Journal of Electronic Research and Application, 2024, Volume 8, Issue 1
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Crossing the Achilles Heel of Algorithms: Identifying the Developmental Dilemma of Artificial Intelligence-Assisted Judicial Decision-Making

Kexin Chen*

School of Law, Macau University of Science and Technology, Macau, China

******Corresponding author:*** Kexin Chen, Delores-chen@foxmail.com

Abstract: In the developmental dilemma of artificial intelligence (AI)-assisted judicial decision-making, the technical architecture of AI determines its inherent lack of transparency and interpretability, which is challenging to fundamentally improve. This can be considered a true challenge in the realm of AI-assisted judicial decision-making. By examining the court's acceptance, integration, and trade-offs of AI technology embedded in the judicial field, the exploration of potential conflicts, interactions, and even mutual shaping between the two will not only reshape their conceptual connotations and intellectual boundaries but also strengthen the cognition and re-interpretation of the basic principles and core values of the judicial trial system.

Keywords: Artificial intelligence; Automated decision-making; Algorithmic law system; Due process; Algorithmic justice

Online publication: January 23, 2024

1. Introduction

In recent years, artificial intelligence has become a pivotal component of judicial decision-making on a global scale. Leveraging advancements in deep learning, machine learning, natural language processing, and other technological innovations, it has propelled the informatization of legal affairs beyond simple electronic retrieval to encompass more intelligent functions. These include class case push, deviation warning, danger assessment, sentencing assistance, sentence prediction, and more.

Despite the significant strides made, the application of artificial intelligence (AI)-assisted judicial decision-making evokes both optimism and pessimism. Among the challenges faced by current AI-assisted judicial decision-making systems in practical use, one that stands out as an essentially insurmountable "Achilles' heel" is the algorithmic black box. This paper seeks to delve deep into the exploration of this issue.

2. Legal black boxes and transparency

The 2016 case of State v. Loomis drew widespread attention to the controversy surrounding risk assessment algorithms in the United States. In this instance, the court imposed a severe sentence based on COMPAS' assessment that the defendant posed a high risk. Subsequently, the defendant moved to reopen the sentencing process, contending that the court's utilization of the COMPAS system violated constitutionally guaranteed due process rights [1].

One of the defendant's arguments was that the COMPAS system, protected by trade secrets, could not be disclosed for an assessment of its scientific accuracy [1]. The most straightforward solution for the "legal black box" issue would be to mandate the judiciary to disclose the original algorithm code when contracting an enterprise to develop an intelligent court system. However, considering that the trade secret clause safeguards developers' rights in highly specialized fields and encourages research and development in emerging technologies, mandatory disclosure might undermine the legitimate interests of algorithm developers. If mandatory disclosure poses a threat to these rights and interests, it should be carried out within reasonable limits, even for due process protection.

Kroll and his colleagues suggested, from a technical standpoint, that the operational steps of algorithmic decision-making can be recorded through specific technical means. For example, zero-knowledge proofs can enable decision-makers to prove that system decisions meet specific requirements without revealing all the contents of cryptographic commitments. Alternatively, cryptographic commitments can be adapted to prove algorithmic decision-making's compliance with specific requirements. In this context, a third party may keep sealed files or numbers to ensure a review of decision-making in a non-public manner, or software verification can mathematically ensure the existence of invariants in the system to guarantee a degree of accountability for the algorithm [2].

In contrast, Contini took a regulatory perspective, suggesting that additional expert committees could be established to review the compliance of algorithmic systems with legal norms, ensuring "qualified transparency" [3]. However, even if algorithms are entirely "transparent," they cannot meet the requirement of accountability based on the understanding of the intelligent court system operation logic. The "technological black box" inherently gives rise to a transparency paradox. In other words, algorithmic transparency does not equate to comprehensibility, let alone accountability [4].

3. Technical black boxes and interpretability

In the Loomis decision, the discussion regarding whether the court's use of the COMPAS system violated due process rights only brought attention to the "legal black box" aspect, neglecting the more pressing issue of the "technological black box" [1]. Indeed, the lack of interpretability stemming from the "technical black box" is the most challenging flaw in fundamentally improving algorithmic systems.

The most direct solution to address the interpretability gap in AI decision-making is to disclose the original program code, allowing for the dissection of the algorithm's logic. However, this approach holds little meaning for non-specialists who cannot glean valuable information by merely reading the code [5]. Moreover, it fails to alleviate the concerns of those affected by the decision regarding the algorithm's result reliability [2]. Additionally, due to the unique nature of deep learning and neural network algorithms, professionals in the field remain in the dark about how the system adjusts the weights of various factors and its operational logic [6]. Furthermore, as various reference factors within the algorithm may dynamically adjust at any time, there is no guarantee of result predictability, let alone repeatability [5].

The issue of algorithmic interpretability cannot be adequately addressed at the current level of existing technologies and specifications. On the technical front, explainable AI technology has gradually developed to alleviate this problem. Gleicher advocated intentionally constructing "interpretable models" in the future to intervene in the algorithmic selection process by controlling variables [7]. However, this approach faces an algorithmic design paradox: the intelligence and accuracy of algorithms cannot seamlessly coexist with interpretability. Only simple algorithms can maintain a high level of interpretability, while the interpretability of complex algorithms always falls short.

On the legal norms front, the General Data Protection Regulation (GDPR) introduced by the European Union (EU) in 2018 provides a more comprehensive norm for algorithm-related responsibility. Although it does not explicitly outline the "right to interpret" automated decision-making. Articles 13-15 and 22 of the GDPR address the issue. Article 13-14 stipulates that the data subject has the right to know the automated decision-making process and information related to expected results. However, the notification obligation is outlined before data collection, not after the decision results are formed, making it challenging to use as the basis of the right to explanation after the algorithm's decision-making results are generated.

Some scholars further argue that the right of access in Article 15 may serve as the jurisprudential basis for the right of interpretation. Still, since the point in time for exercising this power is not explicitly stated, it can be understood, at most, as the right of interpretation for the system's functionality rather than the basis for requesting interpretation in automated decision-making. Article 22, addressing a party's right to reject fully automated decision-making with a significant impact on its rights, is not essentially an affirmative claim but more akin to a passive right of defense. The "technological black box" has yet to provide a proper response to the interpretability problem it gives rise to.

4. How to address the impact of algorithmic black boxes on due process

The primary question that requires attention is whether the lack of transparency and interpretability resulting from the "algorithmic black box" is an issue that cannot be overlooked. Techno-optimists, such as Eugene Volokh, argued that when exploring the appropriateness of AI algorithms in judicial decisions, the focus should be on the outcomes rather than understanding the decision-making process. However, this view has faced opposition from some scholars. They contended that if the emphasis is solely on the results of smart court outputs rather than the process, it not only infringes upon the parties' right to contest adverse evidence but, more significantly, undermines the court's potential role in dispute resolution. The court plays a crucial role not only in soothing people's concerns but also in repairing relationships between parties and fostering reconciliation in society.

Given the critical importance of transparency and interpretability, and acknowledging the difficulty of resolving the issue at the technological source level, it might be feasible to address it through rights remedies. Firstly, the scope of due process rights violations resulting from AI-assisted judicial decision-making must be defined. According to Loomis's defense on appeal, due process rights were violated because COMPAS, safeguarded by trade secrets, prevented a challenge to the scientific validity of the risk assessment. In other words, the court encroached on its "right to interpret" based on State v. Skaff, established by the court. This right entails the defendant's ability to review the accuracy of the judgment and factors influencing it, including understanding the risk assessment algorithm and how danger assessment results were generated. AI fundamentally and directly affects the litigants' rights, especially their right to effective defense.

In the future, in criminal judgments, courts may be required to explicitly state whether they have considered decision-making aided by the risk assessment system and disclose specific risk factors analyzed by the system. For instance, the "Sentencing Trend Suggestion System" developed by the Judicial Yuan of Taiwan publicizes reference factors on its website. Taking Article 221 of the Criminal Law of Taiwan on the crime of compulsory sexual intercourse as an example, the system lists selected reference factors such as the perpetrator's age at the time of the act, previous cases of obstruction of sexual autonomy, confession consistency, and the relationship with the victim. Such transparency measures enhance the understanding and accountability of AI-assisted decision-making in the judicial context.

5. Conclusion

Given the inevitability of embracing the tide of technological development, confronting the impact of AI algorithms on the existing judicial system demands a nuanced approach. Treating AI technology is not a zero-sum game of weighing "advantages" against "shortcomings." Instead, the impact of AI technology on core values and the basic principles of the current judicial system requires careful examination and contemplation.

For instance, the lack of transparency and interpretability resulting from the algorithmic black box can significantly impact due process. Therefore, examinations should extend to understanding how the current judicial system accepts, integrates, and negotiates the trade-offs of AI technology. The goal is to ensure that human judges maintain trial independence and that the rights and obligations of the parties are effectively safeguarded. Simultaneously, science and technology may be harnessed to enhance the quality and efficiency of judicial proceedings, striving for a delicate balance and self-consistency within the system.

Disclosure statement

The author declares no conflict of interest.

References

[1] Liu H-W, Lin C-F, Chen Y-J, 2019, Beyond State v. Loomis: Artificial Intelligence, Government Algorithmization, and Accountability. International Journal of Law and Information Technology, 27(2): 122–141.

[2] Kroll JA, Huey J, Barocas S, et al., 2017, Accountable Algorithms. University of Pennsylvania Law Review, 165(3): 633–705.

[3] Contini F, 2020, Artificial Intelligence and the Transformation of Humans, Law and Technology Interactions in Judicial Proceedings. Law, Technology and Humans, 2(1): 4-18. https://doi.org/10.5204/lthj.v2i1.1478

[4] Schönberger D, 2019, Artificial Intelligence in Healthcare: A Critical Analysis of the Legal and Ethical Implications. International Journal of Law and Information Technology, 27(2): 171–203. https://doi.org/10.1093/ijlit/eaz004

[5] McGregor L, Murray D, Ng V, 2019, International Human Rights Law as a Framework for Algorithmic Accountability. International and Comparative Law Quarterly, 68(2): 309 – 343. https://doi.org/10.1017/S0020589319000046

[6] Ananny M, Crawford K, 2016, Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability. New Media & Society, 20(3): 973–989. https://doi.org/10.1177/1461444816676645

[7] Gleicher M, 2016, A Framework for Considering Comprehensibility in Modeling. Big Data, 4(2): 75–88. http://doi.org/10.1089/big.201$^{6.0007}$

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.