Artificial Intelligence in Brazilian Law Courts: risks and governance standards

Tainá T. C. dos Santos[1], **Renata Wassermann**[1], **Juliano Maranhão**[2]

[1]Instituto de Matemática e Estatística Universidade de São Paulo (IME-USP)
Rua do Matão, 1010 — CEP 05508-090 – São Paulo – SP

[2]Faculdade de Direito da Universidade de São Paulo (FDUSP)
Largo do São Francisco – São Paulo – SP

taina.turella@ime.usp.br, renata@ime.usp.br, julianomaranhao@usp.br

Abstract. *This paper examines the present application of Artificial Intelligence (AI) systems in Brazilian law courts, focusing on potential risks and ethical concerns associated with its integration. While AI has the potential to enhance productivity, its application in sensitive domains like criminal justice demands rigorous Verification and Validation (V&V) processes to mitigate biased outcomes. The study argues for classifying AI tools used in law enforcement as high-integrity systems and advocates adherence to standards such as IEEE 1012-2016. It underscores the need for comprehensive regulation and specialized training to address issues related to bias and privacy breaches, ensuring the responsible deployment of AI systems in the activities of law courts.*

1. Artificial Intelligence and its Applications

Technology profoundly influences modern society, shaping nearly every aspect of our lives. The impact of algorithms and automated systems varies depending on the context in which they are deployed. In some domains, the consequences of unregulated technology may be minor, but in others, the implications can be severe, specially when automated decisions or recommendations based on AI impact individual rights.

For example, AI-based scoring systems, providing risk indexes of default or accidents, may result in denial of a loan, or increase in the price of an insurance. Those decisions affect individuals' rights of access to goods or services and therefore should be deployed responsibly to ensure fairness, transparency, reliability, privacy and data protection. Transparency is particularly relevant since it enables contesting such automated decisions or decisions based on machine recommendations. Transparency is also instrumental to verify bias in the model or in the data, the compliance with privacy and data protection demands, the degree of accuracy and the very integrity of the system with respect to failures.

This concern is particularly present in applications of AI to support or enhance the efficiency of law courts. One may assume that applications of AI in law courts have the potential to impact individual rights since it is the very role of those institutions to decide about those rights. And such concern escalates in criminal courts, where the individual right to freedom is at stake [O'Neil 2016].

Given the variety of AI applications and the risks associated with it, one should conduct rigorous Verification and Validation (V&V) processes, tailored to each system's integrity level, to assess the potential impact of system failures. For instance, aircraft systems are classified as high-integrity software [Hatton 1995], while a local store's sales tracking algorithm is low-integrity. High-stakes systems like COMPAS (Corretional Offender Management Profiling for Alternative Sanctions) [O'Neil 2016], used in criminal justice, should also be considered high-integrity due to the significant harm they could cause, particularly if biased data affects their decisions.

2. AI in Brazilian Law Courts

In Brazil, where 60.8% of the incarcerated population is Black[1], algorithms that predict recidivism, determine sentencing, or influence prison terms must be critically assessed to avoid perpetuating existing social injustices. Research has consistently shown that structural biases—such as racism and misogyny—can be embedded in technology, raising concerns about algorithmic bias. Scholars like Safiya Noble [Noble 2018], Tarcízio Silva [Silva 2020], and Cathy O'Neil [O'Neil 2016] stress the importance of incorporating ethical considerations into computing, highlighting the need to address bias to foster more equitable outcomes.

AI's presence in Brazilian law courts is expanding, facilitating tasks such as case classification, court administration, transcription, decision drafting, and tax enforcement [Salomão 2022]. However, these AI tools risk entrenching societal biases, particularly when viewed as neutral by the public [Viezzer 2022], while unintentionally reinforcing pre-existing prejudices under the guise of objectivity [Almada and Zanatta 2024].

Generative AI systems like ChatGPT are increasingly being integrated into judicial workflows, providing legal professionals with quick access to information, synthesizing complex legal proceedings, and drafting decisions. However, these tools raise concerns about misinformation, data privacy risks from file uploads, and biased legal decisions. A study comparing human and AI-generated legal decisions found that while AI-produced outcomes were not poorly formulated, they lacked the depth and rigor typically found in human-authored judgments [Batista 2023].

While AI currently serves as a support tool for magistrates, aiding them in decision-making while maintaining human oversight, there remains a risk that exposure to biased data could influence harsher sentencing, especially for marginalized groups. A recent study [Maranhão 2024] on magistrates' views of AI in courts revealed that although they acknowledged the productivity gains, they also expressed concerns regarding the social and ethical implications of relying on AI in legal work.

3. Regulatory and Ethical Challenges

AI tools are being rapidly developed in Brazilian courts, with 66% of courts reportedly adopting some form of AI, amounting to around 147 tools designed primarily to extract statistical patterns. However, most of these systems lack the capacity for nuanced

[1] https://agenciabrasil.ebc.com.br/radioagencia-nacional/direitos-humanos/audio/2024-07/estudo-70-da-populacao-carceraria\-no-brasil-e-negra

legal reasoning. Despite this, there is a growing trend toward the informal adoption of Generative AI (GAI) tools within the legal domain [Maranhão 2024].

As previous reports [Maranhão 2024] highlight the legal risks of AI, this paper emphasizes that these concerns are equally critical from a technical perspective. While guidelines for AI—issued by organizations such as the Institute of Electrical and Electronics Engineers (IEEE), the National Institute of Standards and Technology (NIST), and the International Organization for Standardization (ISO)—exist, they are often not adhered to by developers and users of AI technologies. These guidelines should be applied based on risk assessments that consider the integrity of the software.

AI tools can improve court productivity, but especially in high-integrity systems, careful scrutiny is required. Any flaws in these systems could cause significant harm. Unlike other studies [Viezzer 2022], this paper stresses the need for rigorous V&V processes for AI systems classified as high-integrity software. It calls on civil society to demand adherence to the IEEE Standard for System, Software, and Hardware Verification and Validation (IEEE 1012-2016) [IEEE 2017] for software influencing legal decisions. Furthermore, robust regulation of AI in the public sector is urgently required.

4. Mitigating Risks and Potential Approaches

The deployment of AI decision-making tools requires vigilant oversight to ensure fairness and minimize risks. This oversight should extend to technologies used to assist in drafting decisions. Forensic tools, especially those imported from other legal systems, must be carefully examined in the context of Brazil's unique social complexities, with comprehensive public reports justifying their use.

In cases involving "unofficial" AI tools, the risks are even more severe. Bias is a critical issue, especially when tools like ChatGPT are used to draft decisions. Additionally, privacy concerns and potential breaches of confidential data pose serious risks. In the pursuit of productivity, professionals might inadvertently violate LGPD[2] regulations, risking the disclosure of sensitive information.

All AI tools involved in law enforcement and sentencing should be classified as high-integrity systems. Proper training for legal professionals using these tools is crucial to minimize risks. While AI can be an invaluable asset, it must be controlled and cautiously integrated, ensuring it is not seen as a replacement for human judgment, at least at this stage.

As AI regulation evolves, a comprehensive review is needed to establish governance frameworks for emerging technologies. Current personal data protection laws may be outdated, highlighting the need for updated regulations to prevent misuse by both public and private entities. Public authorities must remain accountable for protecting citizens' rights, while private entities should be regulated to avoid monopolizing personal data on a large scale.

5. Conclusion

AI has the potential to significantly enhance productivity and efficiency in law courts, but its integration requires careful and responsible management. Classifying AI tools

[2]General Personal Data Protection Law, in Portuguese, *Lei Geral de Proteção de Dados.*

used in law enforcement and sentencing as high-integrity systems is crucial due to their profound impact on individuals' lives. Implementing rigorous V&V processes, adhering to established standards like IEEE 1012-2016, and enacting comprehensive regulations are essential to ensure that AI applications are fair and just. Proper training for legal professionals is also critical to mitigate risks associated with misinformation, bias, and privacy breaches.

As Artificial Intelligence continues to advance, it should be regarded as a powerful aid rather than a substitute for human judgment, with stringent oversight to maintain justice and equity in society. It is encouraging that magistrates recognize the benefits of these tools while remaining cautious about their ethical implications. This paper does not seek to dismiss the positive aspects of introducing AI, including Generative AI, into the legal system. Instead, it advocates for a thorough evaluation of these tools to ensure their responsible and ethical use.

References

Almada, M. and Zanatta, R. A. F. (2024). Inteligência artificial, direito e pesquisa jurídica. *Revista USP*, 141:51–64.

Batista, C. B. L. (2023). Impactos da utilização de algoritmos e inteligência artificial no sistema de justiça criminal brasileiro. Technical report, Pontifícia Universidade Católica de Goiás. Undergraduate Thesis, Accessed: 2024-08-27.

Hatton, L. (1995). *Safer C: Developing Software for in High-Integrity and Safety-Critical Systems*. McGraw-Hill, Inc., USA.

IEEE (2017). IEEE Standard for System, Software, and Hardware Verification and Validation. Standard, IEEE Computer Society, Los Alamitos, CA.

Maranhão, J. (2024). O uso da inteligência artificial generativa no poder judiciário brasileiro. Technical report, Conselho Nacional de Justiça (CNJ). To appear.

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group, USA.

Salomão, L. F. (2022). Artificial intelligence: Technology applied to conflict management within the brazilian judiciary. Technical report, Centro de Inovação, Administração e Pesquisa do Judiciário, Fundação Getulio Vargas (FGV), São Paulo, Brazil.

Silva, T. (2020). *Comunidades, Algoritmos e Ativismos Digitais: Olhares Afrodiaspóricos*. LiteraRUA, São Paulo.

Viezzer, M. (2022). O uso da inteligência artificial pelo sistema jurídico brasileiro, classificação da inteligência artificial e análise de seu uso. *Revista Ibero-Americana de Humanidades, Ciências e Educação*, 8(1):1193–1213.