Who Writes the Judgment—and Who Cares? Public Assessments of the Advantages and Challenges in LLM-Assisted Judgment Writing

Written by Yixiao Wen[1]**, Sinong Lu**[2]

[1]Shanghai University, [2] University of Macau

Abstract

Courts have begun experimenting with Large Language Model (LLM)-assisted systems capable of generating judicial opinions, raising important questions about how the public perceives such involvement. We conducted a small, rapid qualitative study with seven individuals who had prior litigation experience and had used LLMs to generate legal content. Participants evaluated LLM-generated judgments through three dimensions—fairness, factual accuracy, professional form. They identified advantages such as more impartial reasoning and more formal legal language, alongside concerns about biased training data, hallucinations, and difficulties in meeting legal reasoning requirements. These findings highlight how perceptions are shaped by users' mental models, prior LLM experiences, and legal expertise, offering early insights for designing trustworthy LLM support in judicial decision-making.

Introduction

The judiciary has traditionally been cautious in adopting LLM, especially in matters regarding judicial decision-making (JDM) (Barysė and Sarel 2024; Fine, Berthelot, and Marsh 2025). This caution stems from the fact that JDM is a fundamental judicial task (Oldfather 2007) and a high-stakes domain under contemporary AI governance frameworks (Mäntymäki et al. 2022).Yet recent pilots have moved further to introduce LLMs-assisted adjudication systems capable of generating judicial opinions. These systems can summarize facts, extract disputed issues, and generate legal reasoning and draft judgments (Pereira et al. 2025; Liu and Li 2024), all of which are integral to JDM.

This shift has sparked theoretical discussions about the implications of LLM involvement in judgment writing,like its impact on the judicial duty to state reasons, a duty central to the legitimacy, transparency, and accountability of JDM (Barry 2024; Hendrickx 2024; Dymitruk 2019). While there is extensive research on LLMs and JDM, particularly regarding public trust as a key concern in both legal and technological domains, it remains unclear whether the public trusts judicial judgment generated with LLMs assistance. Since current court practices of LLMs-assisted opinion generation are inaccessible to external users such as lawyers and litigants (Liu and Li 2024; Barysė and Sarel 2024), we adopt a prospective approach to explore public perceptions and ask:

RQ: How do people assess the advantages and challenges of LLM involvement in judgment writing?

To address our research question, we conducted a small qualitative study with seven individuals who had participated in legal proceedings, read judicial judgments, and used LLMs to generate legal content. Through semi-structured interviews, we explored their experiences with judicial documents, their interactions with LLM tools, and their views on LLM involvement in judgment writing. All interviews were transcribed and thematically analyzed by two researchers with strong inter-coder agreement.

Our findings show that participants evaluated LLM-generated judicial judgments along three dimensions: fairness, factual accuracy, professional form. They perceived advantages such as more impartial reasoning, broader argumentative support, and more formal and readable language, while also identifying challenges including bias risks from training data, hallucinated facts, and gaps in doctrinal reasoning.

Overall, this study provides an initial empirical glimpse into how people make sense of LLM-generated judicial judgment. It reveals the evaluative dimensions the public uses, the factors shaping these perceptions, including mental models of LLMs, prior LLM experiences, and legal knowledge background, and why challenges such as bias and hallucination carry particular weight in the judicial context. These insights offer an early foundation for future work on designing trustworthy and context-appropriate LLM support for judicial decision-making process.

Background and Related Work

LLMs Involvement in Judgment Writing

LLMs initially entered courts to support administrative tasks such as e-filing and speech-to-text transcription (Laptev and Feyzrakhmanova 2024; Pinna et al. 2024). Their efficiency gains made them widely accepted, particularly in response to growing caseloads (Bielen et al. 2018; Shi, Sourdin, and Li 2021). More recently, LLM-based systems have begun supporting functions closer to JDM, including risk assessment, analysis of case materials, legal reasoning, and even pilot projects for judgment drafting (Xu 2022; Liu and Li 2024; Pereira et al. 2025).

Since human decision-making is a defining element of the judiciary (Re and Solow-Niederman 2019), this shift has intensified debates about the boundaries between LLM and JDM. Existing studeis highlights concerns about transparency, legitimacy, accountability, fairness, bias, access to justice, and even the prospect of "LLM judges" (Selçuk, Konca, and Kaya 2025; Nowotko 2021; Vidaki and Papakonstantinou 2025; Kim and Peng 2025).

Yet within this broader discourse, LLM-assisted judgment writing remains understudied, despite being a concrete and practice-oriented issue. Opinion writing is central to judicial justification (Oldfather 2007) and to the implementation stage of JDM (Barysė and Sarel 2024). Requirements of accountability and legitimacy further demand that legal reasons be formulated by a responsible human (Pasquale 2019). Because the duty to give reasons is fundamental across legal systems (Ho 2000), LLM-generated judicial opinions constitute a significant but insufficiently examined development. This gap motivates our study.

.

Trust in LLM-Assisted Judgment Writing

Trust is a fundamental concern both in judicial decision-making (Siau and Wang 2018; Jamieson and Hennessy 2006) and in the adoption of LLM across domains (Afroogh et al. 2024). In human–AI interaction research, trust is commonly conceptualized as an attitude—an internal evaluation that cannot be directly observed (Lee and See 2004; Schrills et al. 2025). Its assessment therefore relies on subjective ratings (Kohn et al. 2021).

Within the context of LLM-assisted judicial decision-making (LLM-JDM), prior studies using Likert-type scales and regression modeling have found that participants' trust in LLM tools increases when they believe judges trust these tools (Fine, Berthelot, and Marsh 2025). Research also shows that trust in judicial algorithms varies across different stages of the decision-making process, with legal professionals perceiving automation during the implementation stage as less fair (Barysė and Sarel 2024). Interview-based studies similarly report that German judges' acceptance of LLM systems varies by system type (Dhungel and Beute 2024).

However, how the public evaluates the use of LLM in judgment writing,a specific and increasingly relevant part of LLM-JDM,remains unexplored. Addressing this gap, we empirically examine public trust in LLM-assisted judgment writing.

Method

To explore public assessments of LLM involvement in judicial decision-making, we conducted a qualitative study using semi-structured interviews with individuals who had prior experience with legal proceedings and judicial opinions. Our aim was to understand how they evaluated LLM-generated judgments and what shaped these perceptions.

Participants were recruited through social media and online legal communities. Eligible participants were adults who had taken part in at least one legal proceeding, had read a judicial judgment, and had used LLMs to generate legal content. Seven participants (ages 24–30) met these criteria and provided informed consent. Their demographic information is summarized in Appendix .

Interviews were conducted in October 2025, either in person or online, and lasted about 60 minutes. Each interview covered three areas: (1) basic demographic and litigation background, (2) experiences with judicial documents, and (3) perceptions of LLM use in judgment writing, including views on credibility, fairness, and legal profession. All interviews were audio-recorded and participants received 100 RMB compensation.

We analyzed the data using thematic analysis (Boyatzis 1998). Two researchers independently coded the transcripts and reconciled differences through regular discussions, developing a shared codebook. Inter-coder reliability was strong (Cohen's Kappa = 0.84 (McHugh 2012)). The researchers then collaboratively refined themes to ensure consistency and conceptual clarity.

Findings

Our findings show how people evaluate LLM-generated judicial judgments and the advantages and challenges they perceive in this context. We first outline the key dimensions guiding their evaluations—fairness, factual accuracy, and professional form—followed by the advantages and challenges identified for each. A summary of these perceptions is presented in Table 1.

Dimensions of Evaluating Judgment

Before analyzing how participants assessed the advantages and challenges of LLM involvement in judgment generation, we first outline the main dimensions they used when evaluating such opinions. Participants consistently focused on four core aspects: fairness, factual accuracy, professional form. These dimensions formed the basis of how they judged the quality of LLM-generated judicial opinions.

Fairness Fairness refers to whether the judgment reflects impartial reasoning and aligns with basic principles of justice. Participants emphasized that fairness is a central criterion,they viewed fairness as the foundation of judicial authority and legitimacy.*"Because I've always believed that the law is something very rigorous, very fair, and highly professional. And in the end, fairness of the judgment is what matters most."*(P5).

Factual Accuracy Factual accuracy in adjudication concerns whether the judgment correctly identifies and represents the facts of the case.*"It's impossible for it not to be based on legal facts—that simply cannot happen. "*(P7).Participants emphasized that judicial opinions must offer a precise and faithful account of the case. Even small factual errors were seen as unacceptable in documents carrying such high stakes.

Professional Form Professional form refers the formal qualities of a judicial opinion, including the use of proper legal terminology, correct citation of laws and precedents,

Table 1: Summary of perceived advantages and challenges of LLM-generated judicial opinions across four evaluation dimensions.

Dimension	Advantages	Challenges
Fairness	Impartial Reasoning Unaffected by Personal Emotions Enriched Reasoning Inspired by LLM-Provided Information	Fairness Risks from Inconsistent or Biased Training Data
Factual Accuracy		Inaccurate Facts Caused by Hallucinations Case-mixing and unreliable citations
Professional Form	More Rigorous and Formal Legal Language	Limitations in Meeting Legal Form Requirements

logical structuring, and adherence to established judicial formats. *"You can't have informal phrasing in it, including casual wording."*(P3).Participants considered professional form as significant because it signals rigor and legal expertise,and through its formalized structure, it directly ensures that opinions meet institutional standards.

Perceived Advantages of LLM-Generated Judicial Judgment

LLM involvement in judicial opinion generation is perceived to bring several notable advantages to the judiciary. These advantages primarily relate to improving fairness, extending the depth and coverage of legal reasoning, strengthening the formal quality of judicial language, and generating opinions that are more consistent and readable. Below, we outline four major advantages identified through the evaluation criteria established earlier.

Impartial Reasoning Unaffected by Personal Emotions LLM-generated judicial opinions are viewed as less influenced by human emotions or personal preferences. Because LLMs do not experience sympathy, frustration, or interpersonal pressure, their reasoning is considered less likely to deviate from principles of procedural justice.

> *"In a judgment like this, there is much less interference from personal feelings… The human touch is weaker, but that actually makes it more just, not too biased in any direction."* (P2)

Enriched Reasoning Inspired by LLM-Provided Information LLMs can retrieve and organize information that users may fail to recall or identify independently. In serving as a reference or idea-generation tool, they introduce additional reasoning pathways and widen the set of arguments that may inform a legal opinion.

> *"For me, its biggest advantage is that it can take the point I raise and give me reasoning and supporting arguments… things I normally would not have thought of, but it can find them from its database."* (P1)

More Rigorous and Formal Legal Language With appropriate prompts, LLMs can produce text with stable legal style, precise terminology, and consistent structure (Deng et al. 2023; Wu et al. 2020). Unlike human drafting, which varies with writing ability, career experiences and educational background (Osbeck 2011), LLM-generated text tends to remain formal and technically accurate.

> *"In terms of reasoning, typos, rigor, and completeness,LLM will definitely do better. When a human writes a judicial document, the language ability or educational background can be limiting, but with LLM, these problems become much less significant."* (P4). *"I say this because I feel that what it writes looks more professional... its wording and phrasing really do appear more professional."*(P1).

Perceived Challenges of LLM-Generated Judicial Judgment

While LLMs offer several advantages, participants also identified notable challenges that impact their acceptance of LLM-generated judicial judgment. These challenges relate to fairness risks stemming from training data, concerns about factual accuracy, limitations in professional legal reasoning, and doubts about whether LLMs can handle complex case distinctions.

Fairness Risks from Inconsistent or Biased Training Data Because LLMs learn from large collections of past cases and publicly available text, their outputs may inherit biases or inaccuracies embedded in the training data. If the dataset includes unfair, low-quality, or historically biased cases, these patterns may influence the model's reasoning and harm fairness. Prior research has shown that machine learning models tend to reproduce the biases present in their training data, and similar concerns apply to LLMs used in legal contexts.

> *"LLM is still too fixed and too formal, and it can carry certain biases."* (P5)

Inaccurate Facts Caused by Hallucinations LLMs are known to produce hallucinations, leading to incorrect or fabricated factual statements. This undermines confidence in their ability to represent case facts accurately. Participants noted that LLM-generated legal analyses are often unusable without careful manual verification, as the model may confuse cases, misattribute facts, or incorrectly summarize legal outcomes.

> *"I asked it about the latest ruling, and it took information from another case, even the property auction record, and attached it to mine. It was just wrong—the people were wrong, the facts were wrong, and the legal judgment was wrong. LLM was not accurate."* (P7).

Limitations in Meeting Legal Form Requirements Although LLM-generated text often appears formal and polished, its professional legal reasoning may still fall short, especially for trained legal practitioners. While some users viewed LLM-generated content as more professional than average human writing, legally trained participants pointed out that LLMs still struggle with legal relationships and doctrinal reasoning. This difference in perception suggests that apparent professional form does not always reflect substantive legal accuracy.

> *"It may look more polished, with more professional wording and tone, but how much of it is actually correct?The facts may not be accurate, and the reasoning from beginning to end may not really hold together."*(P1).

Discussion

Factors Shaping Perceptions of LLM-Generated Judicial Opinions

Our findings show that perceptions of LLM-generated judicial opinions arise from multiple intertwined factors that jointly shape how people evaluate fairness, reliability, and quality.

These perceptions are influenced first by how individuals conceptualize LLMs—some see them as objective machines, while others view them as data-driven systems shaped by historical bias. They are also shaped by direct experiences with output quality, as encounters with reliable or unreliable responses strongly affect trust. In addition, differences in legal expertise matter: legal professionals focus more on doctrinal precision, while non-experts prioritize clarity and readability. Together, these factors create varied and sometimes conflicting interpretations of the value and risk of LLM-generated judicial opinions.

The Unique Nature of Judicial Opinions and the Distinct Challenges Faced by LLMs

Compared with other domains of LLM-assisted writing—such as creative writing, educational content generation, or scientific text drafting—judicial opinions carry unique normative, procedural, and societal functions (Dothan and Maucec 2025). These unique functions introduce distinct challenges for LLM-generated judgemnt that do not appear, or appear with less intensity, in other domains.

First, judicial judgment are tightly coupled to individuals' rights and interests, making fairness sensitivity significantly higher than in most other writing contexts (Kirby 1990). Even small perceived deviations in fairness or impartiality can fundamentally undermine public trust in the judiciary. This heightened fairness sensitivity explains why participants scrutinized LLM-generated opinions more strictly than other LLM-generated content.

Second, writing judgment requires high levels of domain-specific expertise, including the precise articulation of legal relationships, application of legal doctrines, and correct referencing of statutes and precedent (Oldfather 2007). At the same time, these opinions must remain accessible to a broad audience whose legal literacy varies widely. This dual demand creates a unique tension: the text must be both technically sophisticated and broadly readable. LLMs struggle to meet this combined requirement: while they may excel in producing clear writing or formal language, they often fall short on doctrinal accuracy and factual correctness, which are indispensable in judicial reasoning.

These characteristics make the judicial domain uniquely challenging for LLM involvement. Errors that may be tolerable in other writing contexts become unacceptable in legal decision-making, and readability demands that might be optional elsewhere become essential here. This dual accountability creates a heightened threshold for acceptable LLM-generated judicial content.

Design Implications

Transparent and Curated Data Inclusion Mechanisms Given participants' concerns about bias in training data, courts deploying LLM-based tools should maintain transparent criteria for including and excluding cases during model fine-tuning. While raw judicial data must be desensitized to protect privacy, partial transparency—such as publishing selection standards, exclusion rules, or dataset summaries—can help mitigate concerns about hidden bias entering the system.

Scenario-Specific Model Design As we discussed, judgment writing involve highly specialized legal reasoning while being consumed by audiences with mixed levels of legal knowledge. Thus, judicial LLM systems should therefore adopt scenario-specific model designs, where different models or modules support different stages of the legal process—for example, one optimized for internal legal reasoning and another for public-facing explanations. Tailoring models to distinct workflow contexts ensures that both professional accuracy and public readability are appropriately addressed.

Conclusion and Future Work

This study offers an initial understanding of how people assess LLM involvement in judgment writing. Through a small set of interviews, we surface both perceived value and key concerns that shape trust in LLM-assisted judgments. While exploratory, this work points to the need for broader empirical studies, evaluations with legal professionals and real case materials, and co-design efforts to ensure transparency, accuracy, and fairness in future judicial LLM systems. Our findings lay groundwork for deeper inquiries into how LLMs can be responsibly integrated into judicial decision-making.

References

Afroogh, S.; Akbari, A.; Malone, E.; Kargar, M.; and Alambeigi, H. 2024. Trust in AI: progress, challenges, and future directions. *Humanities and Social Sciences Communications*, 11(1): 1–30.

Barry, B. M. 2024. AI for assisting judicial decision-making: Implications for the future of open justice. *Australian Law Journal*, 98(9): 656–669.

Barysė, D.; and Sarel, R. 2024. Algorithms in the court: does it matter which part of the judicial decision-making is automated? *Artificial intelligence and law*, 32(1): 117–146.

Bielen, S.; Peeters, L.; Marneffe, W.; and Vereeck, L. 2018. Backlogs and litigation rates: Testing congestion equilibrium across European judiciaries. *International Review of Law and Economics*, 53: 9–22.

Boyatzis, R. E. 1998. *Transforming qualitative information: Thematic analysis and code development*. sage.

Deng, W.; Pei, J.; Kong, K.; Chen, Z.; Wei, F.; Li, Y.; Ren, Z.; Chen, Z.; and Ren, P. 2023. Syllogistic reasoning for legal judgment analysis. In *Proceedings of the 2023 conference on empirical methods in natural language processing*, 13997–14009.

Dhungel, A.-K.; and Beute, E. 2024. AI systems in the judiciary: amicus curiae? Interviews with judges on acceptance and potential use of intelligent algorithms.

Dothan, S.; and Maucec, G. 2025. It is Our Flaws That Make Us Humane: How Technology Ruined Judicial Craft. *SMU Science and Technology Law Review (2025)*.

Dymitruk, M. 2019. The right to a fair trial in automated civil proceedings. *Masaryk University Journal of Law and Technology*, 13(1): 27–44.

Fine, A.; Berthelot, E. R.; and Marsh, S. 2025. Public Perceptions of Judges' Use of AI Tools in Courtroom Decision-Making: An Examination of Legitimacy, Fairness, Trust, and Procedural Justice. *Behavioral Sciences*, 15(4): 476.

Hendrickx, V. 2024. AI and the Judicial Duty to State Reasons.

Ho, H. L. 2000. The judicial duty to give reasons. *Legal Studies*, 20(1): 42–65.

Jamieson, K. H.; and Hennessy, M. 2006. Public understanding of and support for the courts: Survey results. *Geo. LJ*, 95: 899.

Kim, T.; and Peng, W. 2025. Do we want AI judges? The acceptance of AI judges' judicial decision-making on moral foundations. *AI & SOCIETY*, 40(5): 3683–3696.

Kirby, M. 1990. On the writing of judgments. *Australian Journal of Forensic Sciences*, 22(3): 104–124.

Kohn, S. C.; De Visser, E. J.; Wiese, E.; Lee, Y.-C.; and Shaw, T. H. 2021. Measurement of trust in automation: A narrative review and reference guide. *Frontiers in psychology*, 12: 604977.

Laptev, V. A.; and Feyzrakhmanova, D. R. 2024. Application of artificial intelligence in justice: current trends and future prospects. *Human-Centric Intelligent Systems*, 4(3): 394–405.

Lee, J. D.; and See, K. A. 2004. Trust in automation: Designing for appropriate reliance. *Human factors*, 46(1): 50–80.

Liu, J. Z.; and Li, X. 2024. How do judges use large language models? Evidence from Shenzhen. *Journal of Legal Analysis*, 16(1): 235–262.

Mäntymäki, M.; Minkkinen, M.; Birkstedt, T.; and Viljanen, M. 2022. Defining organizational AI governance. *AI and Ethics*, 2(4): 603–609.

McHugh, M. L. 2012. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3): 276–282.

Nowotko, P. M. 2021. AI in judicial application of law and the right to a court. *Procedia computer science*, 192: 2220–2228.

Oldfather, C. M. 2007. Writing, cognition, and the nature of the judicial functions. *Geo. LJ*, 96: 1283.

Osbeck, M. K. 2011. What is" good legal writing" and why does it matter? *Drexel L. Rev.*, 4: 417.

Pasquale, F. 2019. A rule of persons, not machines: the limits of legal automation. *Geo. Wash. L. Rev.*, 87: 1.

Pereira, J.; Assumpcao, A.; Trecenti, J.; Airosa, L.; Lente, C.; Cléto, J.; Dobins, G.; Nogueira, R.; Mitchell, L.; and Lotufo, R. 2025. Inacia: Integrating large language models in brazilian audit courts: Opportunities and challenges. *Digital Government: Research and Practice*, 6(1): 1–20.

Pinna, G.; Tugnoli, D.; Bartole, M.; Manzoni, L.; and De Lorenzo, A. 2024. From Courts to Comprehension: Can LLMs Make Judgments More Accessible? In *2024 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 297–304.

Re, R. M.; and Solow-Niederman, A. 2019. Developing artificially intelligent justice. *Stan. Tech. L. Rev.*, 22: 242.

Schrills, T.; Franke, T.; Hoesterey, S.; and Roesler, E. 2025. Questioning trust in AI research: exploring the influence of trust assessment on dependence in AI-assisted decision-making. *Behaviour & Information Technology*, 1–17.

Selçuk, S.; Konca, N. K.; and Kaya, S. 2025. AI-driven civil litigation: Navigating the right to a fair trial. *Computer Law & Security Review*, 57: 106136.

Shi, C.; Sourdin, T.; and Li, B. 2021. The smart court-a new pathway to justice in china? In *IJCA*, volume 12, 1. HeinOnline.

Siau, K.; and Wang, W. 2018. Building trust in artificial intelligence, machine learning, and robotics. *Cutter business technology journal*, 31(2): 47.

Vidaki, A. N.; and Papakonstantinou, V. 2025. Democratic legitimacy of AI in judicial decision-making. *AI & SOCIETY*, 1–11.

Wu, Y.; Kuang, K.; Zhang, Y.; Liu, X.; Sun, C.; Xiao, J.; Zhuang, Y.; Si, L.; and Wu, F. 2020. De-biased court's view generation with causality. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 763–780.

Xu, Z. 2022. Human judges in the era of artificial intelligence: challenges and opportunities. *Applied Artificial Intelligence*, 36(1): 2013652.

Appendix

Participants Demographic

The demographic information of the interview participants is presented in Table 2. The table summarizes key characteristics, including age, occupation, dispute types experienced, and the LLM tools they reported using. The participants represent a range of ages and professional backgrounds and were involved in different categories of civil disputes. They also reported varying levels of prior interaction with multiple LLM-based tools.

Table 2: Demographics of interview participants.

ID	Age	Occupation	Dispute Type	LLM Tools Used
P1	24	Legal intern	Traffic Accident	Deepseek,ChatGPT
P2	28	Property management	Personal Injury	Doubao,Deepseek
P3	28	Internet operations / marketing planning	Civil Contract	Doubao,kimi,Deepseek
P4	29	Civil court assistant	Civil Contract	Kimi,Deepseek,ChaGpt
P5	25–28	Furniture after-sales service	Product Responsibility	Doubao
P6	30	Independent hand-drawing studio owner	Civil Contract	Deepseek,Dreamia
P7	25–30	Corporate administrative staff	Civil Contract	Doubao