NBP. Nauka, bezbednost, policija
Published by
University of Criminal Investigation and Police Studies
2026, Vol. 31, Issue 1, pp. 1–19
https://doi.org/10.5937/nabepo31-62922
Original Scientific Paper

Artificial Intelligence in Criminal Justice: Predictive Tools, Evidentiary Challenges and Human Rights Implications

Olga Koshevaliska[1]
Goce Delcev University in Stip, Faculty of Law, North Macedonia

Submitted: 2025-11-20 • Accepted: 2025-12-28 • Published: 2026-01-16

Abstract: Integrating artificial intelligence (AI) technologies into criminal justice systems introduces both transformative opportunities and profound legal dilemmas. This paper critically examines the use of AI in crime prediction, risk assessment, evidence analysis and sentencing, with particular attention to its impact on fundamental procedural rights. Focusing on predictive policing algorithms, facial recognition systems, and AI-assisted evidence review, the research explores how these tools reshape police, prosecutorial and judicial discretion. Key challenges include transparency, explainability and risks of systemic bias or "automated justice", contrasted with constitutional guarantees such as the presumption of innocence, the right to a fair trial and the principle of legality. The study concludes that while AI can enhance efficiency and accuracy, its uncritical adoption may jeopardize essential human rights protections unless accompanied by robust procedural safeguards. Artificial intelligence should serve as an instrument of human progress, not as a substitute for human judgment.

Keywords: AI in criminal law, algorithmic evidence, risk assessment tools, predictive policing, fair trial, legal safeguards, procedural rights.

INTRODUCTION

In recent years the integration of artificial intelligence (AI) into criminal justice systems has moved beyond theoretical exploration toward practical implementation. Courts, prosecutors and law enforcement authorities increasingly employ algorithmic tools to enhance efficiency, reduce bias and predict future risks, thereby reshaping the very architecture of justice administration. From automated risk assessment and predictive policing to AI-assisted forensic analysis, the criminal law domain is undergoing a deep technological transformation. The accelerating incorporation of AI in criminal justice has redefined how crimes are predicted, investigated, prosecuted and judged. Predictive algorithms, data-driven sentencing recommendations and digital evidence analytics promise unprecedented levels of efficiency and consistency, yet simultaneously challenge the foundational

[1] Corresponding author: olga.kosevaliska@ugd.edu.mk • https://orcid.org/0000-0001-5288-8492 • Phone: +389 78 33 69 42.

eISSN 2620-0406

principles of fairness, transparency, non-discrimination and accountability (Završnik, 2020; Strikwerda, 2020). No longer distant innovations, AI now possesses measurable influence over real-world outcomes, changing procedural and evidentiary standards in both investigative and judicial contexts (Mugari & Obioha, 2021).

This paper examines the deployment of AI across various stages of the criminal process, with a particular focus on its practical implementation, legal safeguards and human rights implications. It focuses on the phases of investigation, prosecution and sentencing – the domains where AI tools are already operational or undergoing pilot testing. The central objective is to evaluate whether AI can support, rather than undermine, the core principles of criminal law: legality, fairness, contradiction, proportionality, non-discrimination and individualised justice.

Through a combination of comparative legal analysis and empirical case studies, this research examines how AI systems operate in practice and identifies the necessary regulatory frameworks to ensure their lawful and ethical deployment. The paper contributes to the ongoing discourse on "Justice 4.0",[2] offering a grounded assessment of AI's potential and its associated risks in the criminal justice domain.

AI technologies are now regularly employed to support decision-making across the criminal justice continuum, ranging from crime prevention and risk assessment to forensic evidence analysis and case management (Situmeang et al., 2024). These systems use machine learning, natural language processing, and predictive analytics to process large datasets, identify behavioural patterns and forecast potential criminal conduct. For instance, predictive policing platforms generate "hot spot" maps indicating locations where crime is statistically more likely to occur (European Crime Prevention Network [EUCPN], 2022). However, this automation introduces complex ethical and legal dilemmas. Scholars have warned that algorithmic systems, when trained on historically biased data, may reproduce or even amplify the existing social inequalities (Parkkavi & Yadharthana, 2024; Završnik, 2020).

At the same time, international policy bodies recognize that responsible AI use can enhance both the effectiveness and transparency of justice systems, provided that adequate safeguards are implemented. The U.S. Department of Justice (2024) has emphasized that while AI can improve accuracy and efficiency, it must remain consistent with fundamental privacy, civil rights and civil liberties protections. Similarly, European institutions, through frameworks such as the EU Artificial Intelligence Act (Regulation 2024/1689) and the CEPEJ European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment (Council of Europe, 2018), stress the importance of human oversight and accountability as basics for the lawful use of AI in judicial settings (EUCPN, 2022).

Building upon these frameworks, this study explores the dual nature of AI in criminal justice as both an instrument of progress and a potential threat to procedural justice and human rights. It provides a systematic analysis of the evidentiary and legal implications arising from AI-based tools, focusing on issues of bias, transparency and the erosion of judicial discretion.

[2] 4.0. stands for the Fourth Industrial Revolution – a digital transformation that involves the intelligent networking of machines and processes through technologies like the Internet of Things (IoT), artificial intelligence (AI) and big data.

METHODS

This study employs a mixed-methods approach, combining comparative legal analysis, doctrinal review of case law and empirical case studies. It analyses the selected cases involving the use of AI tools in criminal proceedings, and an assessment of real-world AI applications for crime prediction and risk assessment, facial recognition software and automated forensic analysis. A legal-tech mapping of the AI platforms currently used in policing and adjudication complements the research. The goal is to identify best practices and regulatory gaps through an evidence-based lens. Additionally, the study employs legal analysis and explores the admissibility and credibility of AI-generated evidence in criminal trials, emphasising the need for demanding evidentiary standards and judicial oversight.

Through a combination of comparative analysis, empirical case study, and ethical evaluation, the research seeks to answer three core questions:

Q1: How does AI reshape evidentiary and procedural standards in criminal justice?

Q2: To what extent do predictive and risk-assessment tools align with the principles of fairness and the presumption of innocence?

Q3: What regulatory and ethical frameworks are necessary to ensure that AI serves justice rather than undermining it?

Ultimately, this paper contributes to the scholarly and policy debate on the responsible integration of AI in criminal law, particularly within the context of Southeast Europe – a region where digital transformation outpaces legislative reform. It argues that, while AI can substantially improve accuracy and efficiency, its uncritical adoption risks transforming criminal justice from a human-centred adjudicative process into a data-driven predictive mechanism, with different implications for the rule of law and the protection of human rights.

LITERATURE REVIEW

Research on the intersection of AI and criminal justice has developed rapidly over the past decade, reflecting broader debates about algorithmic governance, predictive analytics and human rights in the digital age. Early scholarship focused primarily on the efficiency and technological capacity of AI tools, but recent studies emphasise the normative and procedural implications of their use in law enforcement and courts. This section reviews the most influential theoretical frameworks and empirical findings, identifying the key research gaps that this paper addresses.

Theoretical Foundations

Several overlapping paradigms have shaped the theoretical debate surrounding AI in criminal justice. Ferguson (2017) introduced the concept of big data policing, describing how predictive algorithms transform traditional investigative methods into systems of continuous surveillance. Završnik (2020, 2021) expanded this perspective through the notion of algorithmic governance, which captures the technocratic shift from rule-based decision-making to data-driven probabilistic control. He warns that this transformation risks undermining fundamental legal principles such as due process, equality and human oversight. Similarly, in his PhD thesis, Diver (2019) invented the term 'computational legalism' to describe the epistemic shift in which prediction and control replace interpretation and deliberation, leading to what he calls "data-driven normativity".[3]

Complementing these theoretical contributions, Calo (2017) argues for a human-in-command model of AI governance, insisting that technological legitimacy in justice systems depends not only on accuracy but also on perceived fairness and moral accountability. His approach is in line with Gless et al. (2016), who examine questions of responsibility and culpability when AI systems contribute to harmful or discriminatory outcomes. These authors highlight the need to integrate normative and ethical dimensions into technological innovation.

Empirical and Policy-Oriented Studies

Empirical research has explored the practical application of AI in law enforcement, prosecution and adjudication. Grimm et al. (2021) analysed how AI-generated evidence challenges traditional evidentiary standards, especially regarding explainability and admissibility. Their findings underscore that algorithmic opacity, often referred to as the "black box problem",[4] hampers the right to challenge evidence under the principles of equality of arms and fair trial.[5] Parkkavi and Yadharthana (2024) examined risk-assessment algorithms such as COMPAS and Public Safety Assessment, revealing how predictive models can perpetuate the existing inequalities and reinforce automation bias among judicial actors.

At the policy level, the European Crime Prevention Network (EUCPN, 2022) and the Council of Europe's CEPEJ European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment (2018) established foundational guidelines for AI deployment in justice systems, emphasising transparency, non-discrimination and human oversight. The U.S. Department of Justice (2024) outlined these principles, calling for internal review boards and bias audits to ensure that AI serves justice rather than undermines it. More recently, Bhatt et al. (2024) and Borgesano et al. (2025) explored the role of AI under the paradigms of Justice 4.0 and Justice 5.0, respectively, highlighting that digital transformation can enhance efficiency only if it is grounded in explainability, human-centred design and continuous judicial training. Similarly, De Araújo et al. (2022) argue that digital-by-default courts can expand access to justice but must maintain constitutional guarantees of impartiality, publicity and due process.

[3] Computational legalism is a critical approach to "computational law" that analyses the potential negative consequences of translating legal and regulatory rules into rigid, immutable computer code.

[4] The risk of an incorrect or biased result is heightened when the AI's process is not understood. For example, the COMPAS risk assessment software was found to incorrectly and disproportionately flag Black defendants as high-risk.

[5] If a jury is presented with AI-generated evidence, there is a risk of unfair prejudice and confusing the jury, as studies show that audio-visual evidence strongly influences perception and memory.

Identified Research Gap

While the existing literature provides extensive insights into AI's technical and ethical dimensions, it remains fragmented across disciplines. Most studies focus on either technological capabilities or broad human-rights implications but rarely integrate these with the procedural logic of criminal justice. There is limited comparative research addressing how AI-driven predictive policing, algorithmic risk assessment and evidentiary automation collectively transform the normative foundations of criminal judgment. Moreover, few works examine how these transformations affect the balance between efficiency and fairness, especially in Southeast European jurisdictions where digitalisation outperforms legal reform.

This paper bridges this gap by offering a rights-based and procedure-centred analysis of AI in criminal justice. It synthesises theoretical insights from algorithmic governance and empirical findings from European and global contexts to evaluate how AI technologies reshape the principles of legality, accountability and human dignity. By bridging doctrinal, empirical and ethical perspectives, the study aims to advance the discourse on responsible AI governance in criminal law grounded in the rule of law and fundamental rights protection.

THE RISE OF ARTIFICIAL INTELLIGENCE IN CRIMINAL JUSTICE

The Evolution of AI Technologies

The increasing presence of artificial intelligence (AI) in criminal justice reflects a global movement toward digital governance and data-driven decision-making. Originally introduced as automated data-processing tools, AI systems have evolved into sophisticated predictive and analytical mechanisms capable of simulating certain forms of human reasoning. They rely on machine-learning algorithms, natural-language processing, neural networks and data-mining techniques to identify behavioural patterns, classify risks and assist in evidentiary evaluation (EUCPN, 2022).

AI's integration into justice institutions began with administrative tasks, digital case-management, database searches and forensic image analysis, but has advanced to autonomous decision-support mechanisms influencing core judicial and prosecutorial functions (Završnik, 2020). Modern courts now employ AI for multiple purposes: predictive policing, risk assessment, facial recognition, automated document generation and review, and even sentence recommendations. In some jurisdictions, AI-based chatbots inform defendants of procedural rights, while algorithmic tools assist prosecutors in case prioritization and evidence screening.

This evolution represents what Završnik (2020) terms the "algorithmic turn" in criminal justice, a transition from rule-based logic to data-driven governance. Decision-making increasingly depends on probabilistic reasoning and predictive analytics rather than traditional judicial interpretation. Although these innovations promise greater efficiency and consistency, they simultaneously shift the soul foundation of justice from normative reasoning to statistical inference (Završnik, 2020).

Hildebrandt (2020) describes this process as the emergence of "computational legalism", a governance model in which prediction and control risk replace deliberation and interpretation. She warns that legal systems must resist data-driven normativity in order to safeguard human autonomy and legality (Hildebrandt, 2020). Likewise, Strikwerda (2020) emphasises that algorithmic risk assessment, when applied to criminal justice, tends to blur the line between prevention and punishment, fostering a "pre-crime logic" that can undermine the presumption of innocence.

Recent scholarship situates this judicial digitalisation within the broader Industry 4.0 transformation of public administration, where AI works alongside the Internet of Things, cloud computing, blockchain and big-data analytics to tackle case backlogs, standardize procedures and enhance transparency. A 2024 systematic review in Humanities and Social Sciences Communications highlights concrete initiatives, such as Switzerland's Justitia 4.0 and Brazil's PJe, and argues that "high-risk" justice applications must include strong oversight aligned with the EU AI Act's risk-based approach emphasising human supervision, documentation, and accountability (Bhatt et al., 2024). Yet, the same literature cautions that efficiency gains could render justice "less humane" unless explainability and rights-protection are fixed by design (Bhatt et al., 2024). As Mugari and Obioha (2021) observe, these tensions between innovation and legality reflect global concerns about fairness and transparency in predictive policing and other algorithmic interventions.

From Automation to Decision Support

AI applications in criminal justice can be understood along a continuum ranging from automation, replacing repetitive human tasks, to decision support, assisting judicial discretion. Early systems merely digitised the existing processes, whereas the current models, especially those in predictive policing and risk assessment, directly influence legal outcomes.

Predictive-policing software analyses historical crime data to forecast potential "hot spots", enabling law-enforcement authorities to allocate resources proactively. Such systems operate in the United States (PredPol), the Netherlands (Crime Anticipation System – CAS) and the United Kingdom (Harm Assessment Tool). Although they have improved efficiency, critics warn that these programs reproduce structural bias when trained on tilted historical datasets (Završnik, 2020; EUCPN, 2022). Risk-assessment algorithms like COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) in the US and PSA (Public Safety Assessment) models evaluate the likelihood of reoffending and influence bail and sentencing decisions (Situmeang et al., 2024).

A leading example of AI integration in the judiciary is Estonia**,** where artificial intelligence has been introduced to handle minor civil disputes. Through its so-called "AI judge", the Estonian legal system resolves straightforward cases such as unpaid bills and simple contractual disagreements within a few hours, producing legally binding rulings. This approach demonstrates how automation can relieve human judges of routine tasks, allowing them to concentrate on more complex legal matters.

Similarly, the Netherlands provides valuable insights through its AI-assisted sentencing system in criminal justice. This system evaluates risk factors, including the probability of reoffending, and offers recommendations to judges. However, the Dutch model preserves human oversight, ensuring that AI serves only as a supportive tool rather than a substitute for judicial reasoning, with the final decision always remaining in human hands (Zhang, 2022).

Likewise, the Supreme Court of India's SUPACE platform (Supreme Court Portal for Assistance in Courts Efficiency) and China's Smart Court systems demonstrate increasing judicial reliance on AI for evidence review and even preliminary drafting of judgments. These examples reveal how algorithmic recommendations increasingly guide human decision-makers, often without full transparency regarding the underlying logic.

In the prosecutorial domain, AI supports case triage, digital-evidence sorting, and identification of relevant precedents. In forensic contexts, AI-based image and voice-analysis tools have accelerated criminal investigations, particularly in cybercrime and terrorism cases (Parkkavi & Yadharthana, 2024). Meanwhile, facial-recognition technologies, though valuable for identifying suspects, have provoked ethical debates about privacy, proportionality and racial misidentification (Grimm et al., 2021).

While automation undeniably increases efficiency, it also risks delegating moral and legal judgment to algorithmic processes. As Završnik (2020) and the U.S. Department of Justice (2024) emphasise, sustained human oversight and algorithmic explainability are essential to ensure that AI remains a tool of justice rather than an autonomous arbiter of guilt. Predictive-policing algorithms, in particular, have drawn criticism for reinforcing historical inequities. Ferguson (2017) argues that these systems frequently recycle biased datasets, effectively "predicting policing rather than crime". By relying on prior arrest patterns, they intensify surveillance of marginalised communities and generate feedback loops that spread discrimination (Ferguson, 2017).

Decision-making, once grounded in human interpretation, is increasingly mediated by data-driven inference, offering opportunities for improvement and risks of dehumanisation. As legal systems become more dependent on algorithmic reasoning, ensuring transparency, accountability and respect for fundamental rights becomes indispensable. The challenge is no longer whether AI should be used in criminal justice, but how it can be used responsibly without undermining the very principles upon which justice itself rests.

PREDICTIVE POLICING AND RISK ASSESSMENT TOOLS

Predictive Models in Law Enforcement

The emergence of predictive policing represents one of the most visible manifestations of artificial intelligence within criminal justice. These systems employ statistical models, historical crime data and geospatial information to forecast potential criminal activity, thereby allowing law enforcement agencies to allocate resources more efficiently and, in theory, prevent crime before it occurs (EUCPN, 2022). Well-known examples include PredPol[6] in the United States, the Crime Anticipation System (CAS) in the Netherlands and the Harm Assessment Tool in the United Kingdom.

[6] PredPol was an artificial intelligence (AI) software for "predictive policing" that used algorithms to forecast where and when property crimes were most likely to occur. It was a pioneer in the field but was discontinued in 2021 amid controversies over its accuracy and biased results.

Across Europe, predictive policing has been implemented in various forms. The Netherlands, Germany, Austria, France, Estonia and Romania have operational systems, while Luxembourg, Portugal and Spain continue to explore their potential use. Most European programs focus on domestic burglary and vehicle theft prevention. The Netherlands stands out as the first country to adopt predictive policing nationwide through the Crime Anticipation System CAS, which integrates demographic and socioeconomic data from multiple public databases to produce heat maps that visualise areas with elevated crime risk (Strikwerda, 2020). Germany's PreCobs (Pre-Crime Observation System) system employs five years of historical burglary data to forecast incidents (Gerstner, 2018), while Austria and France use long-term crime statistics and geospatial mapping to identify property-crime hotspots. Estonia's approach is broader, integrating event-based, area-based and person-based prediction models using data from criminal records, border-crossing statistics and traffic-fatality reports. Romania's pilot models similarly assess area-based and individual risks, demonstrating the EU's regional convergence toward data-driven policing (Mugari & Obioha, 2021; Hardyns & Rummens, 2018).

While predictive policing has enhanced police efficiency and situational awareness, it remains controversial for reinforcing structural bias and discriminatory practices. Algorithms trained on incomplete or biased data tend to replicate social inequalities, disproportionately targeting marginalized or low-income populations (Parkkavi & Yadharthana, 2024). Ferguson (2017) argues that such systems "predict policing rather than crime", as they rely heavily on prior arrest data rather than independent indicators of criminality, thereby perpetuating surveillance loops and stigmatisation. Završnik (2020) similarly describes predictive policing as embodying a "technocratic illusion", the mistaken belief that technological objectivity can eliminate human bias. In reality, algorithmic models frequently insert historical prejudice into digital form, and their cloudy "black-box" design obstructs both judicial scrutiny and democratic oversight.

From a legal perspective, predictive policing raises fundamental questions about due process, privacy and proportionality. When an individual becomes the subject of state surveillance solely based on algorithmic inference, constitutional guarantees, such as personal liberty and the presumption of innocence, are jeopardised (Završnik, 2020; EUCPN, 2022). These risks have prompted several European jurisdictions to introduce stronger oversight mechanisms, including requirements for algorithmic transparency, public accountability, and data-protection impact assessments (Gstrein et al., 2019).

Empirical studies on digital-justice platforms further demonstrate that successful AI integration depends not only on technological capacity but also on institutional reform. A 2021 IEOM study on Indonesia's administrative courts found that e-litigation systems improved efficiency and transparency but required continuous training and human oversight to maintain legitimacy (Sari et al., 2021).

In Europe, predictive profiling continues to evolve within diverse legal and cultural contexts. Germany's PreMap and PreCobs systems employ near-repeat theory to forecast residential burglary (Gerstner, 2018), while the Netherlands' CAS integrates statistical and socioeconomic indicators to identify "hot zones". The United Kingdom has moved toward internally developed tools, including the Gang Matrix and the National Data Analytics Solution (NDAS), designed to assess the likelihood of violent behaviour (Amnesty International, 2018; Jansen, 2018). These developments demonstrate the continent's growing adoption of algorithmic profiling but also highlight enduring concerns about privacy, fairness and the proportionality of automated decision-making (Couchman, 2019; Mugari & Obioha, 2021). Scholars like Zedner (2007) emphasise that such predictive models risk transforming criminal law from a backward-looking system of culpability into a forward-looking mechanism of risk control.

Risk Assessment in Judicial Decision-Making

AI-based risk-assessment systems represent another major frontier of algorithmic governance in criminal justice. Designed to estimate the likelihood of reoffending, these systems guide judicial decision-making regarding bail, sentencing and parole (Grimm et al., 2021). Among the most widely used is Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), developed in the United States and implemented across several state jurisdictions. The algorithm calculates risk scores using variables such as age, gender, criminal history, employment status and social background.

Although such tools were introduced to promote consistency and objectivity, several landmark cases have exposed systemic limitations. In State v. Loomis (2016), the Wisconsin Supreme Court upheld the use of COMPAS while simultaneously acknowledging concerns about explainability and fairness. Likewise, in Kansas v. Walls (2020), defence attorneys argued that algorithmic assessments effectively transferred judicial discretion to private software developers (Grimm et al., 2021). Although Walls was sentenced to presumptive probation, the district court relied on the LSI-R's finding that he was a "high-risk, high-needs" candidate to place him on a more highly supervised form of probation. Walls was given a summary page of his LSI-R scores but was denied access to the full assessment, including his specific answers.

Similar systems are now used beyond the United States. The Public Safety Assessment (PSA) in the United Kingdom, India's Supreme Court Portal for Assistance in Courts Efficiency (SUPACE) and China's Smart Court initiative all employ AI to evaluate risk factors and assist judges. While such systems claim to improve efficiency, critics argue that they foster automation bias, the human tendency to over-trust algorithmic outputs (Parkkavi & Yadharthana, 2024). Even small standardisation errors can produce severe consequences, particularly in pre-trial detention and sentencing, where inflated risk scores may result in unjustified incarceration and undermine individualised justice (Završnik, 2020).

The U.S. Department of Justice (2024) underscores that risk-assessment tools should serve as decision-support mechanisms, not as replacements for judicial reasoning. Maintaining human oversight, algorithmic auditing, and data-source transparency remains essential to align these systems with fundamental rights.

Ethical and Legal Implications

Predictive and risk-assessment algorithms reveal the delicate balance between efficiency and fairness in modern justice systems. While AI can minimise human error and enhance consistency, it may also erode procedural safeguards and human discretion. Regulatory frameworks must therefore prioritise transparency, accountability and explainability while equipping judges and prosecutors with the skills to interpret algorithmic outputs responsibly (Gstrein et al., 2019; Grimm et al., 2021).

Algorithmic governance should function as a complementary mechanism, enhancing, not substituting, human adjudication. As Calo (2017) argues, legitimacy in justice derives not only from accuracy but from citizens' belief that decisions remain rooted in human moral judgment. As AI integration accelerates, ensuring transparency, accountability, verifiability and human oversight will be crucial to preventing the erosion of fundamental rights in the pursuit of technological progress.

AI AND EVIDENTIARY CHALLENGES IN CRIMINAL PROCEDURE

Algorithmic Evidence and Admissibility

The growing use of AI in criminal proceedings has disrupted traditional evidentiary paradigms. As algorithms begin to generate or analyse key forms of evidence, courts must confront new questions of authenticity, admissibility and credibility. The central challenge is determining whether algorithmic outputs, such as pattern recognition, data correlations, or forensic predictions, can satisfy the same evidentiary standards historically applied to human testimony and expert opinion (Grimm et al., 2021).

A persistent obstacle is the "black-box problem", where even developers cannot fully explain how a machine-learning model reaches its conclusions. This opacity undermines the adversarial principle because neither the prosecution nor the defence can meaningfully interrogate or contest the reliability of AI-generated evidence (Das &Rad, 2020). For instance, facial-recognition software may produce high-confidence matches based on faulty datasets, and risk-assessment tools may classify individuals as "high risk" without revealing their weighting parameters (Parkkavi & Yadharthana, 2024).

Traditionally, expert testimony has served as the vehicle for introducing scientific or technical evidence. Yet, accountability becomes diffuse when the "expert" is an algorithm. As Grimm et al. (2021) note, machine-generated outputs must not be presumed objective merely because they are automated. Courts must instead apply demanding examinations requiring independent validation, full disclosure of the algorithmic model, and documentation of data provenance, to ensure that such evidence meets the principles of relevance, competence and fairness.

Furthermore, evidentiary standards must adapt to the probabilistic logic underlying algorithmic reasoning. While legal doctrines centre on categorical proof, AI operates on likelihoods and predictive scores. This epistemic gap threatens the moral and legal threshold of beyond a reasonable doubt. We have to have in mind that "probability is not proof", and statistical certainty cannot replace normative judgment (König & Krafft, 2021). Goodman and Flaxman (2017) similarly oppose that neural-network opacity challenges admissibility itself, advocating for a legally enforceable right to explanation in any courtroom where AI influences outcomes. They argue that interpretability functions as a procedural safeguard against automation bias and a precondition for a fair trial.

Chain of Custody and Data Integrity

AI-driven evidence collection introduces further complexity concerning the chain of custody and data integrity. When algorithms collect, filter or transform digital evidence, they may alter its format or metadata in ways that are difficult to audit, raising doubts about authenticity and continuity (Situmeang et al., 2024; Ehsanpour, 2025). For example, automated image and voice-recognition systems used in cybercrime and terrorism investigations can process vast quantities of data but are prone to false positives, especially when dealing with diverse demographics or low-resolution material (U.S. Department of Justice, 2024). Once AI modifies digital evidence through enhancement, reconstruction or categorization, courts must verify that the altered data faithfully represent their source, otherwise evidentiary integrity is compromised.

To mitigate such risks, European legal frameworks have begun to introduce procedural safeguards. The Council of Europe's CEPEJ Ethical Charter on AI in Judicial Systems and the EU Artificial Intelligence Act[7] both emphasise human oversight, traceability and documentation throughout data processing. These instruments require that AI-assisted evidence remain independently verifiable and subject to adversarial review (EUCPN, 2022). A parallel debate concerns whether AI tools should be open-source or at least subject to judicial disclosure obligations. Without insight into the algorithmic logic, defence attorneys face substantial barriers to exercising the right to contest evidence, an essential component of procedural justice (Grimm et al., 2021). Thus, transparency must be understood not as a technical preference but as a constitutional necessity for maintaining due process.

Judicial Discretion and Human Oversight

The increasing reliance on algorithmic evidence evaluation raises concerns about diminishing judicial discretion. Judges may unconsciously treat algorithmic outputs as objective truth, falling victim/defendant to automation bias, the tendency to over-trust machine-generated information (Parkkavi & Yadharthana, 2024). If left unchecked, such bias risks converting courts into passive ratifiers of 'computational determinations'.[8] However, genuine human oversight remains achievable through judicial literacy, procedural safeguards and mandatory algorithmic audits. As the U.S. Department of Justice (2024) emphasises, algorithms must function as assistive instruments, not as autonomous decision-makers. Training programs educating judges on interpreting algorithmic reasoning in conjunction with disclosing of error rates and limitations can help maintain equilibrium between technological efficiency and procedural fairness.

Ultimately, evidentiary standards in the era of AI must reaffirm human accountability. Grimm et al. (2021) stress that no matter how advanced algorithms become, the legitimacy of judgement depends on human responsibility and the capacity to justify, explain, and morally defend decisions. A doctrinal critique from Brazil's "Law 4.0" debate reinforces this point: automation in judging, evidence review and biometric identification threatens due

[7] The EU Artificial Intelligence (AI) Act is the world's first comprehensive legal framework to regulate AI, aiming to foster trustworthy, safe and transparent AI systems. The regulation took effect on August 1, 2024, with various provisions and obligations phasing in over a multi-year period.

[8] "Computational determination" refers to using computer-based methods, algorithms and simulations to find a specific value, solve a problem or predict a result.

process whenever exclusive or non-explainable models are used. Da Costa-Abreu and Silva (2020) advocate mandatory disclosure of algorithmic logic, data provenance, and documented error rates whenever AI outputs inform legal conclusions. Only through transparency and explainability can AI coexist with the adversarial principle and equality of arms.

HUMAN RIGHTS IMPLICATIONS

The Presumption of Innocence and the Right to a Fair Trial

The presumption of innocence lies at the core of every democratic legal system, protecting individuals from arbitrary state power and premature moral judgment (Buzarovska et al., 2015). However, the increasing integration of AI into criminal justice threatens to blur this fundamental safeguard. Predictive algorithms and risk-assessment models routinely generate probabilistic scores that pre-judge an individual's potential for criminal behaviour, thereby undermining the assumption of innocence until proven guilty.

When predictive policing systems classify individuals or neighbourhoods as "high risk", such designations often influence how law enforcement interacts with them. In judicial contexts, algorithmic risk scores, produced by systems like COMPAS or the Public Safety Assessment, may shape bail and sentencing decisions before defendants have a full opportunity to present their case (Grimm et al., 2021). This shift from evidentiary evaluation to probabilistic reasoning embodies what Završnik (2020) calls the "probabilistic turn" in criminal law.

The right to a fair trial, enshrined in Article 6 of the European Convention on Human Rights (ECHR), presupposes equality of arms between the prosecution and defence (Kosevaliska, 2015). Yet, when the defence lacks access to the algorithm's code, training data or error rates, the ability to challenge such evidence becomes merely formal rather than substantive (EUCPN, 2022). As Parkkavi and Yadharthana (2024) note, transparency and explainability are not optional ethical ideals but procedural requirements for safeguarding fairness. Without them, algorithmic judgement risks replacing reasoned deliberation with automated interpretation, threatening the very legitimacy of justice.

The Right to Privacy and Data Protection

AI systems in criminal justice are driven by massive datasets that include biometric identifiers, communication logs and behavioural records. Their use frequently entails continuous surveillance, facial recognition, and predictive monitoring, each directly engaging the right to privacy protected under Article 8 of the ECHR (Council of Europe, 1950) and Article 17 of the ICCPR (United Nations, 1966, Art. 17). Facial-recognition technologies and similar AI tools are extremely valuable for identifying suspects. However, they often operate without consent, store data indefinitely and exhibit disproportionate error rates for women and ethnic minorities (U.S. Department of Justice, 2024).

When used without strong oversight, these tools risk creating a regime of mass surveillance inconsistent with democratic accountability. The General Data Protection Regulation (GDPR) and the EU Artificial Intelligence Act seek to counter these dangers by imposing principles such as purpose limitation, data minimisation and human oversight. However, as Situmeang et al. (2024) emphasise, effective enforcement remains limited by insufficient institutional capacity and technical expertise within justice systems. Ensuring privacy protection in the era of AI thus requires not only legal safeguards but also judicial literacy, algorithmic auditing and interagency cooperation.

Equality and Non-Discrimination

Perhaps the most persistent human-rights challenge linked to AI in criminal justice is algorithmic bias. Machine-learning models trained on historical crime data inherently reflect the prejudices of those data, perpetuating systemic inequalities (Parkkavi & Yadharthana, 2024). Predictive-policing systems relying on arrest records from over-policed neighbourhoods often reinforce surveillance cycles, effectively criminalising social vulnerability.

Risk-assessment algorithms likewise reproduce structural hierarchies. Grimm et al. (2021) observe that variables such as income, education and neighbourhood act as proxies for race and class, producing outcomes that disproportionately affect marginalised groups. Ferguson (2017) captures this phenomenon concisely: predictive models frequently "predict policing rather than crime", generating feedback loops that sustain discrimination. In regions such as Southeast Europe, where social disparities and weak oversight persist, the risk of digital discrimination is particularly acute (Kosevaliska et al., 2024; Kosevaliska, 2023).

A further manifestation of algorithmic inequality arises from ethnic and racial profiling embedded in AI-based mass surveillance systems. Automated License Plate Recognition (ALPR) technologies, facial recognition networks and border-control algorithms increasingly rely on datasets that misrepresent non-European populations. Studies indicate that vehicle-registration surveillance systems frequently yield false positives for non-European license plates, flagging them as suspicious or irregular (Bennett Moses, 2023). Such systemic bias not only distorts enforcement priorities but also normalises disproportionate monitoring of ethnic minorities, migrants and cross-border workers. When combined with predictive-policing or immigration-control algorithms, these technologies risk constructing a digital architecture of suspicion where certain identities are algorithmically "pre-criminalized". Addressing these harms requires rigorous validation of input data, mandatory public disclosure of error rates, and participatory oversight that includes affected communities in algorithmic-governance processes.

Promoting equality in AI-driven justice requires transparent datasets, independent bias auditing and accessible mechanisms for redress. The U.S. Department of Justice (2024) recommends regular bias testing while European regulators advocate for a human-in-command model that ensures human review at every stage of AI deployment. Both frameworks reject the myth of technological neutrality and affirm that equality must be actively protected through law and ethics, not assumed through code.

Balancing Innovation with Rights Protection

The integration of AI into criminal justice is both inevitable and transformative, yet its legitimacy depends on inserting human-rights principles into every stage of design and implementation. As Završnik (2020) and Grimm et al. (2021) argue, technology must serve justice, not replace it. A rights-based approach should ensure that efficiency never conceals fairness and that automation never erodes accountability.

This equilibrium requires a shift from reactive regulation to proactive governance: regular algorithmic audits, interdisciplinary ethics committees and international cooperation in standard-setting. De Araújo et al. (2022) emphasise that digital-by-default judicial models such as Brazil's Juízo 100% Digital can enhance accessibility and speed only if accompanied by guarantees of publicity, impartiality and a reasonable time. Likewise, Gless et al. (2016) advocate for algorithmic accountability rooted in transparency, traceability and proportionality review, noting that regulation must be normatively grounded in fundamental rights.

Ultimately, preserving the human dimension of justice requires institutionalising oversight and public transparency. Set in ethical impact assessments, audit trails and judicial review within AI systems can ensure that technological innovation reinforces rather than replaces the rule of law.

CONCLUDING REMARKS
REGULATORY, ETHICAL AND POLICY FRAMEWORKS FOR RESPONSIBLE AI IN CRIMINAL JUSTICE

As artificial intelligence becomes increasingly embedded in criminal justice, legislators and international institutions are striving to ensure legality, transparency and protection of human rights. The European Union's Artificial Intelligence Act (2024) represents the most comprehensive regulatory framework to date. Following a risk-based model, it classifies AI systems used in law enforcement, border control and judicial decision-making as high-risk, subjecting them to strict requirements of documentation, explainability and human oversight (EUCPN, 2022). Complementing the Act, the Council of Europe's CEPEJ Ethical Charter on AI in Judicial Systems articulates five guiding principles: respect for fundamental rights, non-discrimination, quality and security, transparency and human control as prerequisites for legitimate AI deployment (Council of Europe, 2018). Harmonising international standards with national capacities demands not only legislative reform but also sustained investment in technical infrastructure, institutional competence and judicial training.

Regulation alone cannot safeguard justice in an algorithmic age; ethical and procedural measures must accompany every stage of design and deployment. Explainability and transparency should be mandatory for any AI system affecting legal rights. Algorithms must be auditable and open to scrutiny by judges and defence counsel, consistent with the CEPEJ's notion of "understandable justice" (Grimm et al., 2021).

Judicial literacy initiatives should train judges and prosecutors to interpret algorithmic evidence critically rather than defer mechanically to machine outputs. Moreover, accountability mechanisms must assign clear responsibility for AI deployment, monitoring and error correction.

Ethical governance also requires public transparency and participation. Citizens must be informed of how AI affects their rights and have access to remedies when those rights are violated. Without democratic oversight, AI regulation risks devolving into technocratic self-governance, eroding the rule of law (Završnik, 2021). Recent scholarship on Justice 5.0 envisions participatory and human-centric digital courts that integrate explainability, bias auditing and capacity-building as core design principles (Borgesano et al., 2025). Bhatt et al. (2024) likewise warn that efficiency gains from digital transformation must not render justice "less humane" but should instead reinforce transparency and accountability.

Comparative analysis confirms that the European AI Act and the CEPEJ Charter together represent the most advanced rights-based framework globally, balancing innovation with fundamental freedoms. Nonetheless, many Southeast European jurisdictions lack the institutional capacity to operationalise these standards effectively (Situmeang et al., 2024). This asymmetry creates "regulatory blind spots" where AI tools operate without adequate legal supervision.

The findings of this research underline that the future of criminal justice depends on a rights-based, human-centred governance model. AI must remain an instrument serving justice, not a substitute for it. As Calo (2017) argues, legitimacy arises not only from procedural accuracy but also from citizens' trust that decisions are humanly reasoned. De Araújo et al. (2022) similarly stress that digital-by-default courts, such as Brazil's Juízo 100% Digital, can enhance accessibility only if constitutional guarantees of publicity, impartiality and reasonable time are preserved.

Ultimately, sustainable digital justice requires embedding ethical reflection into both technological design and regulatory practice. AI governance should combine legal enforceability with moral responsibility, an equilibrium where innovation advances efficiency while the rule of law continues to safeguard dignity, equality and liberty.

Artificial intelligence is reshaping the architecture of criminal justice at every level, from police investigations and prosecutorial discretion to judicial deliberations and sentencing. However, this transformation presents a fundamental moral and legal dilemma: how to connect the power of intelligent technologies without compromising the rule of law and the core principles of justice.

The findings of this paper demonstrate that the integration of AI into criminal procedure, while potentially beneficial for efficiency and consistency, carries profound implications for fairness, transparency and human accountability. While predictive policing and risk-assessment algorithms have shown the capacity to optimise resource allocation, they risk reinforcing the existing social biases and eroding the presumption of innocence. Likewise, AI-generated evidence and automated decision-support systems challenge traditional evidentiary standards and judicial discretion, creating new demands for explainability, traceability and oversight.

Ensuring that AI serves justice, rather than undermines it, requires a multidimensional response. First, regulatory alignment must embed transparency, human oversight and rights-based evaluation into all stages of AI design and implementation. Second, ethical governance should cultivate institutional responsibility, promoting collaboration between developers, policymakers and judicial actors to ensure accountability and non-discrimination. Third, judicial empowerment through education and algorithmic literacy is vital to preserving the integrity of human reasoning in an age of automation.

In the context of Southeast Europe, where digital transformation often outperforms legal reform, these imperatives are particularly urgent. Building resilient institutions that can balance innovation with rights protection will determine whether AI becomes a tool of empowerment or a mechanism of exclusion.

Ultimately, the future of criminal justice depends on reaffirming a simple truth: technology must remain subordinate to the rule of law. Artificial intelligence should serve as an instrument of human progress, not as a substitute for human judgment. Only by embedding ethical reflection and human oversight into technological design and regulatory practice can societies ensure that the evolution of justice remains aligned with its most essential mission, to uphold truth, protect liberty and safeguard human dignity.

ACKNOWLEDGEMENTS

Language editing assistance was provided using Grammarly (AI-based writing support tool).

REFERENCES

Amnesty International. (2018). *Trapped in the Matrix: Secrecy, stigma, and bias in the Met's Gangs Database*. Amnesty International United Kingdom Section.

Bennett Moses, L. (2023). Oversight of police intelligence: A complex web, but is it enough? *Osgoode Hall Law Journal*, *60*(2), 289–324. https://ssrn.com/abstract=4248480

Bhatt, H., Bahuguna, R., Swami, S., Singh, R., Gehlot, A., Akram, S. V., Gupta, L. R., Thakur, A. K., Priyadarshi, N., & Twala, B. (2024). Integrating industry 4.0 technologies for the administration of courts and justice dispensation: A systematic review. *Humanities & Social Sciences Communications*, *11*, 1076. https://doi.org/10.1057/s41599-024-03587-0

Borgesano, F., De Maio, A., Laghi, P., & Musmanno, R. (2025). Artificial intelligence and justice: A systematic literature review and future research perspectives on Justice 5.0. *European Journal of Innovation Management*, *28*(11), 349–385. https://doi.org/10.1108/EJIM-01-2025-0117

Buzarovska-Lazetić, G., Kalajdžiev, G., Misoski, B., & Ilik, D. (2015). *Criminal procedure law*. Justinianus Primus Law Faculty.

Calo, R. (2017, August 8). *Artificial intelligence policy: A primer and roadmap*. SSRN. https://doi.org/10.2139/ssrn.3015350

Couchman, H. (2019). *Policing by machine: Predictive policing and the threats to our rights*. Liberty.

Council of Europe. (1950, November 4). *Convention for the protection of human rights and fundamental freedoms, as amended by Protocols No. 11 and 14*. https://www.echr.coe.int/documents/convention_eng.pdf

Council of Europe. (2018, December 4). *European ethical charter on the use of artificial intelligence in judicial systems and their environment*. European Commission for the Efficiency of Justice (CEPEJ). https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c

Da Costa-Abreu, M., & Silva, B. (2020). A critical analysis of 'Law 4.0': The use of automation and artificial intelligence and their impact on the judicial landscape of Brazil. *Revista de Direitos Fundamentais e Tributação*, *1*(3), 1–16. https://shura.shu.ac.uk/27336/

Das, A., & Rad, P. (2020). *Opportunities and challenges in explainable artificial intelligence (XAI): A survey*. arXiv (arXiv:2006.11371v2). Cornell University. https://doi.org/10.48550/arXiv.2006.11371

De Araújo, V. S., de Paiva, A., & Porto, F. R. (2022). O futuro da Justiça e o Mundo 4.0 [The future of justice and the world 4.0]. *Revista do Ministério Público do Estado do Rio de Janeiro*, (84), 207–231.

Diver, L. E. (2019). *Digisprudence: The affordance of legitimacy in code-as-law* [Doctoral dissertation, University of Edinburgh]. https://era.ed.ac.uk/handle/1842/36567

Ehsanpour, S. R. (2025). The importance and role of artificial intelligence in crime prevention. *Applied Criminology Research*, *3*(7), 59–80. https://doi.org/10.22034/aqcr.2025.2054758.1053

European Crime Prevention Network (EUCPN). (2022). *Artificial intelligence and predictive policing: Risks and challenges*. https://eucpn.org/sites/default/files/document/files/PP%20%282%29.pdf

Ferguson, A. G. (2017). *The rise of big data policing: Surveillance, race, and the future of law enforcement*. NYU Press.

Gerstner, D. (2018). Predictive policing in the context of residential burglary: An empirical illustration on the basis of a pilot project in Baden-Württemberg, Germany. *European Journal for Security Research*, *3*, 115–138. https://doi.org/10.1007/s41125-018-0033-0

Gless, S., Silverman, E., & Weigend, T. (2016). If robots cause harm, who is to blame? Self-driving cars and criminal liability. *New Criminal Law Review: An International and Interdisciplinary Journal*, *19*(3), 412-436. https://doi.org/10.1525/nclr.2016.19.3.412

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation." *AI Magazine*, *38*(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Grimm, P. W., Grossman, M. R., & Cormack, G. V. (2021). Artificial intelligence as evidence. *Northwestern Journal of Technology and Intellectual Property*, *19*(1), 10–105. https://scholarlycommons.law.northwestern.edu/njtip/vol19/iss1/2/

Gstrein, O. J., Bunnik, A., & Zwitter, A. (2019). Ethical, legal and social challenges of predictive policing. *Católica Law Review, Direito Penal*, *3*(3), 77–98. https://ssrn.com/abstract=3447158

Hardyns, W., & Rummens, A. (2018). Predictive policing as a new tool for law enforcement? Recent developments and challenges. *European Journal on Criminal Policy and Research*, *24*, 201–218. https://doi.org/10.1007/s10610-017-9361-2

Hildebrandt, M. (2020). *Law for computer scientists and other folk*. Oxford University Press.

Jansen, F. (2018, May 7). *Data-driven policy in the context of Europe*. datajusticeproject.net. https://datajusticeproject.net/wp-content/uploads/2019/05/Report-Data-Driven-Policing-EU.pdf

König, P. D., & Krafft, T. D. (2021). Evaluating the evidence in algorithmic evidence-based decision-making: The case of US pretrial risk assessment tools. *Current Issues in Criminal Justice*, *33*(3), 359–381. https://doi.org/10.1080/10345329.2020.1849932

Kosevaliska, O. (2015). The 'equality of arms' in Macedonian criminal procedure. *SEEU Review*, *11*(1), 2015. 123–130. https://doi.org/10.1515/seeur-2015-0015

Kosevaliska, O. (2023) Ethnic and racial profiling in criminal cases (as part of non-discrimination). In: *Sustainability and Law – optional course*, Miskolc, Hungary. (Unpublished). https://eprints.ugd.edu.mk/32997/

Kosevaliska, O., Poposka, Ž., & Maksimova, E. (2024). Prevalence of hate crime and hate incidents in municipalities in North Macedonia. In G. Meško, S. Kutnjak Ivković, & R. Hacin (Eds.), *The UN Sustainable Development Goals and provision of security, responses to crime and security threats and fair criminal justice system* (pp. 91–121). University of Maribor Press. https://doi.org/10.18690/um.fvv.7.2024.4

Mugari, I., & Obioha, E. E. (2021). Predictive policing and crime control in the United States of America and Europe: Trends in a decade of research and the future of predictive policing. *Social Sciences*, *10*(6), 234. https://doi.org/10.3390/socsci10060234

Parkkavi, E., & Yadharthana, K. (2024). Artificial intelligence in criminal justice: Balancing efficiency with fairness and accountability. *Indian Journal of Integrated Research in Law*, *4*(6), 483–491.

Regulation (EU) 2024/1689. *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)*. European Parliament and Council. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Sari, E., Rahman, A., Saputra, J., & Bon, A. T. (2021, November 3–5). Optimising and digitalising the technology-based electronic justice in the 4.0 era: A judicial reform. *Proceedings of the International Conference on Industrial Engineering and Operations Management*, ID 39, 242–246. https://ieomsociety.org/proceedings/2021monterrey/39.pdf

Situmeang, S. M. T., Harliyanto, R., Zulkarnain, P. D., Nahdi, U., & Nugroho, T. (2024). The role of artificial intelligence in criminal justice. *Global International Journal of Innovative Research*, *2*(8), 1966–1981. https://doi.org/10.59613/global.v2i8.264

State v. Loomis: Wisconsin Supreme Court requires warning before use of algorithmic risk assessments in sentencing. (2017). *Harvard Law Review*, *130*(5), 1530. https://harvardlawreview.org/print/vol-130/state-v-loomis/

Strikwerda, L. (2020). Predictive policing: The risks associated with risk assessment. *The Police Journal: Theory, Practice and Principles*, *94*(3), 422–436. https://doi.org/10.1177/0032258X20947749

United Nations. (1966, 16 December). *International covenant on civil and political rights*. https://treaties.un.org/doc/treaties/1976/03/19760323%2006-17%20am/ch_iv_04.pdf

U.S. Department of Justice. (2024, December 3). *Artificial intelligence and criminal justice: Final report*. https://www.justice.gov/olp/media/1381796/dl

Završnik, A. (2020). Criminal justice, artificial intelligence systems, and human rights. *ERA Forum*, *20*, 567–583. https://doi.org/10.1007/s12027-020-00602-0

Završnik, A. (2021). *Algorithmic governance and governance of algorithms*. Edward Elgar Publishing.

Zedner, L. (2007). Pre-crime and post-criminology? *Theoretical Criminology*, *11*(2), 261–281. https://doi.org/10.1177/1362480607075851

Zhang, S. (2022). AI in China's judicial system: Efficiency at what cost? *Asian Journal of Law and Technology*, *15*(4), 210–214.