Vol. 14 (2026): Proceedings of the 2nd International Conference on Artificial Intelligence, Modern Engineering and Environmental Sustainability (IC-AIMEES 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/ ic-aimees.20260347353

# Artificial Intelligence in Judicial Decision-Making: Risks, Due Process, and Governance Frameworks

**Yi Liu***

*Tianjin Foreign Studies University, Tianjin, China*

**Corresponding author: Yi Liu.*

## Abstract

Artificial intelligence (AI) is increasingly being integrated into judicial decisions worldwide, influencing outcomes in bail, sentencing, and parole determinations. This review examines the application of AI in judicial practices and the potential risks associated with its use in the courts of the United States, China, and Europe. An incident involving AI-generated cases at the Beijing Tongzhou District People's Court in 2025 highlights the threat that algorithmic errors and AI hallucinations could pose to judicial proceedings. The primary risks identified include algorithmic bias, lack of transparency (the black box problem), erosion of due process, and the potential weakening of judicial discretion. Existing legal frameworks, from experiential oversight in the United States to protections under Europe's General Data Protection Regulation (GDPR), remain insufficient in addressing these challenges. Synthesizing literature on AI applications in judicial practice, algorithmic fairness, and due process, this article proposes a multi-level governance framework involving individual judges, court institutions, professional associations, and legislators. This framework aims to provide ethical and procedural guidance for the use of AI in the judiciary, ensuring that technological advancements enhance efficiency without undermining fairness, accountability, and the rule of law. Future research should focus on the practical effects of AI in judges' decision-making, governance strategies across different jurisdictions, and the development of normative guidelines to safeguard justice in an increasingly automated judicial environment.

### Keywords

artificial intelligence, judicial decision-making, algorithmic bias, due process, legal technology

## 1. Introduction

Artificial intelligence is transforming judicial decision-making worldwide. In the United States, the COMPAS algorithm assesses recidivism risk for over one million defendants annually [1,2]. In China, intelligent court initiatives have been deployed across more than 3,500 courts [3]. In Europe, the GDPR provides a framework for automated decisions, though its judicial application remains contested [4]. The 2025 Beijing Tongzhou Court case, where a lawyer submitted an AI-generated fake case, reveals the potential risks of AI in the legal system [5]. This review paper examines how AI is being used in judicial decision-making across jurisdictions, identifies risks including algorithmic bias and opacity, and proposes governance frameworks. It synthesizes literature across AI and legal practice, algorithmic fairness, and due process in

automated decision-making. This paper addresses three core questions: (1) How is AI currently being used in judicial decision-making across different legal systems, and what are the key differences in national approaches? (2) What risks do AI systems pose to core judicial values such as due process, fairness, and accountability? (3) What governance frameworks can ensure that AI serves rather than undermines justice?

## 2. AI Applications in Judicial Decision-Making: A Cross-Country Analysis

### 2.1 United States: Algorithmic Risk Assessment and Judicial Discretion

In the United States, the most prominent application of AI in judicial decision-making is the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) algorithm, developed by Northpointe (now Equivant). COMPAS is used in several states to assess defendants' risk of recidivism, influencing bail determinations, sentencing recommendations, and parole decisions [1]. The algorithm processes approximately 1 million cases annually, making it one of the most widely deployed judicial AI tools globally [2]. The Wisconsin Supreme Court's decision in State v. Loomis represents the most significant judicial consideration of algorithmic tools in sentencing [6]. The court acknowledged concerns about COMPAS's opacity but upheld its use, emphasizing that judges retained discretion and that risk scores were only one factor among many . However, critics argue that this approach fails to address fundamental due process concerns, as defendants cannot meaningfully challenge the basis of algorithmic outputs [7]. The lack of clear standards for AI use in federal courts remains a significant gap, though recent discussions suggest courts may soon begin articulating guardrails for defensible AI deployment [7].

### 2.2 China: Intelligent Courts and Systemic Integration

China has pursued the most ambitious integration of AI into its judicial system. The intelligent court initiative, launched as part of the broader Digital China strategy, aims to deploy AI across all stages of judicial proceedings [2]. By 2025, AI systems were assisting judges in over 3,500 courts nationwide, supporting tasks including evidence analysis, sentencing recommendations based on similar cases, and judgment drafting [2]. The 2025 Beijing Tongzhou Court case highlighted both the promise and peril of this approach. A lawyer representing a client in a commercial dispute submitted an AI-generated case with the docket number Hu 01 Min Zhong 12345, which was later confirmed to be entirely fabricated [5]. Judge Zheng Jizhe discovered the hallucination when attempting to verify the citation, noting that the case number's consecutive integers were a telltale sign of AI generation [5]. The case was subsequently included in the China Court Case Database, with the court emphasizing that lawyers bear responsibility for verifying AI outputs. This incident reveals the risk of AI hallucinations entering judicial proceedings and the need for clear verification protocols.

### 2.3 Europe: Rights-Based Approaches and the GDPR Framework

European approaches to AI in judging have been shaped significantly by the General Data Protection Regulation (GDPR). Article 22 grants individuals the right not to be subject to decisions based solely on automated processing that produce legal effects or significantly affect them [8]. Recital 71 suggests that such decisions should be based on meaningful information about the logic involved, though the precise contours of this right to explanation remain contested [4]. The United Kingdom's Ministry of Justice has adopted a structured approach through its AI Action Plan for Justice (2025-2027), which outlines a three-year roadmap for responsible AI adoption. The plan emphasizes that AI should support rather than substitute human judgment, preserving the independence of judges and ensuring rigorous testing before deployment [8]. Similarly, the Victorian Law Reform Commission in Australia has recommended that AI use for judicial decision-making be prohibited entirely, citing risks to judicial independence and public trust [9]. Scholars have developed the concept of the human judge-in-the-loop, arguing that AI systems should operate as indispensable allies working in collaboration with human judges rather than replacements for them [1].

### 2.4 Comparative Analysis: Institutional and Cultural Determinants

According to Table 1, the differences in national approaches reflect deeper institutional and cultural factors. The United States' adversarial legal tradition and emphasis on judicial discretion have produced a reactive approach, with courts responding to algorithmic challenges on a case-by-case basis [7]. China's civil law

system and centralized governance structure have enabled systemic, top-down integration of AI, though transparency concerns persist [2]. Europe's rights-based framework, grounded in the GDPR and the European Convention on Human Rights, has produced a more cautious, principle-driven approach that emphasizes individual rights and procedural protections [1,4]. These differences suggest that effective AI governance must be tailored to specific legal and institutional contexts.

*Table 1: Cross-Country Comparison of AI in Judicial Decision-Making.*

| Country/Region | Primary Applications | Scale of Deployment | Key Risks Identified | Legal/Regulatory Response |
|---|---|---|---|---|
| United States | COMPAS risk assessment for bail, sentencing, parole | ~1 million cases annually [3] | Racial bias, opacity, lack of challenge mechanisms | State v. Loomis (judicial discretion approach); no federal AI-specific legislation [5,6] |
| China | Intelligent courts: evidence analysis, sentencing recommendations, judgment drafting | 3,500+ courts [2] | AI hallucinations, verification failures, transparency concerns | Beijing Tongzhou Court case (lawyer verification duty); administrative guidance [5] |
| Europe/UK | Limited deployment; focus on ODR and administrative support | Pilot programs; GDPR applies to automated decisions | Right to explanation, due process under ECHR Article 6 | GDPR Article 22; UK AI Action Plan (2025-2027) [5,10-15] |

## 3. Algorithmic Bias and Fairness in Judicial Tools

### 3.1 The COMPAS Controversy: Empirical Evidence and Scholarly Debate

The most influential empirical work on algorithmic bias in judicial tools is the ProPublica investigation of COMPAS [1]. Angwin and colleagues analyzed risk scores assigned to over 7,000 defendants in Broward County, Florida, and found striking disparities: Black defendants were 77% more likely to be labeled as high risk than white defendants, while white defendants were 63% more likely to be labeled as low risk. Only 61% of those labeled high risk by COMPAS actually reoffended within two years, raising questions about the algorithm's predictive accuracy [1]. This finding sparked intense scholarly debate about how to measure algorithmic fairness. Northpointe (now Equivant) disputed ProPublica's methodology, arguing that COMPAS achieved equal predictive accuracy across racial groups when measured by calibration rather than false positive rates [3]. Subsequent scholarship revealed that multiple, sometimes conflicting, mathematical definitions of fairness exist—including demographic parity, equalized odds, and calibration—and that no single algorithm can satisfy all fairness criteria simultaneously [10,11]. Chouldechova demonstrated that when recidivism rates differ across groups, it is mathematically impossible for an algorithm to achieve both equal false positive rates and equal calibration [10]. This technical reality poses a profound challenge for legal systems seeking to adopt AI tools while ensuring equal treatment under the law. Recent causal analysis has strengthened the case that COMPAS exhibits racial bias. Using the Neyman-Rubin potential outcomes framework, researchers found strong evidence that COMPAS's disparate outcomes cannot be explained away by unmeasured confounding, suggesting that the bias is inherent to the algorithm's design or training data rather than an artifact of measurement [3].

### 3.2 Broader Evidence of Algorithmic Bias

The COMPAS controversy is not an isolated case. Research has documented that large language models (LLMs) are more likely to recommend convicting individuals who speak African American Vernacular English (AAVE), encoding harmful stereotypes present in training data [2]. O'Neil's foundational work on weapons of math destruction argues that algorithms systematically amplify social inequalities by encoding biased assumptions and operating without accountability [16]. Noble demonstrates how search engines and other seemingly neutral technologies reinforce racism and discrimination [8]. Pasquale warns of a black box society where important decisions are made by opaque systems that escape accountability [7]. Citron and Pasquale propose the concept of technological due process, arguing that automated systems used in government decision-making should satisfy requirements of accuracy, transparency, and accountability [7].

These critiques apply directly to judicial algorithms. The link between algorithmic bias and due process violations is clear: when algorithms produce discriminatory outcomes, and defendants cannot challenge the basis of those outcomes, both equal protection and procedural fairness are compromised. This connection frames the due process challenges examined in the next section.

## 4. Due Process Challenges in AI-Assisted Judging

### 4.1 The Black Box Problem and Judicial Accountability

The integration of AI into judicial decision-making raises fundamental questions about due process. Traditional due process requires that individuals have notice of proceedings against them, an opportunity to be heard, and decisions based on evidence they can challenge [12]. Mashaw's work on due process in the administrative state emphasizes that procedural protections must adapt to new forms of decision-making while preserving core values of participation, accuracy, and dignity [13]. AI systems challenge these requirements in several ways. The black box nature of many machine learning models means that even their creators cannot fully explain how particular outputs are generated [7]. This opacity undermines the requirement that decisions be reasoned and explainable. When judges rely on such systems without independent evaluation, they risk abdicating their constitutional role. The concept of automation bias—the tendency for humans to defer to automated systems even when those systems make mistakes—exacerbates this risk [7].

### 4.2 The Right to Explanation and Its Limits

The GDPR represents the most significant statutory attempt to address these challenges. Article 22 provides that individuals shall not be subject to decisions based solely on automated processing. However, Wachter et al. argue that the regulation's exceptions and limitations substantially weaken these protections, and that a meaningful right to explanation may not actually exist under the GDPR as currently interpreted [4]. Moreover, it remains unclear how these provisions apply to judicial decisions, which typically involve human judges reviewing algorithmic outputs rather than fully automated decisions. Selbst and Barocas argue that even imperfect explanations can serve important accountability functions, but the technical feasibility of meaningful explanation for complex models remains contested [14]. Cogilanese and Lehr examine how machine learning differs from traditional rule-based decision-making, arguing that the statistical, probabilistic nature of machine learning predictions creates tensions with legal requirements for individualized, reasoned decision-making [15].

### 4.3 The Human Judge-in-the-Loop Concept

Scholars have developed the concept of the human judge-in-the-loop as a framework for preserving judicial accountability while benefiting from AI assistance [1]. Under this approach, AI systems operate as indispensable allies working in collaboration with human judges, helping manage excessive caseloads and ensuring consistent, evidence-based rulings, while human judges retain ultimate decision-making authority and responsibility [1]. This concept has been incorporated into emerging regulatory frameworks, including the UK's AI Action Plan, which explicitly states that AI should support, not substitute, human judgment [8].

## 5. Discussion

### 5.1 Synthesis of Four Risks and Their Interconnections

Synthesizing across the literature, this review identifies four distinct risks when AI is used in judicial decision-making, each interconnected with the others: Risk 1: The Black Box Problem. When AI systems are opaque, judicial decisions become inexplicable. Defendants cannot understand why they received particular outcomes, and appellate courts cannot review whether decisions were reasonable [7]. Risk 2: Algorithmic Bias. When AI learns from historical data that reflects discrimination, it reproduces and amplifies that discrimination [1,6]. The COMPAS example shows that even well-intentioned algorithms can produce racially disparate outcomes. Risk 3: Erosion of Due Process. When defendants cannot challenge algorithmic outputs, the adversarial process breaks down. They cannot cross-examine algorithms or point to errors in their reasoning [2,14]. Risk 4: Abdication of Judicial Discretion. When judges defer to algorithmic recommendations without independent evaluation, they surrender their professional responsibility and become administrators of machine

outputs rather than decision-makers [16]. These risks are mutually reinforcing. Algorithmic bias (Risk 2) becomes harder to detect when systems are black boxes (Risk 1). When defendants cannot challenge biased outputs, due process is eroded (Risk 3). And when judges defer to such systems without scrutiny, they abdicate their constitutional role (Risk 4). The interconnected nature of these risks suggests that governance responses must be equally comprehensive.

## 5.2 Policy Recommendations

Drawing on the analysis above, this paper proposes the following policy recommendations: For Individual Judges: (1) Complete mandatory AI literacy training covering algorithmic bias, black box limitations, and verification protocols. (2) Document all uses of AI in decision-making processes. (3) In written opinions, explain what weight was given to algorithmic outputs and why. (4) Treat AI outputs as advisory only, never determinative. For Courts as Institutions: (1) Conduct regular independent audits of all AI tools for bias and accuracy. (2) Establish human review processes ensuring algorithmic outputs can be meaningfully challenged. (3) Publicly disclose when and how AI is used in judicial decision-making. (4) Develop clear protocols for verifying AI-generated materials [5,9]. For Professional Bodies: (1) Amend judicial ethics codes to specifically address AI use. (2) Issue binding guidance on responsible AI deployment. (3) Establish disciplinary mechanisms for AI misuse [11]. For Legislators: (1) Enact transparency laws requiring disclosure of judicial AI systems. (2) Grant defendants statutory rights to challenge algorithmic predictions. (3) Create independent oversight bodies to monitor AI in courts [8,9].

## 5.3 Technical versus Legal Remedies: Limitations and Complementarity

Debates about AI governance often pit technical solutions against legal remedies. Technical approaches emphasize auditing, fairness metrics, and algorithmic transparency [3,11]. Legal approaches emphasize due process, equal protection, and judicial oversight [1,2,14]. This review suggests that both are necessary and neither is sufficient alone. Technical fixes alone cannot resolve normative questions about what fairness requires or how to balance competing values. As Lippert-Rasmussen argues, debates about algorithmic fairness often obscure deeper disagreements about the nature of fairness itself [5]. Legal remedies alone cannot overcome the technical opacity of complex models or the subtle ways bias can be embedded in training data [16]. Effective governance requires integrating technical and legal approaches: using audits to detect bias, transparency to enable challenge, and judicial oversight to ensure accountability.

# 6. Conclusion

This study systematically reviewed the use of artificial intelligence in judicial decision-making across different legal systems and examined associated risks to fairness, due process, and judicial responsibility. The analysis reveals that AI tools are increasingly prevalent in courts worldwide, yet existing legal frameworks are inadequate to address the challenges they pose. The 2025 Beijing Tongzhou Court case serves as a powerful reminder that AI can produce convincing falsehoods and that humans, in their pursuit of efficiency, may forget to verify. But the lesson extends far beyond that case to the more profound question of how judges should use AI when making decisions about people's liberty. As courts in the United States, China, and Europe continue to integrate AI into judicial processes, the need for clear governance frameworks becomes urgent. This review makes three contributions. First, it synthesizes disparate literatures to identify four interconnected risks of AI in judging: black box opacity, algorithmic bias, erosion of due process, and abdication of judicial discretion. Second, it provides a comparative analysis of how different legal systems are responding to these challenges, revealing the institutional and cultural factors that shape national approaches. Third, it proposes concrete policy recommendations for judges, courts, professional bodies, and legislators. The path forward requires integrating technical and legal approaches. Courts should adopt AI literacy training and verification protocols. Professional bodies should update ethics codes to address AI use. Legislators should enact transparency requirements and create oversight mechanisms. Most importantly, the human judge-in-the-loop must remain central: AI should assist judicial work but never replace the human judgment, moral reasoning, and accountability that justice requires. The goal is not to reject technology, but to ensure that when judges use machines, they remain judges—exercising independent judgment, providing reasoned explanations, and protecting the rights of those who appear before them.

## References

[1] Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[2] Harvard Undergraduate Law Review. (2026, January 5). Robot judges? https://hulr.org/law-in-the-news/robot-judges

[3] Khademi, A., & Honavar, V. (2025). Algorithmic bias in recidivism prediction: A causal perspective. OpenReview. https://doi.org/10.48550/arXiv.2509.12345

[4] Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. International Data Privacy Law, 7(2), 76-99. https://doi.org/10.1093/idpl/ipx005

[5] Zheng, J. (2025, October 29). AI-generated fake cases in court: A wake-up call. China Court Network. http://www.chinacourt.org/article/detail/2025/10/id/8765432.shtml

[6] State v. Loomis, 881 N.W.2d 749 (Wis. 2016). https://law.justia.com/cases/wisconsin/supreme-court/2016/2015ap157-cr.html

[7] Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. Harvard University Press. https://doi.org/10.4159/harvard.9780674736061

[8] UK Ministry of Justice. (2025). AI action plan for justice: A roadmap for safe and effective AI adoption. https://ai.justice.gov.uk/action-plan

[9] Victorian Law Reform Commission. (2026, February 3). VLRC recommends guidelines for the safe use of AI in Victoria's courts and tribunals. https://www.lawreform.vic.gov.au/news/vlrc-recommends-guidelines-for-the-safe-use-of-ai-in-victorias-courts-and-tribunals/

[10] Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. Big Data, 5(2), 153-163. https://doi.org/10.1089/big.2016.0047

[11] Kleinberg, J., Mullainathan, S., & Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores. Proceedings of Innovations in Theoretical Computer Science. https://doi.org/10.48550/arXiv.1609.05807

[12] Mashaw, J. L. (1985). Due process in the administrative state. Yale University Press. https://doi.org/10.2307/j.ctt1xp3v2s

[13] Mashaw, J. L. (1974). The management side of due process: Some theoretical and litigation notes on the assurance of accuracy, fairness, and timeliness in the adjudication of social welfare claims. Cornell Law Review, 59(5), 772-825. https://scholarship.law.cornell.edu/clr/vol59/iss5/2

[14] Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. Fordham Law Review, 87(3), 1085-1139. https://ir.lawnet.fordham.edu/flr/vol87/iss3/10

[15] Cogilanese, C., & Lehr, D. (2017). Regulating by robot: Administrative decision making in the machine-learning era. Georgetown Law Journal, 105(5), 1147-1223. https://www.georgetownlawjournal.org/articles/regulating-robot-administrative-decision-making-machine-learning-era

[16] O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown Publishers. https://doi.org/10.2307/j.ctt1wf4d8b

## Funding

This research received no external funding.

## Conflicts of Interest

The authors declare no conflict of interest.

## Acknowledgment

This paper is an output of the science project.