CAN AN ALGORITHM DECIDE HUMAN FREEDOM? ARTIFICIAL INTELLIGENCE, CRIMINAL JUSTICE AND THE PROTECTION OF FUNDAMENTAL RIGHTS IN ALBANIA AND EUROPE

Valbona HIKMET XHAFERAJ

Research paper | International Congress "From Research to Application" | University of Tetova | 18–19 November 2026

Abstract

Artificial intelligence (AI) is increasingly discussed as a means of supporting decisions in criminal justice, including the assessment of reoffending risks, evidential analysis and judicial research. Yet the apparent precision of an algorithmic score may obscure methodological uncertainty, unequal error rates and limitations on procedural rights. This paper asks under which legal and institutional conditions algorithmic risk assessment could be compatible with fundamental rights in Albania, taking European regulatory developments as a comparative benchmark. Using doctrinal legal analysis and a structured comparison of European standards and selected international research, the paper distinguishes prohibited individual criminal-risk prediction from potentially permissible high-risk applications, examines the requirements of human oversight and contestability, and develops a precautionary governance framework for Albanian criminal justice and probation practice. It argues that an algorithmic recommendation should never substitute for individualized reasoning, and that formal human involvement is insufficient without meaningful understanding, documentation and an effective opportunity to challenge the result. The contribution is a proposed set of minimum legal and organizational safeguards for any future consideration of such tools in Albania. The study is conceptual and normative; it does not claim to measure actual deployment or effectiveness of AI systems in Albanian institutions.

Keywords: artificial intelligence; criminal justice; algorithmic risk assessment; fundamental rights; probation; due process; Albania.

1. Introduction

The criminal justice system routinely makes decisions under uncertainty: whether a measure is proportionate, whether supervision should be intensified, or how to assess an individual's circumstances. AI-based systems promise to detect patterns in large datasets and standardize selected analytical tasks. Such promises are particularly attractive where institutions face heavy workloads. However, a predictive model trained on historical outcomes is not an impartial judge. It reflects choices about labels, sampling, data quality, objectives and acceptable errors (Barocas, Hardt, & Narayanan, 2023).

The core dilemma is not whether machines can calculate faster than human decision-makers. It is whether an institution may rely on an opaque or inadequately validated inference when the consequences can include deprivation of liberty, more restrictive supervision or stigmatization. This question becomes acute when a risk score is treated as objective evidence rather than as a contestable statistical estimate. The paper adopts the principle that criminal justice decisions must remain grounded in law, individualized facts and reasons open to scrutiny.

Albania provides a useful setting for forward-looking legal analysis. As a Council of Europe member and a state pursuing European legal approximation, it must consider both the safeguards of the European Convention on Human Rights and relevant European regulatory developments. This paper does not assume that Albanian courts or probation authorities currently deploy predictive AI for individual risk scoring. Instead, it asks what a responsible legal response should look like before any such deployment is contemplated.

2. Research question, objectives and method

The principal research question is: Under what legal, procedural and institutional conditions, if any, may algorithmic risk-assessment systems assist criminal justice decisions in Albania without undermining fundamental rights? Three subsidiary questions guide the analysis: (i) which AI uses are prohibited or designated high-risk under the EU Artificial Intelligence Act; (ii) what makes human oversight and the right to challenge an output meaningful; and (iii) what minimum safeguards would be required in Albanian courts and probation practice?
The study employs doctrinal analysis of the European Convention on Human Rights, the EU Artificial Intelligence Act (Regulation (EU) 2024/1689), and the European Commission for the Efficiency of Justice (CEPEJ) Ethical Charter. It complements this with critical examination of research on algorithmic fairness and recidivism prediction, including Dressel and Farid (2018). The comparison is normative, not empirical: no interviews, confidential case files, original dataset or field experiment were used. Accordingly, proposed safeguards should be tested through stakeholder consultation and impact assessment before operational adoption.

3. The conceptual limits of algorithmic prediction

An algorithmic risk score usually estimates the statistical association between observed characteristics and an outcome defined by the system designer. It does not establish that a particular person will commit a future offence. Furthermore, "recidivism" may be operationalized as rearrest, reconviction or reincarceration; these are different outcomes and can be influenced by policing, reporting and administrative practices. Predictive accuracy depends on the definition of the target, the population and the period of observation.
Error distribution matters as much as aggregate performance. A system can appear reasonably accurate overall while generating unequal false-positive rates across groups. In criminal justice, a false positive may expose a person to unnecessarily restrictive treatment; a false negative may create public-safety concerns. These costs cannot be resolved solely by technical optimization. They require transparent legal and ethical choices. Dressel and Farid (2018) challenged assumptions about the superior accuracy of a prominent recidivism prediction tool, illustrating why independent validation is necessary rather than presumed.
The appearance of numerical precision can also produce automation bias: a professional may defer to a score despite contradictory contextual evidence. A human signature at the end of a workflow is therefore not synonymous with human judgment. Oversight must include competence to question the model, discretion to depart from its output and a documented explanation for the actual decision.

4. European legal and ethical standards

The EU AI Act establishes a differentiated regulatory approach. Article 5(1)(d) prohibits certain AI systems that assess or predict an individual's risk of committing a criminal offence based solely on profiling or assessment of personality traits and characteristics; the provision contains a carefully framed exception for support of a human assessment already based on objective, verifiable facts directly linked to criminal activity. This prohibition must not be confused with a general ban on all algorithmic risk analysis. The Act also lists certain permitted law-enforcement and administration-of-justice applications as high-risk in Annex III, subject to their precise intended purpose and applicable law (European Union, 2024).
In particular, Annex III addresses specified uses in law enforcement, including certain assessments of offending or reoffending risk, and systems assisting judicial authorities in researching, interpreting and applying law to facts. High-risk classification does not itself

authorize a use: a separate lawful basis, purpose limitation and applicable procedural protections remain necessary. Because Albania is not an EU member state, the EU AI Act should not be described as automatically and generally directly applicable to its domestic institutions; it is an important comparative and approximation benchmark, with potential relevance depending on territorial and market-related scope.

The CEPEJ Ethical Charter (2018) offers five interconnected principles: respect for fundamental rights; non-discrimination; quality and security; transparency, impartiality and fairness; and user control. These principles are especially valuable because they treat technology as subordinate to judicial legitimacy rather than as an independent source of authority. The European Convention on Human Rights further provides the overarching context for liberty, fair trial, privacy and effective remedies, depending on the particular proceeding and interference.

Taken together, these sources support a distinction between assistance and substitution. A tool that helps organize non-sensitive information poses a different risk from one that ranks persons for restrictive decisions. The greater the effect on liberty or legal position, the stronger the need for legality, necessity, proportionality, verification, reason-giving and review.

5. Implications for probation and individualized justice

Probation supervision illustrates both the appeal and danger of predictive tools. Staff must assess dynamic circumstances, compliance, support needs and public-safety concerns. An appropriately designed digital tool could help organize appointments, identify missing administrative information or prompt a structured review. By contrast, automatically assigning a person to a restrictive category on the basis of an unexplainable score could undermine individualization and procedural fairness.

Crucially, risk and need are not interchangeable. A person's need for employment assistance, treatment or education should not automatically be interpreted as proof of dangerousness. Social disadvantage can be both a support need and a source of bias if proxies for poverty or neighborhood are used to predict criminality. Responsible probation practice requires attention to rehabilitation and reintegration, not merely surveillance and classification.

A future Albanian framework should therefore separate low-impact administrative tools from consequential risk-scoring systems. Before introducing the latter, an institution should demonstrate a clear legal basis, necessity compared with less intrusive alternatives, independent validation on relevant populations, a documented error profile, accessible explanation, trained human oversight, an individual challenge procedure and independent audit. A moratorium on operational use would be appropriate where these conditions cannot be demonstrated.

6. Proposed safeguards for Albania: a rights-first framework

This paper proposes a preliminary six-part governance test. First, legality and purpose: legislation or another sufficiently clear lawful basis must identify the permitted purpose, responsible authority and prohibited uses. Second, necessity and proportionality: the authority must explain why the intended benefit cannot reasonably be achieved by a less intrusive measure. Third, evidence and validity: data provenance, target definitions, local validation, performance limitations and subgroup errors must be independently examined.

Fourth, meaningful human control: authorized professionals must understand the tool's purpose and limitations, retain genuine discretion and record independent reasons for consequential decisions. Fifth, procedural contestability: affected persons and their legal representatives should receive an intelligible account of material algorithmic involvement and

a practical route to challenge errors, subject to narrowly justified lawful restrictions. Sixth, accountability: procurement documentation, access logs, incident reporting, periodic audits, data-protection review and clear institutional responsibility must be established.
These requirements should operate cumulatively, not as a checklist that can be satisfied by generic policy statements. A pilot, if legally permissible, should begin with non-consequential or simulated use, independent ethics and legal review, and pre-defined stopping criteria. Any later use affecting liberty should require stronger evidence and oversight. This proposal is a conceptual framework for debate and is not presented as a validated Albanian regulatory instrument.

7. Discussion

The main finding of this normative analysis is that algorithmic efficiency is not a sufficient justification for integrating AI into criminal justice. The legal question concerns the allocation of authority, burden of explanation and avenues of challenge. Where a score influences an adverse decision, it must be possible to identify the source of the recommendation, examine its relevant limitations and determine who remains accountable. Otherwise, automation risks transforming uncertainty into an apparently objective label that is difficult to contest.
A second implication is that importing a model developed elsewhere is especially problematic. Differences in legal definitions, institutional practice, population characteristics and data collection can undermine validity. Even a model that performs acceptably in one jurisdiction cannot be presumed appropriate for Albania. Moreover, a technically transparent system can still be legally unacceptable if its purpose is prohibited or its use disproportionate.
The study has important limitations. It does not evaluate a deployed Albanian AI system, quantify discrimination or claim that proposed controls will eliminate error. Further research should map the existing legal and technological environment, interview stakeholders with ethical approval, and test any proposed instrument through independent legal, statistical and human-rights assessment. These steps are prerequisites for moving from research to responsible application.

8. Conclusion

Can an algorithm decide human freedom? In a rights-based criminal justice system, the answer must be no: a statistical output cannot itself exercise legitimate judicial authority. Carefully bounded AI tools may assist human professionals, but only where their specific use is lawful, necessary, validated, explainable, contestable and accountable. The European framework provides a valuable set of warnings and design principles, not a shortcut to deployment. For Albania, the priority should be legal preparedness and institutional capacity before procurement or operational experimentation with consequential risk scoring. The measure of innovation is not how quickly an algorithm produces a recommendation, but whether the justice system remains capable of explaining, reviewing and correcting decisions that affect a human life.

References

[1] Barocas, S., Hardt, M., & Narayanan, A. (2023). Fairness and machine learning: Limitations and opportunities. MIT Press. https://fairmlbook.org/

[2] Council of Europe, European Commission for the Efficiency of Justice. (2018). European Ethical Charter on the use of artificial intelligence in judicial systems and their environment. https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment

[3] Council of Europe. (1950). Convention for the Protection of Human Rights and Fundamental Freedoms. https://www.echr.coe.int/european-convention-on-human-rights
[4] Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. Science Advances, 4(1), eaao5580. https://doi.org/10.1126/sciadv.aao5580
[5] European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj
[6] European Commission. (2026). AI Act Service Desk: Article 5, prohibited AI practices; Annex III, high-risk AI systems. https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-5