ARTIFICIAL INTELLIGENCE IN CRIMINAL JUSTICE: LEGAL CHALLENGES TO FAIR TRIAL GUARANTEES AND PROCEDURAL SAFEGUARDS

DOI: https://doi.org/10.53486/dri2026.87
UDC: 343.1:004.8

Kamran KHALILOV
Baku State University
Baku, Azerbaijan
Email account: xalilovkamran75@gmail.com
ORCID ID: 0000-0002-0086-009X

Abstract: *The increasing use of artificial intelligence in criminal justice is no longer a distant prospect but an emerging reality that is gradually reshaping procedural practices. Algorithmic tools are increasingly involved in risk assessment, data analysis, and evidentiary evaluation, particularly in cases involving complex digital environments. The main aim of this study is to examine the legal implications of integrating artificial intelligence into criminal proceedings, with a particular focus on its impact on fair trial guarantees and procedural safeguards. The analysis draws on international human rights standards, selected jurisprudence of the European Court of Human Rights, and recent developments in European Union regulatory approaches to artificial intelligence. The research is based on a doctrinal and comparative legal methodology, allowing for an assessment of how existing legal frameworks respond to the challenges posed by algorithmic decision-making. Particular attention is given to issues of transparency, contestability, and the ability of parties to effectively challenge AI-generated outcomes. The findings indicate that the opacity of algorithmic systems creates significant difficulties for maintaining equality of arms, particularly in situations where procedural rules do not adequately ensure meaningful access to the reasoning underlying automated outputs. It is argued that without clearly defined safeguards ensuring transparency, accountability, and proportionality, the use of artificial intelligence risks undermining the fairness and legitimacy of criminal proceedings.*

Key words: *artificial intelligence, criminal justice, fair trial, procedural safeguards, legal accountability.*

JEL: K14, K40, O33.

Introduction

Over the past decade, the use of artificial intelligence within criminal justice systems has moved beyond theoretical discussion and entered practical application. Law enforcement authorities and judicial actors increasingly rely on algorithmic tools to process large datasets, identify patterns, and support decision-making. This development is often framed as a response to the growing complexity of digital environments, where traditional procedural mechanisms struggle to manage the scale and speed of information flows.

At first glance, such technologies appear to enhance efficiency and objectivity. Yet their integration into criminal proceedings does more than accelerate existing processes. It gradually reshapes how decisions are formed and justified. Elements that were once grounded in human reasoning are now influenced by systems whose internal logic cannot always be fully accessed or explained within the courtroom. In this sense, the issue is not limited to technical reliability but extends to the conditions under which legal judgments are produced.

A central difficulty arises from the tension between algorithmic opacity and the requirements of procedural transparency. Criminal procedure has long been built on the assumption that evidence can be examined, challenged, and interpreted by all parties involved. Where algorithmic tools are used to generate or evaluate information, this assumption becomes less stable. The reasoning underlying automated outputs may not be meaningfully accessible, which can affect the ability of the defense to contest such material and, consequently, disturb the balance between the parties.

These concerns are not merely theoretical. Empirical and doctrinal discussions on algorithmic decision-making have shown that such systems may reproduce patterns embedded in historical data, including forms of structural bias (Angwin et al., 2016). At the same time, European human rights law consistently emphasizes that any interference with individual rights must satisfy the requirements of foreseeability, proportionality, and effective judicial control, as reflected in the jurisprudence of the European Court of Human Rights (Klass and Others v. Germany, 1978; Roman Zakharov v. Russia, 2015).

From a doctrinal perspective, this creates a structural tension between efficiency-oriented technological development and the normative foundations of criminal justice. Judicial decision-making is not merely a technical exercise, but a process grounded in legal reasoning, accountability, and the obligation to provide reasons that can be scrutinized by the parties and, where necessary, by higher courts. As noted in the literature, the increasing reliance on automated systems challenges the traditional understanding of judicial responsibility and the transparency of adjudication (Mitsilegas, 2022).

In this context, it becomes increasingly difficult to treat artificial intelligence as simply another technical instrument. The more complex question concerns whether existing legal frameworks are capable of accommodating such technologies without weakening core procedural guarantees. This includes not only the right to a fair trial, but also broader principles such as equality of arms, legal certainty, and the integrity of judicial reasoning.

The analysis that follows adopts a doctrinal and comparative legal perspective. Rather than focusing on the technical architecture of artificial intelligence systems, it examines their legal implications within criminal proceedings. Particular attention is given to the relationship between algorithmic decision-making and fundamental procedural safeguards, especially in light of evolving European legal standards and the emerging regulatory framework on artificial intelligence.

Basic content

The use of artificial intelligence in criminal justice.

The integration of artificial intelligence into criminal justice has not occurred through a single coordinated reform, but rather through the gradual introduction of tools addressing specific practical needs. Today, algorithmic systems are used at various stages of criminal proceedings, including policing, pre-trial decision-making, and the handling of evidence. Although these systems do not operate independently of legal actors, they increasingly shape the informational environment within which decisions are formed.

In the context of policing, predictive models are used to identify spatial and temporal patterns of crime and to support the allocation of resources. Systems such as PredPol rely on historical crime data to generate forecasts about potential future incidents. While these models are often presented as neutral and data-driven, their outputs remain closely tied to the structure of the underlying data. Where past enforcement practices have reflected existing social inequalities, predictive systems may reproduce these patterns rather than correct them (Ferguson, 2017). As a result, algorithmic predictions may reinforce pre-existing forms of selective attention within policing strategies (Angwin et al., 2016).

A similar set of issues arises in the use of risk assessment tools within judicial decision-making. Instruments such as COMPAS are designed to estimate the likelihood of reoffending and are used in decisions related to bail, sentencing, and parole. Although these tools are intended to improve consistency and reduce subjective bias, empirical studies have shown that their predictive performance is contested and, in certain contexts, does not significantly exceed that of human judgment (Dressel & Farid, 2018). More importantly, the variables on which such systems rely may indirectly reflect socio-economic conditions, leading to outcomes that are difficult to reconcile with the principle of equal treatment before the law.

Artificial intelligence also plays an increasingly significant role in the handling of digital evidence. In investigations involving large volumes of electronic data, algorithmic systems are used to filter,

classify, and prioritize information. This allows investigators to process material at a scale that would otherwise be impractical. At the same time, this development alters the evidentiary landscape. The interpretation of information becomes partially dependent on automated processes, raising questions about how such outputs should be assessed within established legal standards and how their reliability can be effectively scrutinized in court.

What unites these different applications is not their technical design, but their influence on the structure of decision-making. Algorithmic systems do not merely provide additional information; they affect which information is considered relevant and how it is organized. This, in turn, shapes the reasoning process itself. Where decision-makers rely on system-generated outputs without full insight into the processes behind them, the relationship between human judgment and technological input becomes increasingly difficult to define (Mitsilegas, 2022).

From a legal perspective, this shift raises concerns that extend beyond questions of efficiency. The use of artificial intelligence affects the evidentiary basis of proceedings, the distribution of informational power between the parties, and the capacity of courts to exercise independent evaluation. In particular, the limited transparency of many AI systems, often described as the "black box" problem, complicates the ability of the defense to effectively challenge the material relied upon by the prosecution. This challenge is increasingly recognized within the European regulatory framework, where AI systems used in criminal justice are classified as high-risk under the EU Artificial Intelligence Act (Regulation (EU) 2024/1689). Where access to the reasoning underlying algorithmic outputs remains restricted, procedural guarantees, including the right to a fair trial under Article 6 of the European Convention on Human Rights, risk becoming formal rather than substantive (Zalnieriute, 2026).

In this sense, the use of artificial intelligence in criminal justice reflects a broader transformation in the conditions under which legal decisions are produced. While these technologies expand institutional capacity, they also introduce new forms of dependency that are not easily accommodated within existing procedural frameworks. Understanding their role therefore requires not only a technical description of their functions, but a legal assessment of their implications for fairness, accountability, and the overall structure of criminal proceedings.

Table 1. Key Applications of AI in Criminal Justice and Associated Legal Concerns

Application Area	Primary Function	Core Legal and Procedural Risks
Predictive policing	Identifying crime-prone areas and temporal patterns based on historical data.	Reproduction of structural biases; risk of circular reasoning in policing strategies.
Risk assessment	Estimating the likelihood of recidivism to support judicial decisions.	Challenges to the principle of equal treatment; opacity in algorithmic scoring (e.g., COMPAS).
Digital evidence handling	Filtering, classifying, and prioritizing massive datasets (e-discovery).	Reliability concerns; difficulties in establishing an accessible chain of evidentiary reasoning.
Decision-support systems	Assisting prosecutors and judges in complex legal evaluations.	"Automation bias"; shifting the balance between human judgment and algorithmic input.

Source: Author's compilation based on Ferguson (2017), Angwin et al. (2016), Dressel & Farid (2018), and Mitsilegas (2022).

Legal challenges of algorithmic decision-making

While the use of artificial intelligence in criminal justice is often justified in terms of efficiency and consistency, its integration raises a set of legal challenges that cannot be addressed within traditional procedural frameworks. These challenges do not arise solely from the technical complexity of algorithmic systems, but from the mismatch between their operational logic and the normative structure of criminal procedure (Hildebrandt, 2016).

One of the most persistent difficulties concerns the issue of transparency. Many algorithmic systems used in criminal justice operate through complex computational models that are not readily interpretable by non-specialists. This lack of interpretability becomes legally significant when such systems influence decisions affecting individual rights. Unlike traditional forms of evidence, which can be examined and contested through established procedural mechanisms, algorithmic outputs often lack a clear and accessible chain of reasoning (Pasquale, 2015). As a result, it becomes difficult to determine on what basis a particular conclusion has been reached.

Closely related to this is the problem of explainability. In legal proceedings, the ability to provide reasons for a decision is not merely a formal requirement, but a substantive safeguard that allows parties to understand and challenge the outcome. Where algorithmic systems are involved, this requirement is not always easily satisfied. Even when some level of explanation is provided, it may not be sufficient to allow meaningful scrutiny within an adversarial process (Selbst and Baracos, 2018). This creates a situation in which the formal existence of procedural rights does not necessarily translate into their effective exercise.

These concerns directly affect the principle of equality of arms. Criminal proceedings are structured around the idea that both parties should have a fair opportunity to present their case under conditions that do not place one side at a substantial disadvantage. Where one party—typically the prosecution—relies on algorithmic tools whose functioning is not fully disclosed or understood, this balance may be disrupted. The defense may be placed in a position where it is required to challenge conclusions without access to the underlying reasoning, which undermines the effectiveness of adversarial argument (Zalnieriute, 2026).

Another important dimension of the problem relates to accountability. In traditional legal frameworks, responsibility for decisions can be attributed to identifiable actors, such as investigators, prosecutors, or judges. The introduction of algorithmic systems complicates this model. Where decisions are influenced by automated tools, it becomes less clear who bears responsibility for potential errors or biases. This diffusion of responsibility raises concerns not only at the level of individual cases, but also in relation to systemic oversight (Binns, 2018).

These issues are increasingly reflected in European legal developments. The EU Artificial Intelligence Act explicitly classifies AI systems used in criminal justice as high-risk, recognizing their potential impact on fundamental rights. However, the regulatory approach remains primarily ex ante, focusing on risk management and compliance requirements, rather than on the procedural implications within individual cases. This creates a gap between regulatory design and procedural reality, where formal safeguards may not fully address the challenges that arise in practice.

From the perspective of human rights law, the central concern is whether existing legal guarantees remain effective in the context of algorithmic decision-making. The jurisprudence of the European Court of Human Rights has consistently emphasized that rights such as the right to a fair trial must be practical and effective, rather than theoretical or illusory. Applying this standard to cases involving artificial intelligence suggests that the mere formal availability of procedural rights is insufficient where the conditions necessary for their meaningful exercise are not met.

In this sense, the legal challenges posed by artificial intelligence are not isolated or technical in nature. They reflect a deeper tension between emerging forms of decision-making and the principles on which criminal procedure is based. Addressing this tension requires more than incremental

adjustments; it calls for a reconsideration of how transparency, accountability, and fairness are ensured in a context where the sources of decision-making are no longer fully visible.

Fair trial guarantees and procedural safeguards

The challenges associated with the use of artificial intelligence in criminal proceedings become particularly visible when assessed through the lens of fair trial guarantees. Unlike broader regulatory concerns, these guarantees are not abstract principles but operational standards that determine whether a proceeding can be considered legally legitimate. The integration of algorithmic tools therefore requires an evaluation not only of their functionality, but of their compatibility with the core requirements of procedural justice.

At the center of this assessment lies the right to a fair trial, as enshrined in Article 6 of the European Convention on Human Rights. The jurisprudence of the European Court of Human Rights has consistently emphasized that this right must be "practical and effective" rather than merely formal (Artico v. Italy, 1980). In cases involving complex or opaque forms of evidence, the Court has stressed the importance of ensuring that the accused is able to understand, challenge, and respond to the material used against them. This requirement becomes significantly more difficult to satisfy where decision-making is influenced by algorithmic systems whose internal logic is not fully accessible. Recent scholarship suggests that the growing reliance on automated processes in adjudication raises broader questions about the continued effectiveness of traditional procedural safeguards (Fikfak and Helfer, 2024). In such contexts, there is a risk that the formal availability of rights does not translate into their effective exercise, particularly where the reasoning underlying a decision cannot be meaningfully examined.

One of the most affected guarantees is the principle of equality of arms. This principle requires that each party be given a reasonable opportunity to present its case under conditions that do not place it at a substantial disadvantage vis-à-vis the opponent (Dombo Beheer B.V. v. Netherlands, 1993). Where the prosecution relies on AI-assisted tools, particularly in the analysis of evidence or risk assessment, the defense may face structural limitations in accessing or interpreting the underlying reasoning. Even where disclosure obligations are formally respected, the technical nature of algorithmic outputs may prevent meaningful contestation, thereby weakening the adversarial balance (Zalnieriute, 2026).

Closely linked to this is the right to an effective defense. The ability to challenge evidence presupposes not only access to information, but also the capacity to understand and scrutinize it. In the context of AI-generated outputs, this presupposition is not always met. Where the reasoning process of a system cannot be reconstructed or explained in a comprehensible manner, the defense is placed in a position where it must contest conclusions without being able to engage with their basis. This risks transforming procedural rights into formal guarantees lacking substantive content (Hildebrandt, 2016).

The requirement of judicial impartiality is also affected, albeit in a more subtle way. Judges are expected to base their decisions on evidence that can be evaluated independently and transparently. Where algorithmic tools are used to organize or interpret information, there is a risk that judicial reasoning becomes indirectly shaped by processes that are not fully subject to scrutiny (Parasuraman & Manzey, 2010). This does not necessarily imply bias in the traditional sense, but it may affect the independence of judgment by introducing elements that cannot be fully assessed within the legal framework.

These concerns must also be understood in light of the broader doctrine of the "quality of law," as developed in the case law of the European Court of Human Rights. According to this doctrine, legal rules that interfere with individual rights must be accessible, precise, and foreseeable in their application (Klass and Others v. Germany, 1978; Sunday Times v. UK, 1979). The use of artificial intelligence complicates this requirement, particularly where the operation of algorithmic systems is

not transparent or where their outcomes are not sufficiently predictable. In such situations, the individual may be affected by decisions whose underlying logic cannot be reasonably anticipated.
In this context, the central issue is not whether artificial intelligence can be used in criminal justice, but under what conditions its use remains compatible with fundamental procedural guarantees. The analysis suggests that the existing framework of fair trial rights is not inherently incompatible with technological development. However, its effectiveness depends on the extent to which procedural safeguards are adapted to address the specific risks associated with algorithmic decision-making.
Ultimately, ensuring compliance with fair trial guarantees in the age of artificial intelligence requires more than the formal recognition of rights. It demands the creation of procedural mechanisms capable of translating these rights into practical protections. This includes, in particular, ensuring meaningful access to the reasoning underlying automated outputs, strengthening the capacity of the defense to challenge such material, and preserving the role of the judge as an independent and fully informed decision-maker (Mitsilegas, 2022).

Legal responsibility and accountability

The increasing reliance on artificial intelligence in criminal justice raises a question that is less visible than issues of fairness or transparency, yet equally fundamental: who bears responsibility when algorithmic systems influence legal outcomes. Unlike traditional decision-making structures, where responsibility can be traced to identifiable actors, the use of AI introduces a more fragmented and diffuse model of accountability, a situation that closely resembles what has been described in legal and political theory as the "problem of many hands" (Thompson, 1980; Nissenbaum, 1996).
In conventional procedural frameworks, legal responsibility is attached to human decision-makers. Investigators, prosecutors, and judges are expected to justify their actions and can be held accountable for errors or violations of rights. The introduction of algorithmic systems complicates this model without fully replacing it. Decisions are rarely made by AI alone; rather, they are shaped through an interaction between human judgment and system-generated outputs. This hybrid structure makes it more difficult to determine whether responsibility lies with the individual relying on the system, the institution deploying it, or the entity responsible for its design (Binns, 2018).
This difficulty becomes particularly apparent in cases where algorithmic outputs prove to be inaccurate, biased, or misleading. If a decision is influenced by a system whose internal logic is not fully understood, attributing fault becomes problematic. Human actors may argue that they relied on a tool presented as reliable, while developers may claim that the system was used outside its intended context (Parasuraman & Manzey, 2010). As a result, responsibility risks being diffused across multiple actors, without any single point of clear accountability.
From a legal perspective, this situation challenges the traditional foundations of responsibility. Criminal procedure is built on the assumption that decisions can be traced back to accountable actors whose reasoning can be examined and, where necessary, contested. Where algorithmic systems intervene in this process, this assumption becomes less certain. The difficulty is not only in identifying responsibility after an error has occurred, but in ensuring that responsibility is structured in a way that prevents such errors in the first place.
European regulatory developments reflect an increasing awareness of these concerns, yet they do not fully resolve them. The EU Artificial Intelligence Act (Regulation (EU) 2024/1689) introduces a set of obligations for both developers and users of high-risk systems, including requirements for human oversight (Article 14). While these measures are designed to mitigate risks at the design and deployment stages, their focus remains largely preventive and compliance-oriented.
In practice, this regulatory approach does not easily translate into clear procedural standards for attributing responsibility within individual criminal cases, particularly where algorithmic outputs are relied upon in evidentiary reasoning. The result is a persistent gap between administrative compliance and the requirements of criminal liability or procedural accountability, a tension increasingly noted in the literature (Koops, 2021).

In this context, the challenge is not to replace human responsibility with technological accountability, but to preserve the primacy of human decision-making while ensuring that the use of AI does not obscure the lines of responsibility. This requires a legal approach in which algorithmic tools remain subject to meaningful oversight, and where the actors relying on such tools retain full responsibility for the decisions they influence.

Ultimately, the question of accountability highlights a broader concern: the risk that technological complexity may weaken the link between decision-making and responsibility. Preventing this outcome is essential for maintaining not only procedural fairness, but also the legitimacy of criminal justice systems as a whole.

Conclusions

The increasing integration of artificial intelligence into criminal justice systems reveals a growing tension between technological efficiency and procedural fairness. While AI systems expand the capacity of institutions to navigate complex digital environments, they also affect the conditions under which legal decisions are formed and justified.

The analysis shows that the main legal challenges arise from algorithmic opacity, the difficulty of effectively contesting automated outputs, and the diffusion of responsibility within hybrid decision-making structures. These factors directly affect the practical effectiveness of fair trial guarantees under Article 6 of the ECHR, in particular the principle of equality of arms and the right to an effective defense.

Although the EU Artificial Intelligence Act (2024) and related regulatory initiatives acknowledge these risks, a gap remains between ex ante risk management and the procedural protection of individual rights in concrete cases. Consequently, ensuring meaningful transparency, preserving judicial autonomy, and maintaining clear lines of accountability should not be seen as technical adjustments, but as necessary conditions for sustaining the legitimacy and integrity of criminal justice systems in a technologically mediated environment.

At the same time, the present analysis is subject to certain limitations. It is primarily based on doctrinal and comparative legal methods and does not include empirical assessment of how AI systems function in practice. Future research could further explore the practical implementation of procedural safeguards in real cases, as well as the interaction between technological design and legal standards, in order to develop more effective models for integrating artificial intelligence into criminal justice.

References

1. Angwin, J., Larson, J., Mattu, S., and Kirchner, L., 2016. Machine Bias. *ProPublica,* [online] Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [Accessed 12 April 2026].
2. *Artico v. Italy,* 1980. (App. no. 6694/74). European Court of Human Rights.
3. Binns, R., 2018. Algorithmic accountability and public trust. *International Journal of Law and Information Technology,* 26(1), pp.72-91.
4. *Dombo Beheer B.V. v. the Netherlands,* 1993. (App. no. 14448/88). European Court of Human Rights.
5. Dressel, J. and Farid, H., 2018. The accuracy, fairness, and limits of predicting recidivism. *Science Advances,* 4(1), eaao5580.
6. European Parliament and Council, 2024. *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).*
7. Ferguson, A.G., 2017. *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement.* New York: NYU Press.
8. Fikfak, V. and Helfer, L.R., 2024. Automated Decision-Making in International Human Rights Adjudication. *European Journal of International Law.*
9. Hildebrandt, M., 2016. *Smart Technologies and the End(s) of Law.* Cheltenham: Edward Elgar Publishing.
10. *Klass and Others v. Germany,* 1978. (App. no. 5029/71). European Court of Human Rights.
11. Koops, B. J., 2021. The trouble with European data protection law. In: *Research Handbook on EU Data Protection*

Law.
12. Mitsilegas, V., 2022. *EU Criminal Law.* 2nd ed. London: Hart Publishing.
13. Nissenbaum, H., 1996. Accountability in a Computerized Society. *Science and Engineering Ethics,* 2(1), pp.25-42.
14. Parasuraman, R. and Manzey, D.H., 2010. Complacency and Bias in Human Use of Automation. *Human Factors,* 52(3), pp.381-410.
15. Pasquale, F., 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information.* Cambridge, MA: Harvard University Press.
16. *Roman Zakharov v. Russia,* 2015. (App. no. 47143/06). European Court of Human Rights.
17. Selbst, A.D. and Barocas, S., 2018. The Intuitive Appeal of Explainable Machines. *Fordham Law Review,* 87, p.1085.
18. *Sunday Times v. UK,* 1979. (App. no. 6538/74). European Court of Human Rights.
19. Thompson, D.F., 1980. Moral Responsibility of Public Officials: The Problem of Many Hands. The American Political Science Review, 74(4), pp.905-916.
20. Zalnieriute, M., 2026. AI Judges in Courts of the Future and the Right to a Fair Trial in the ECHR. In: *The Cambridge Handbook of AI and Technologies in Courts.* Cambridge: Cambridge University Press.