of whether algorithms are at present capable of outperforming human judge decisions and judgments. It is also an ontological and a moral matter about: the determination of what kind of society we want to construct on the basis of whose value; where to place human element in it; who should be the ultimate arbiter to resolve disputes between humans; whether justice for humans could be delegated to AI, which lacks of factors peculiar to human beings such as emotion, empathy, intuition, discretion, common sense, conscience, value judgments and sense of justice/fairness. The latter matter arises as such despite the fact that no one could contrarily argue that the existing criminal justice system operates perfectly without any bias, discrimination, arbitrariness and injustice.

In that regard, the human-in-the-loop approach reinforces the idea that AI should support, but never supplant human expertise and ethical judgment. In that regard as declared by the Council, AI must not interfere with the decision-making power of human judges or judicial independence and a court decision cannot be delegated to an AI tool and must always be made by a human being.[110] In that respect, especially Recital 61 of EU AI Act expresses that '[t]he use of AI tools can support the decision-making power of judges or judicial independence, but should not replace it: the final decision-making must remain a human-driven activity'. To enforce this principle, legal frameworks should implement mandatory AI impact assessments before deployment in judicial settings and human-in-the-loop mechanisms, ensuring that human judges remain, with effective discretion, in the centre of judicial decision-making and AI outputs are reviewed and contextualised by legal professionals. This should be fostered by transparent auditing procedures for AI-generated recommendations, allowing external oversight and accountability. Additionally, there should be training programs for judges and legal professionals to enhance AI literacy, preventing uncritical acceptance of or overreliance on algorithmic outputs.

Furthermore, Article 22 of the GDPR, similarly to Article 15 of Data Protection Directive[111] and Article 11 of Law Enforcement Directive, gives the data subject 'the right

---

[106] M Patrão Neves and A Betâmio de Almeida, 'Before and Beyond Artificial Intelligence: Opportunities and Challenges' in Henrique Sousa Antunes et al (eds), *Multidisciplinary Perspectives on Artificial Intelligence and the Law*, (Springer 2024) 123.
[107] Galli and Sartor (n 96).
[108] Jesper Ryberg, 'Artifcial intelligence at sentencing: when do algorithms perform well enough to replace humans?' (2024) AI and Ethics.
[109] Greenstein (n 39).
[110] Council of the European Union, 'Council Conclusions – Access to Justice – Seizing the Opportunities of Digitalisation' 2020/C 342 I/01.
[111] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L281/31.

not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her'. This right, accompanied with the right to obtain human intervention and to contest the decision in order to maintain human oversight over AI systems, however, is subject to three exceptions: if it is necessary for contractual purposes; if it is authorised by Union or Member State law laying down safeguards for the data subject; and if it is based on the data subject's explicit consent. Recital 71 of the GDPR entails that automated processing should be subject to suitable safeguards for the data subject to obtain an explanation of the decision reached after such assessment and to challenge the decision. Article 13(2) of the GDPR also provides for the data subjects with the information of the existence of automated decision-making and meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject. The right not to be subject to automated decision-making, the right to obtain human intervention and the right to challenge such decisions are also recognised by the CJEU.[112]

Digital justice on the one hand may offer an algorithmic decision that replaces a human decision within the context of supporting judges with certain advantages in terms of effectiveness, efficiency, speed and margin of error[113] with capabilities of investigation, massive amount of data-processing, analysing information, bias-detecting, enhancing legal cognition, ensuring human judges access to widespread relevant precedents, identifying patterns, generating predictive risk assessments and identification of certain crimes such as cybercrimes or deepfakes. On the other hand, it may pose risks to fundamental rights, such as biases and discrimination, and to judicial impartiality and independence and human-centric judicial decision-making. Given the compensating performance of AI systems in the administration of justice it would not be plausible to raise a categorical objection to deploying AI system for assisting, but merely to any form of automated judicial decision-making replacing human judges. To be precise, human-centric conception of justice requires both categorical rejection to automated decision-making and precautionary utilisation of the assistive dimension of AI.

For the foregoing reasons, a human(-centric) component should be a must in the criminal justice system and so human-centric, human-made, human-supportive/complementary and human-controlled AI as declared by the European Parliament should be preserved in the system.[114] Given the certain advantages stemming from the use of AI in the criminal justice system, a hybrid model seems to be the best to ensure with the firm reservation of non-elimination of the human factor from decision-making in the criminal justice system. In such assistive form, AI should merely enable human judges to concentrate in their case analysis more on substantive legal issues and help judges with drafting legal documents and decisions. AI may collect and interpret data, process the information derived from them, find patterns in them and make predictions on the basis of those patterns. As declared by the Council, AI may 'improve the functioning of justice systems for the benefit of citizens and businesses by assisting judges and judicial staff in their

[112] Case C-634/21 *SCHUFA Holding* EU:C:2023:957 paras. 54-56, 66.
[113] Angela Busacca and Melchiorre Monaca, 'Using AI for Justice: Principles and Criteria of the "European Ethical Charter on the Use of AI in Judicial Systems"' in Domenico Marino and Melchiorre Monaca (eds), *Artificial Intelligence and Economics: the Key to the Future* (Springer 2023) 157-172.
[114] European Parliament Resolution (n 21).

activities, accelerating court/tribunal proceedings and helping enhance the comparability, consistency and, ultimately, the quality of judicial decisions'.[115] Argument-mining capability of AI may propose to human judges nuanced perspectives from precedents and so may provide a foundational basis for robust and well-informed decision-making.[116] Summarisation and analysis tools distil extensive legal documents and case law into concise and digestible insights and facilitate quicker comprehension of complicated cases.[117] Identifying similar cases may provide judges with a broader and holistic comprehension of legal issues.[118] For instance, evidence-based judicial decision-making would indeed be improved by the use of AI.[119] While leaving the human judge as the ultimate judicial decision-maker, it would thus be reasonable to use AI in the criminal justice system insofar as it replaces labour-intensive and paper-based systems.[120] Information technology could accordingly be used to facilitate the judicial task.[121] In using IT this way, judges certainly require technical expertise to efficiently use and evaluate outcomes of AI systems on the basis of AI specialised educational and training programs.

On the other hand, human judges should be able to distance themselves from AI outputs. Human judges should refrain from the blind pursuit of automated outputs. In that regard, accuracy, precision, recall, effectiveness, fairness, security, robustness, traceability, explicability and so trustworthiness and reliability are parameters to be taken into account when assessing algorithms to keep track of false positives and false negatives engendered by predictive models.[122] Ensuring the trustworthiness of AI is a significant step to achieve both individual and collective human wellbeing, the ultimate aims for using AI.[123] The principle of control by the user articulated in the Ethics Guidelines[124] thus enables the centrality of the human in the judicial decision. Human oversight therefore keeps the human at the centre and provides for the supportive operation of AI in compliance with fundamental rights and ethical values to draw public confidence and support. Human oversight is significant for the protection of fundamental rights and human autonomy against AI/machine autonomy.[125] Accountability for abuses and errors committed in automated decision-making processes and the possibility to review and overturn mistaken judicial

[115] Council of the European Union, 'Seizing the Opportunities of Digitalisation' (n 110).
[116] Galli and Sartor (n 96).
[117] ibid.
[118] ibid.
[119] Hartmann and Wenzelburger (n 26).
[120] Teodor Manea and Dragos Lucian Ivan, 'AI Use in Criminal Matters as Permitted under EU Law and as Needed to Safeguard the Essence of Fundamental Rights' (2022) 1(1) International Journal of Law in Changing World 1.
[121] Mikael Rask Madsen and Robert Spano, 'Authority and Legitimacy of the European Court of Human Rights: Interview with Robert Spano, President of the European Court of Human Rights' (2021) iCourts Working Paper Series, no. 236.
[122] Federico Boggia, 'Artificial intelligence in the Criminal Justice System The Role of decision-makers and how big data tools support them' (2 June 2022) <https://drive.binatomy.com/AIJustice.pdf> accessed 10 July 2024; Francisco J Castro-Toledo, Fernando Miró-Llinares, and Jesús Carreras Aguerri, 'Data-Driven Criminal Justice In The Age Of Algorithms: Epistemic Challenges And Practical Implications' (2023) 34 Criminal Law Forum 295.
[123] AI-HLEG (n 10) 9-11.
[124] ibid., 26-27.
[125] Riikka Koulu, 'Proceduralizing control and discretion: Human oversight in artificial intelligence policy' (2020) 27(6) Maastricht Journal of European and Comparative Law 720.

decisions made by or with the support of AI with the chance to challenge them accordingly may reduce negative consequences of algorithmisation.[126]

4 CONCLUSION

The EU's Human-Centric AI Framework represents a pioneering vision for the responsible deployment and development of AI technologies. By prioritising ethical principles and fundamental values and rights at the core of its AI strategy, the EU aims to foster an ecosystem where AI can be a force for good, enhancing societal well-being while mitigating risks. As this framework is put into practice, particularly in critical areas like criminal justice, it will likely evolve in response to emerging challenges and technological advancements, maintaining its core commitment to placing humans at the centre of the AI (r)evolution.

The EU's human-centric AI strategy in that regard offers a blueprint for the future of integrating AI into the criminal justice system in a way that upholds human rights, promotes fairness and maintains public trust. As countries around the world grapple with the challenges and opportunities presented by AI in criminal justice, the implications arising from the EU's approach therefore appear both timely and instructive. The EU's AI strategy can be a model for balancing innovation with fundamental rights and values. With international collaboration, the EU can lead global efforts towards trustworthy AI practices. By prioritising ethical considerations, transparency and inclusivity, the criminal justice system therefore can harness the power of AI to improve outcomes without compromising fundamental values and rights. Moreover, to prevent and rectify biases in AI algorithms, the EU rigorously scrutinises the implementation of AI, which may perpetuate historical biases and injustices leading to discriminatory outcomes. Additionally, the EU advocates for explainable AI, as it enables stakeholders to understand and evaluate the logic behind AI-driven decisions.[127] This approach builds public trust and ensures that AI is used ultimately in compliance with the rule of law and EU fundamental values.

By prioritising human oversight, the EU stands as a guardian against the de-humanisation, de-subjectivation, de-individualisation of justice or legal reasoning absent justice and underscores the importance of keeping human judgment at the core of AI systems, especially those designated as high-risk. The strategy's focus on reducing bias and enhancing fairness addresses critical ethical concerns, aiming to ensure AI tools supporting equitable justice rather than perpetuating existing disparities. Transparency and accountability form another cornerstone of the EU's framework, advocating for explainable AI systems in fostering trust and enabling ethical and responsible use. The strong emphasis on privacy and data protection aligns with the EU's broader commitment to fundamental rights, ensuring that AI applications in criminal justice safeguard sensitive personal information. The call for multi-stakeholder engagement reflects the EU's recognition that the development and deployment of AI in criminal justice require a collaborative effort, drawing on the expertise and perspectives of a diverse range of actors. This inclusive approach not only enriches the AI development process but also ensures that these powerful technologies are aligned with societal values, ethical and legal norms. Categorical objection

[126] Carrera, Mitsilegas, and Stefan (n 62).
[127] CEPEJ (n 47) 25.

to substitution of human judges by AI also keeps the human component always at the centre of the judicial decision-making, especially in criminal justice.

As AI continues to evolve and its application in criminal justice becomes more pervasive, the implications arising from the EU AI strategy offers timely and essential guidance for the development of AI systems that are not only to be technologically advanced, but also to be ethical, equitable, human-centred and aligned with fundamental rights. The EU's framework accordingly sets a benchmark for trustworthy AI practice. Ultimately, the EU's AI human-centred AI strategy emphasises that the path to a safe, technology-integrated criminal justice system must be navigated with a commitment to human dignity and the common well-being of humans.

LIST OF REFERENCES

Adams-Prassl J, Binns R, and Kelly-Lyth A, 'Directly Discriminatory Algorithms' (2023) 86(1) The Modern Law Review 144
DOI: https://doi.org/10.1111/1468-2230.12759

Bagaric M et al, 'The Solution to the Pervasive Bias and Discrimination in the Criminal Justice System: Transparent and Fair Artificial Intelligence' (2022) 59(1) American Criminal Law Review 95

Barysė D and Sarel R, 'Algorithms in the court: does it matter which part of the judicial decision-making is automated?' (2024) 32 Artificial Intelligence and Law 117
DOI: https://doi.org/10.1007/s10506-022-09343-6

Biber SE, 'Between Humans and Machines: Judicial Interpretation of the Automated Decision-Making Practices in the EU' (2023) University of Luxembourg Law Research Paper Series 2023-19

Bradford A, *Digital Empires - The Global Battle to Regulate Technology* (Oxford University Press 2023)
DOI: https://doi.org/10.1093/oso/9780197649268.001.0001

Brodsky T, 'Artificial Intelligence in the Criminal Justice System: The Ethical Implications of Lawyers Using AI' (2023) Hofstra Law Student Works

Bruno Lepri, Nuria Oliver, and Alex Pentland 'Ethical machines: The human-centric use of artificial intelligence' (2021) 24(3) iScience, Article 102249
DOI: https://doi.org/10.1016/j.isci.2021.102249

Carrera S, Mitsilegas V, and Stefan M, 'Criminal Justice, Fundamental Rights and the Rule of law in the Digital Age – Report of CEPS and QMUL Task Force' (Centre for European Policy Studies (CEPS) Brussels, May 2021)

Castro-Toledo FJ, Miró-Llinares F, and Aguerri JC, 'Data-Driven Criminal Justice In The Age Of Algorithms: Epistemic Challenges And Practical Implications' (2023) 34 Criminal Law Forum 295
DOI: https://doi.org/10.1007/s10609-023-09454-y

Chiao V, 'Fairness, accountability and transparency: notes on algorithmic decision-making in criminal justice' (2019) 15(2) International Journal of Law in Context 126
DOI: https://doi.org/10.1017/s1744552319000077

Çınar OH, 'The current case law of the European Court of Human Rights on privacy: challenges in the digital age' (2021) 25(1) The International Journal of Human Rights 26
DOI: https://doi.org/10.1080/13642987.2020.1747443

David Restrepo Amariles and Pablo Marcello Baquero, 'Promises and limits of law for a human-centric artificial intelligence' (2023) 48 Computer Law & Security Review, Article 105795
DOI: https://doi.org/10.1016/j.clsr.2023.105795

Davis JP, 'Of Robolawyers and Robojudges' (2022) 73(5) Hastings Law Journal 1173.

Dimitrijević Ž, 'Smart Algorithms as a Prerequisite for the Use of Artificial Intelligence in Judicial Decision-Making' (2023) Year XII Issue 2023, Harmonious Journal of Legal and Social Studies in South East Europe 78
DOI: https://doi.org/10.51204/harmonius_23104a

G'sell F, 'AI Judges' in DiMatteo LA, Poncibò C, and Cannarsa M (eds), *The Cambridge Handbook of Artificial Intelligence, Global Perspectives on Law and Ethics* (Cambridge University Press 2022)
DOI: https://doi.org/10.1017/9781009072168.032

Galli F and Sartor G, 'AI Approaches to Predictive Justice: A Critical Assessment' (2023) 5 Humanities and Rights, Global Network Journal 165

Garrett BL and Rudin C, 'The Right to A Glass Box: Rethinking the Use of Artificial Intelligence in Criminal Justice' (*SSRN*, 22 November 2022) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4275661#> accessed 25 August 2024
DOI: https://doi.org/10.2139/ssrn.4275661

Gentelet K and Mizrahi SK, 'A Human-Centered Approach to AI Governance: Operationalizing Human Rights through Citizen Participation' in Régis C et al (eds), *Human-Centered AIA Multidisciplinary Perspective for Policy-Makers, Auditors, and Users* (CRC Press 2024)
DOI: https://doi.org/10.1201/9781003320791-24

Gentile G, 'Artificial Intelligence and the Crises of Judicial Power: (Not) Cutting the Gordian Knot?' (*SSRN*, 22 February 2024) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4731231> accessed 10 July 2024.

Gravett WH, 'Judicial Decision-Making in the Age of Artificial Intelligence' in Sousa Antunes H et al (eds), *Multidisciplinary Perspectives on Artificial Intelligence and the Law* (Springer 2024)
DOI: https://doi.org/10.1007/978-3-031-41264-6_15

Greenstein S, 'Preserving the rule of law in the era of artificial intelligence (AI)' (2022) 30 Artificial Intelligence and Law 291
DOI: https://doi.org/10.1007/s10506-021-09294-4

Gunning D et al, 'XAI - Explainable Artificial Intelligence' (2019) 37(4) Science Robotics aay7120
DOI: https://doi.org/10.1126/scirobotics.aay7120

Hartmann K and Wenzelburger G, 'Uncertainty, risk and the use of algorithms in policy decisions: a case study on criminal justice in the USA' (2021) 54 Policy Sciences 269
DOI: https://doi.org/10.1007/s11077-020-09414-y

Kearns M and Roth A, *The Ethical Algorithm – The Science Of Socially Aware Algorithm Design* (Oxford University Press 2019)

Komoda J, 'Designing AI for Courts' (2023) 29(3) Richmond Journal of Law & Technology 145

Koulu R, 'Proceduralizing control and discretion: Human oversight in artificial intelligence policy' (2020) 27(6) Maastricht Journal of European and Comparative Law 720
DOI: https://doi.org/10.1177/1023263x20978649

Krištofík A, 'The Role of Privacy in the Establishment of the Right Not to Be Subject to Automated Decision-Making' (2024) 2/2024 TLQ 236

Lelieur J et al, 'General Report' in Lelieur J (ed), *Artificial Intelligence and Administration of Criminal Justice* (International Colloquium, Buenos Aires, Argentina, 28-31 March 2023) 94 Revue Internationale de Droit Pénal 11

Malek MA, 'Criminal courts' artificial intelligence: the way it reinforces bias and discrimination' (2022) 2 AI and Ethics 233
DOI: https://doi.org/10.1007/s43681-022-00137-9

Manea T and Ivan DL, 'AI Use in Criminal Matters as Permitted under EU Law and as Needed to Safeguard the Essence of Fundamental Rights' (2022) 1(1) International Journal of Law in Changing World 1
DOI: https://doi.org/10.54934/ijlcw.v1i1.15

Neves MP and de Almeida AB, 'Before and Beyond Artificial Intelligence: Opportunities and Challenges' in Sousa Antunes H et al (eds), *Multidisciplinary Perspectives on Artificial Intelligence and the Law*, (Springer 2024)
DOI: https://doi.org/10.1007/978-3-031-41264-6_6

Notaro L, 'Predictive Algorithms and Criminal Justice: A Synthetic Overview from An Italian and European Perspective' (2020) 2 Roma Tre Law Review 49.

Ozmen Garibay O et al, 'Six human-centered artificial intelligence grand challenges' (2023) 39(3) International Journal of Human-Computer Interaction 391
DOI: https://doi.org/10.1080/10447318.2022.2153320

Peršak N, 'Automated Justice and its Limits: Irreplaceable Human(E) Dimensions of Criminal Justice' in Vermeulen G, Peršak N, and Recchia N (eds), *Artificial Intelligence, Big Data and Automated Decision-Making in Criminal Justice* (2021) 92/1 Revue Internationale de Droit Pénal 13

Pirozzoli A, 'The Human-centric Perspective in the Regulation of Artificial Intelligence' (2024) 9 European Papers 105

Rahnama K, 'Science and Ethics of Algorithms in the Courtroom' (2019) 1 Journal of Law, Technology & Policy 169

Rask Madsen M and Spano R, 'Authority and Legitimacy of the European Court of Human Rights: Interview with Robert Spano, President of the European Court of Human Rights' (2021) iCourts Working Paper Series, no 236

Ríos PM, 'Predictive algorithms and criminal justice: expectations, challenges and a particular view of the Spanish VioGén system' (2024) 2/2024 Rivista italiana di informatica e diritto 547

Ryberg J, 'Artifcial intelligence at sentencing: when do algorithms perform well enough to replace humans?' (2024) AI and Ethics
DOI: https://doi.org/10.1007/s43681-024-00442-5

Sargiotis D, 'Fostering Ethical and Inclusive AI: A Human-Centric Paradigm for Social Impact' (2025) 6(1) International Journal of Research Publication and Reviews 3754
DOI: https://doi.org/10.55248/gengpi.6.0125.0529

Shi C, Sourdin T, and Li B, 'The Smart Court – A New Pathway to Justice in China?' (2021) 12(1) International Journal for Court Administration 4
DOI: https://doi.org/10.36745/ijca.367

Shi J, 'Artificial Intelligence, Algorithms and Sentencing in Chinese Criminal Justice: Problems and Solutions' (2022) 33 Criminal Law Forum 121
DOI: https://doi.org/10.1007/s10609-022-09437-5

Siapka A, 'The Ethical and Legal Challenges of Artificial Intelligence: The EU response to biased and discriminatory AI' (Thesis, Panteion University of Athens, 2018)
DOI: https://doi.org/10.2139/ssrn.3408773

Sigfrids A et al, 'Human-centricity in AI governance: A systemic approach' (2023) 6 Frontiers in Artificial Intelligence 2
DOI: https://doi.org/10.3389/frai.2023.976887

Sprenger J and Brodowski D, "Predictive policing', 'Predictive Justice', and the use of 'Artificial Intelligence' in the Administration of Criminal Justice in Germany' (2023) A-02 Association internationale de droit penal 5

Ulenaers J, 'The Impact of Artificial Intelligence on the Right to a Fair Trial: Towards a Robot Judge?' (2020) 11(2) Asian Journal of Law and Economics 1
DOI: https://doi.org/10.1515/ajle-2020-0008

Volokh E, 'Chief Justice Robots' (2019) 68(6) Duke Law Journal 1135

Xenidis R and Senden L, 'EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination' in Bernitz U et al (eds), *General Principles of EU law and the EU Digital Order* (Kluwer Law International 2020)

Yeung K, 'Why worry about decision-making by machine?' in Yeung K and Lodge M (eds), *Algorithmic Regulation* (Oxford University Press 2019)
DOI: https://doi.org/10.1093/oso/9780198838494.003.0002

— —, Howes A, and Pogrebna G, 'AI Governance by Human Rights-Centred Design, Deliberation and Oversight: An End to Ethics Washing' in Dubber MD, Pasquale F, and Das S (eds), *The Oxford Handbook of AI Ethics* (Oxford University Press 2020)
DOI: https://doi.org/10.1093/oxfordhb/9780190067397.013.5

Završnik A, 'Criminal justice, artificial intelligence systems, and human rights' (2020) 20 ERA Forum 567
DOI: https://doi.org/10.1007/s12027-020-00602-0

— —, 'Algorithmic justice: Algorithms and big data in criminal justice settings' (2021) 18(5) European Journal of Criminology 623
DOI: https://doi.org/10.1177/1477370819876762