

Ethical use of AI in forecasting domestic violence recidivism

Helder Pinto & Isabel Fidalgo
ISTEC Porto

Executive Summary

The growing application of Artificial Intelligence (AI) to public safety, as in the project “AI for the Prediction of Recidivism in Domestic Violence”, offers transformative potential. However, the absence of a robust legal framework in Portugal for AI in criminal and judicial contexts raises concerns regarding discrimination, the violation of fundamental rights, and a lack of accountability. It is essential to establish clear guidelines and governance mechanisms before the widespread deployment of these systems.

This Policy Brief proposes recommendations to ensure that AI is used in an ethical, transparent, and accountable manner, protecting citizens and strengthening public trust in the justice system.

Recommendations

- Establish a robust legal and regulatory framework for the use of AI in public safety;
- Introduce a mandatory protocol for the independent validation and auditing of predictive models;
- Ensure transparency and explainability;
- Implement mandatory human oversight in all critical decisions;
- Ensure full compliance with the GDPR and data protection legislation.

Target Audience of the Policy Brief

Magistrates, the Public Prosecution Service, police authorities (GNR/PSP), the Commission for Citizenship and Gender Equality (CIG), Non-Governmental Organisations (NGOs), victims, and civil society.

Introduction and Problem Framing

Domestic violence remains a social vulnerability in Portugal, with a devastating impact on thousands of lives each year. The complexity of cases and the need to protect victims and prevent recidivism call for innovative approaches.

In this context, Artificial Intelligence (AI) emerges as a promising tool, capable of analysing large volumes of data to identify patterns and predict the risk of recidivism, thereby optimising the allocation of resources and the intervention of the authorities.

Nevertheless, the deployment of AI systems in areas as sensitive as public safety and the judicial system is not without challenges. These systems may inadvertently perpetuate or amplify existing biases, discriminate against individuals on the basis of protected characteristics, and influence judicial and operational decisions without the appropriate legal and ethical safeguards.

The project “AI for the Prediction of Recidivism in Domestic Violence in Portugal”, developed under Science4Policy 2025, is currently in the development phase, making it urgent to establish a robust legal and governance framework before its widespread deployment.

Problem Definition

Portugal faces a significant gap in its legal framework concerning the regulation of Artificial Intelligence in criminal and judicial contexts. At present, there is no specific legislation addressing the unique challenges posed by predictive AI in public safety. This absence creates fertile ground for substantial risks, including the potential discrimination of individuals on the basis of gender, ethnic origin, or social class, the violation of fundamental human rights, and a troubling lack of clarity regarding accountability in the event of algorithmic errors or biases.

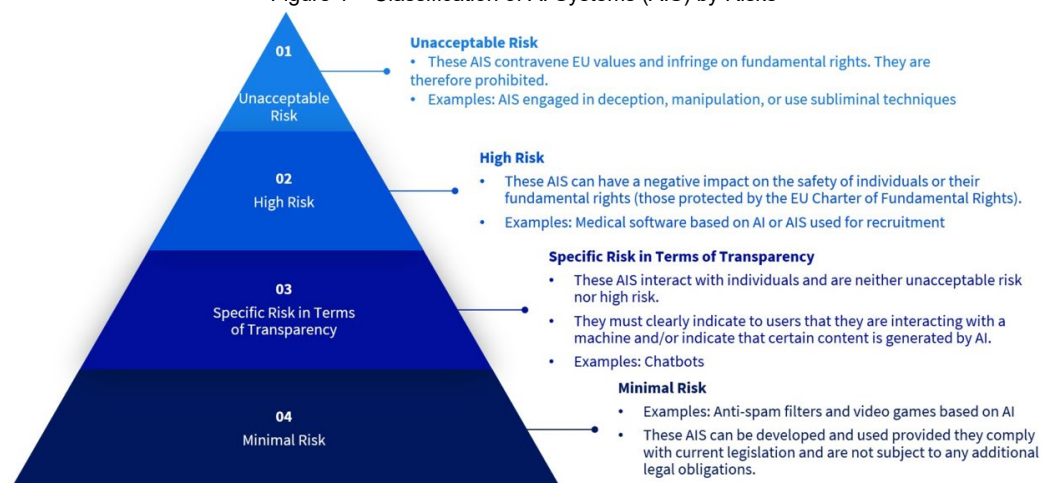
The impact of this gap cuts across a range of stakeholders. Magistrates and prosecutors require clear guidelines on how to integrate and interpret AI predictions in their decisions, since the absence of such guidance may compromise procedural fairness. Police forces operate in an environment of legal uncertainty, without defined protocols for the use of AI tools. Victims of domestic violence, whom these systems are intended to protect, may become even more

vulnerable should the technology fail or operate unfairly.

The policy relevance of this problem is undeniable, aligning with the National Strategy for Equality and Non-Discrimination and with the principles of the GDPR.

At the European level, the proposed AI Regulation (AI Act) already signals the urgency of addressing these issues, advancing the classification of high-risk AI systems, as shown in Figure 1, including those used in law enforcement and the administration of justice. The GDPR already establishes relevant data protection principles but does not specifically address the challenges of predictive AI in criminal contexts. Other countries, such as the United Kingdom and the United States, are likewise debating and implementing guidelines for the ethical and responsible use of AI in public safety, with a focus on transparency and accountability.

Figure 1 – Classification of AI Systems (AIS) by Risks



Source: https://ec.europa.eu/commission/presscorner/detail/en/qanda_21_1683

In Portugal, several critical gaps can be identified, such as:

- The **absence** of a clear and mandatory protocol for the **technical and ethical validation** of AI models prior to their use in sensitive contexts.
- A **lack of specific guidelines** on how to **integrate AI predictions into judicial decisions**, or on the evidential weight to be attributed to such analyses.
- The **chain of accountability for algorithmic errors or biases is ambiguous**, hindering the attribution of fault and the redress of harm.
- Although the **GDPR** is fundamental, it **is not sufficiently specific** to address the complexities of predictive AI in

criminal contexts, where sensitive data are processed to make inferences about future behaviour.

The technical risks are considerable. The historical data used to train these models may reflect and perpetuate social and institutional biases, resulting in algorithmic discrimination.

The **lack of explainability** in “black-box” models **prevents an understanding** of how decisions are reached, jeopardising the right to a fair hearing. Nonetheless, the opportunities are vast: if properly regulated, AI can significantly improve the protection of victims of domestic violence, reduce recidivism through more effective interventions, and optimise the scarce resources of the authorities.

Analysis & Key Findings

To address the question of governance and accountability in AI systems for public safety, three options are considered according to the level of risk:

Minimal Risk: Do Nothing (Status Quo)

This allows AI systems to be deployed without any specific regulation or clear guidelines, thereby ensuring the rapid adoption of new technologies.

Specific Risk: Light-Touch Regulation

This sets out non-binding guidelines and promotes self-regulation by the institutions that

develop and use AI systems. In this way, it affords flexibility for technological innovation, allowing organisations to adapt their practices in an agile manner.

High Risk: Robust Regulation

This designs and implements a comprehensive and binding legal framework that establishes clear requirements for the design, development, validation, deployment, and monitoring of AI systems in public safety.

Policy Options and Recommendations

For the Prediction of Recidivism in Domestic Violence in Portugal, classified as a high-risk system, the application of robust regulation is recommended.

Advantages

- Protects fundamental rights through binding requirements that prevent algorithmic discrimination and abusive surveillance;
- Establishes public trust and transparency through explainability mechanisms;

- Provides legal certainty to organisations, reducing uncertainty and facilitating investment;
- Promotes clear guidelines for responsible innovation;
- Mandates human oversight in all critical decisions;
- Ensures fairness, procedural justice, and institutional accountability.

Disadvantages and Mitigation

- Complex and time-consuming legislative development, involving legal

experts, technical specialists, and diverse stakeholders.

- A slowing of innovation, given that mitigation entails periods of

progressive compliance, allowing for gradual adaptation in which organisations test solutions under supervision.

Conclusion

Artificial Intelligence holds undeniable potential across a range of areas, for example in transforming the prevention of domestic violence in Portugal, offering tools for more effective prediction of recidivism and an optimised allocation of resources. However, this potential can only be fully realised if it is framed by robust governance and a clear legal framework. The urgency of establishing this regulatory structure is pressing. The absence of clear rules exposes citizens to the risks of discrimination and the violation of rights, and exposes institutions to legal uncertainty.

The adoption of a comprehensive and binding legal framework, encompassing independent validation, human oversight, and accountability mechanisms, is the only way to ensure that AI serves the public interest in an ethical and fair manner. This approach not only protects victims and guarantees equitable justice, but also fosters responsible innovation and public trust in the technology. The Government should lead a multidisciplinary consultation, involving legal experts, technical specialists, security forces, victims, and NGOs, with the aim of developing and implementing this legal framework, ensuring that Portugal positions itself at the forefront of responsible AI.

References

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on Artificial Intelligence (AI Act).

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (GDPR).

Law No. 32/2008 of 17 July, transposing into national law Council Framework Decision 2008/977/JHA of 27 November 2008 on the protection of personal data

processed in the framework of police and judicial cooperation in criminal matters.

Science4Policy 2025 project documents (available at [project URL, if applicable]).

National Strategy for Equality and Non-Discrimination 2018–2030 (ENIND).

Studies on algorithmic bias in predictive systems and their social impacts (e.g., the work of Joy Buolamwini and Cathy O’Neil).



HOW TO CITE THIS DOCUMENT

Pinto, H. & Fidalgo, I. (2025). *Ethical use of artificial intelligence in decision-making*. S4P-25 Policy Brief 8040/2025. PLANAPP – Centre for the Planning and Evaluation of Public Policies.

CONTACT

science4policy@planapp.gov.pt

COPYRIGHT

© PLANAPP, 2026



[Science for Public Policy](#)



[PLANAPP](#)



[Newsletter](#)



[PLANAPP](#)



[@planapp](#)



[PLANAPP podcasts](#)



PLANAPP



Instituto Superior
de Tecnologias
Avançadas do Porto



This policy brief was developed under Science4Policy 2025 (S4P-25): annual science for policy project calls, an initiative by the Centre for the Planning and Evaluation of Public Policies (PLANAPP) in collaboration with the Foundation for Science and Technology (FCT), financed by Portugal's Recovery and Resilience Plan. S4P-25 thematic line 12 – Use of Artificial Intelligence to predict recidivism in domestic violence.

The content is the sole responsibility of its authors and does not bind or commit PLANAPP nor FCT.