

Fostering Trust with Prudence: A Judicial Trust Model in the Era of Artificial Intelligence

JunQiang Xu

PhD Student, School of Politics and Public Administration, China University of Political Science and Law, Beijing, China. Email: Xujq2024@qq.com

Abstract: Judicial trust serves as the micro-foundation for the formation of judicial credibility, as well as the process of its concretization. Against the backdrop of the intelligent transformation of the judiciary, judicial trust manifests a novel model of "fostering trust with prudence." This model does not simply replace traditional ones but represents a complex evolution driven by the dual forces of technological empowerment and technological risks. Examining its underlying logic must not be confined to the dichotomous perspective of "human" versus "institution"; rather, it requires placing "technological systems" and their interactions with humans at the core of analysis. On the one hand, the "fostering trust with prudence" model positions technology as a "facilitator" of judicial fairness and efficiency, accumulating public trust in technology by enhancing transparency, reliability, and procedural regularity. On the other hand, this trust model also guards against technology becoming a new source of judicial bias and injustice, mitigating risks through algorithmic governance, checks and balances, and the enhancement of judicial personnel's capabilities. While these two aspects may appear contradictory on the surface, they are ultimately aligned with the fundamental goal of "ensuring that technology serves judicial fairness."

Keywords: Judicial trust; Artificial Intelligence; Fostering trust with prudence; Technology trustworthiness

Introduction

The advent of artificial intelligence (AI) is precipitating a profound transformation across the judicial landscape. Technologies such as smart adjudication, AI-assisted sentencing, and case law recommendation systems are reshaping the operational dynamics of judicial power, altering not only procedural efficiency but also the very foundations upon which public trust in the judiciary is built. Historically, judicial trust has been cultivated through interpersonal interactions between judges and litigants, as well as systemic confidence in legal institutions and procedures—a paradigm often summarized as establishing trust while maintaining necessary skepticism.

However, the integration of AI as a participant, assistant, and influencer within judicial processes fundamentally expands and complicates this trust equation. The object of trust extends beyond judges and institutions to encompass the technology systems themselves, and the interaction shifts from purely human-to-human dynamics to include complex human-computer interactions and collaborations. This evolution raises critical questions: In a justice system deeply embedded with technology, how is trust generated? What form does skepticism take, and how does it function as a vital safeguard?

The following discussion will first elaborate on the new propositions of judicial trust in the AI era. It will then dissect the dual logic of this model: the "establishing trust" logic focused on constructing the credibility of technology, and the "harboring concerns" logic dedicated to preventing technological risks. Through this exploration, the document argues that a mature and wise digital justice system must expertly balance these two seemingly contradictory yet fundamentally unified stances, ensuring that technological advancement ultimately serves the core judicial value of justice.

1. The Proposition of Judicial Trust in the Era of Artificial Intelligence

Trust is a key cornerstone for the stability of human society, a social phenomenon that emerged alongside human social interactions. Individuals exhibit significant differences in their modes of trust, a phenomenon rooted in the diversity of geographical environments and cultural backgrounds, particularly the divergence in educational levels and value systems. Human trust models continuously evolve with the expansion of interpersonal interactions and the transformation of social structures, as evidenced by the shift from interpersonal trust based on kinship ties to institutional trust gradually constructed with the development of industrial society, both highlighting the dynamic nature of trust models. ^[1]In the judicial field, as big data and artificial intelligence technologies are applied deeply and extensively (e.g., intelligent adjudication, sentencing assistance, similar case recommendation, and judicial management), new technologies are reshaping the operational posture of judicial power. Consequently, the model of judicial trust is undergoing a transformation, "shifting from a singular institutional trust towards a composite trust that integrates both institution and technology." ^[2]

In traditional research, the production of judicial trust primarily revolves around interpersonal and systemic interactions between judges and parties, as well as between courts and the public. Its core logic lies in judges employing strategies such as role-playing, discourse dominance, and emotion management to "build trust," while simultaneously "maintaining suspicion" based on professional characteristics and risk prevention. This constitutes the dualistic logic of action described as "fostering trust while upholding prudence." ^[3] However, when artificial intelligence becomes a participant, assistant, and even a partial influencer in judicial decision-making, the

^[1] See Han Ye, *A Holistic Study on the Transformation of Trust Models in the Digital-Intelligence Era*, Academic Exploration, No. 3, 2025.

^[2] Zhao Yang, *Judicial Trust in the Era of Artificial Intelligence and Its Construction*, Journal of the East China University of Political Science and Law, Vol. 4, 2021.

^[3] See Huang Rui, *Fostering Trust While Upholding Prudence: The Logic of Producing Judicial Trust from the Judge's Perspective*, Jurist, No. 2, 2023.

structure of judicial trust production undergoes a fundamental transformation: the object of trust expands from "humans and institutions" to include "technology and systems," and trust interactions partially shift from "human-human interaction" to "human-machine interaction" and "human-machine collaboration."

In the era of artificial intelligence, the production of judicial trust involves at least three dimensions: first, the public's trust in courts or technology companies that serve as technology providers and operators; second, the trust and acceptance of AI technological tools by judicial participants such as judges and parties; third, mediated by technology, the public's and parties' trust in the final judicial decisions that incorporate technology-assisted judgments. ^[4]The traditional connotation of "system trust" is now infused with the reliability and ethics of technological systems, and the perception of "procedural justice" is profoundly influenced by technical features such as algorithmic transparency and explainability. Therefore, exploring the logic of judicial trust in the AI era must address a core question: in a judicial process deeply embedded with technology, how is trust produced, and in what form does necessary suspicion exist and function?

The author argues that in the era of artificial intelligence, judicial trust manifests a novel logical form characterized by "fostering trust with prudence." This does not constitute a mere replacement of traditional trust models but rather represents a complex evolution under the dual context of technological empowerment and technological risks. Its dimension of "fostering trust" can be viewed as the judicial system's purposeful construction and demonstration of the trustworthiness of intelligent technology. Conversely, its dimension of "prudence" is reflected in a clear awareness of the inherent limitations of technology, the establishment of institutional safeguards against potential risks in its application, and vigilance against the strategic gaming of litigation tactics facilitated by technology. While these two aspects may appear contradictory, they are ultimately unified under the paramount goal of "ensuring that technology serves judicial fairness."

2. The Logic of "Fostering Trust" in Judicial Trust in the AI Era: Constructing Technological Trustworthiness

From a traditional perspective, judges build public judicial trust through "on-stage" performances such as courtroom conduct and reasoning in legal documents, combined with their professional competence. In the era of artificial intelligence, however, measures for "fostering trust" within the judicial system have largely shifted towards cultivating and demonstrating the trustworthiness of technology. This is not an abandonment of the human factor, but rather an effort to establish technological systems as trustworthy "agents" or "assistants." In this process, trust-production strategies exhibit new characteristics.

2.1 The Role of Technological Systems: Demonstrating Explainability and Reliability

The role undertaken by artificial intelligence in the judicial field resembles that of a "digital assistant" with professional functions. To establish trust in AI-assisted judicature, the primary requirement is its successful fulfillment of the role of a fair, reliable, and practical assistant. The construction of its trustworthiness relies on the

^[4] See Guo Chunzhen & Huang Sihan, *Trust in Artificial Intelligence for Criminal Justice and Its Construction*, *Jilin University Journal Social Sciences Edition*, No. 2, 2023.

strategic, conscious operation by judicial institutions. In practice, judicial institutions employ various strategies to enhance the perceived trustworthiness of the technological role: First, proactive transparency initiatives. Judicial organs actively manage public perception by disclosing the scope, principles, and inherent limitations of technological applications. A landmark example is the civil ruling (2025) Min 72 Min Te No. 76 by the Xiamen Maritime Court, which, for the first time in a judicial ruling, explicitly required litigation representatives to fully and continuously disclose their use of AI technology during litigation. This measure brings AI application under the supervision of procedural transparency, directly addresses public concerns about the "algorithmic black box," and represents an institutional exploration in constructing trust in intelligent judicial technology. Second, proof of reliability. Demonstrating empirical data on how technology-assisted systems enhance efficiency, unify standards, and reduce errors, thereby basing trust on quantifiable performance improvements. For instance, the AI-assisted adjudication system launched by the Shenzhen Intermediate People's Court in June 2024 reportedly reduced the average case-closing time for civil and commercial cases by 25.6% within six months of operation.^[5] Systems like Beijing courts' "Judge Rui," by integrating legal knowledge bases and similar-case recommendations, assist judges in unifying adjudication standards.^[6] Third, ceremonial demonstration of human-machine collaboration. Publicly or to the parties involved, demonstrating how technology provides assistance and how judges make final decisions, thereby re-emphasizing the "human-centric" principle of "technology reviews, judge decides." This further establishes the judge's authoritative image as the ultimate responsible subject and master of the technology, making the trustworthiness of intelligent technology a new foundation for the extension of judicial credibility.

2.2 Reconstruction of the Discourse System: The Fusion and Dominance of Legal and Computational Language

In the process of judicial digital transformation, a portion of discursive authority is shifting from traditional legal interpretation logic towards technological logic. Algorithms, by translating elements such as legal provisions and case characteristics into computable data models to generate predictions, classifications, and recommendations, have created a new form of "computational discourse." Faced with this transformation, the key for judges to establish trust lies in becoming effective "translators" and "directors" of this discourse, which is manifested in the practice of two core competencies. On one hand, it is necessary for judges to comprehend the conclusions output by technology, such as analyses of disputed issues and alerts for evidentiary contradictions, and to judiciously integrate them into legal reasoning, thereby enhancing the logicity and persuasiveness of judgments. An example is the innovative measure by the Jianshan District People's Court in Shuangyashan City, Heilongjiang Province, which introduced the Implementation Measures for the "Appending Effective Similar Case Judgments" Work Mechanism, deeply integrating similar-case judgments into the entire process before, during, and after litigation. The presiding judge conducts legal interpretation and reasoning based on the results of similar-case judgments.^[7] This initiative not only aligns with the requirements of the Supreme People's Court's Guiding Opinions on Unifying the Application of Law and Strengthening Similar-Case Retrieval (for Trial Implementation) for responding to

^[5] See Li Feiyan, *Security Dilemmas and Governance Pathways for Large Language Model Applications in Judiciary*, People's Tribune · Academic Frontiers, No. 7, 2025.

^[6] See Lao Jiaqi, *AI Assistance for More Efficient Judicial Justice*, Guangming Daily, April 26, 2025, at p.5.

^[7] See "Meticulous Cultivation, Responding to Public Expectations," People's Court Daily, March 3, 2026, at p.1.

parties' demands regarding similar cases, but also enhances the scientific nature and credibility of judgments by demonstrating consideration for a broader spectrum of legal practice. On the other hand, judges need to control the direction of the discourse to prevent intelligent technology from "usurping the leading role." This requires judges to possess the ability to critically examine algorithmic suggestions, identifying potential biases or logical flaws therein, and to adjust and correct them using professional legal terminology. Through such integration and direction, judges can leverage technology to strengthen the "scientific" nature of their own judgments to gain trust, while simultaneously retaining core authority over legal interpretation and discretion, preventing trust from becoming entirely reliant on uncontrollable technological output.

2.3 Optimization of Process Control: Enhancing the Perceived Fairness of Procedure

Through mechanisms such as process standardization, node-based reminders, and real-time recording, artificial intelligence technology has significantly increased the standardization, traceability, and predictability of judicial proceedings. For instance, intelligent case complexity stratification and distribution systems automatically categorize cases as "complex", "standard" or "simple" using pre-defined factors such as the cause of action, amount in dispute, number of parties, and complexity of claims. These systems then direct cases to corresponding trial teams. This algorithmic application aims to make procedural choices more objective and standardized, thereby reducing the arbitrariness of human discretion. ^[8] Another example is the fully paperless case handling model adopted by Shanghai courts, which achieves "end-to-end digital traceability" from the submission of materials by parties to the generation of judgments. Through real-time recording and visual supervision of each procedural node, this system ensures that the process is both immutable and reviewable, thereby constructing the "infrastructure" for technological trust. ^[9] Furthermore, intelligent judicial technology possesses the capability to generate legal documents from standardized templates, which can unify output formats, minimize formatting errors, and reduce variations between individuals, thereby enhancing the stability and predictability of judicial outputs. The meticulous and granular control over the procedure itself creates and establishes a critical foundation for institutional trust. By virtue of its impersonal nature, intelligent judicial technology transforms into a "rigid" guarantor of procedural fairness, allowing parties to more readily perceive a stable, equitable, and strictly enforced judicial process, which in turn substantially strengthens their sense of judicial trust.

3. The Logic of "Prudence" in Judicial Trust in the AI Era: Guarding Against Technological Risks

Unreserved trust in technology is perilous. The inherent conservatism of the judiciary and its relentless pursuit of justice necessitate an instinctive caution towards every new technology. The logic of "prudence" in the AI era is grounded in the inherent limitations of technology itself, the latent risks in its application, and the potential for its misuse. It leverages this prudence to regulate technology and defend against risks.

^[8] See "Case Complexity Stratification and Distribution: Accelerating the 'Realization of Justice'," sourced from the Supreme People's Procuratorate official website, at https://www.spp.gov.cn/spp/zdgz/202501/t20250107_678735.shtml, (last accessed March 5, 2026).

^[9] See Ge Xiang, *The Reality and Prospects of Artificial Intelligence Application in Judicial Practice: Referring to the Intelligent Case Handling Assistance System for Administrative Cases in Shanghai Courts*, Journal of the East China University of Political Science and Law, No. 5, 2018.

3.1 Distrust in Technology Itself: Algorithmic Risk Prevention and Control

The judiciary's institutional-level cautious stance and prudence logic towards AI technology constitute an indispensable prerequisite for preventing the potential systemic risks posed by AI. This prudence does not stem from a rejection of technology but from a clear and thorough understanding of its inherent limitations and external risks. In practice, such prudence manifests primarily in three areas: algorithmic bias and discrimination, the inexplicability of the "algorithmic black box," and technological failures and security risks. Accordingly, the judicial system transforms the stance of "prudence" into a series of proactive risk prevention and control mechanisms aimed at delineating safe boundaries for the application of technology. For example, in the *State v. Loomis* case in Wisconsin, where the COMPAS tool was used for sentencing, the defendant challenged the procedural due process and appealed for the disclosure of the algorithm. Thus, judicial organs should establish mechanisms for algorithmic auditing and compliance review, conducting independent assessments of a technology's fairness, security, and effectiveness prior to its deployment. Furthermore, it is essential to ensure that judicial personnel retain ultimate decision-making authority, clearly defining AI's role as an assistive tool whose outputs serve only as references, while the ultimate decision-making power and corresponding responsibility for judgments must remain with human judges. The case where an Indian judge was held accountable for relying on AI to generate fictitious precedents serves as a typical negative example of blind trust in AI output and the abdication of independent review duties.^[10] Simultaneously, the judicial system should construct a full-process data governance and security protection system. The Gansu Provincial People's Procuratorate formulated the Rules for Data Security Use and the Implementation Measures for Building a Data Sharing System, establishing a data classification and hierarchical protection system and a "four-layer accountability framework," achieving end-to-end security control from legality review at data collection to compliant disposal upon destruction. This provides an operational normative model for cross-departmental data sharing.^[11] In addition, the judicial system should establish contingency plans for technological failures, promote the development of "Explainable Artificial Intelligence" (XAI) technologies, and explore pathways such as introducing expert assistant systems. The intent of this institutionalized prudence and its corresponding prevention and control system is not to hinder technological innovation, but to build an immune mechanism for the healthy and sustainable integration of technological innovation into the judicial system. A mature judicial system does not readily trust statements from parties but relies on rigorous rules of evidence and cross-examination. Similarly, a rational digital judicial system should not readily trust algorithmic outputs but must depend on a stringent set of rules for technical scrutiny, procedural checks and balances, and accountability attribution.

3.2 Distrust Towards Litigants' Use of Technology: Countering Strategic Gaming

While intelligent judicial technology enhances the efficiency of judicial operations, it

^[10] See *Indian Judge to Face Consequences for Using AI in Case Adjudication*, sourced from *Guangming Online*, at https://m.gmw.cn/2026-03/05/content_1304365188.htm, (last accessed March 5, 2026).

^[11] See *"Top Ten Data Rule of Law Incidents in China 2025 (Digital Procuratorate Category), One from Gansu Selected"*, sourced from the official website of the Gansu Provincial People's Procuratorate, at https://mp.weixin.qq.com/s?__biz=MjM5OTg3ODkwNw==&mid=2689509632&idx=1&sn=8bbac75ba944c1c8714bbc2299ec66cc&chksm=83cf2b77b85f475072393837c86c2c78b056a599ce98b78acd072cc32cec632dd0d7e0dd40d#rd, (last accessed March 5, 2026).

also profoundly reconstructs the capability composition of litigation participants, granting parties and their representatives unprecedented technological empowerment. This empowerment manifests not only in efficiency gains for procedural tasks such as document generation and evidence organization but also extends to the formulation and execution of litigation strategies. For instance, leveraging Generative Artificial Intelligence (AIGC) can rapidly produce highly formatted legal documents, intelligently catalog and analyze large volumes of evidentiary materials, and even utilize big data models to predict the adjudicative tendencies of specific judges, thereby devising more targeted litigation strategies. Confronted with such "technology-empowered" litigation participants, judges must possess and maintain a professionally-oriented cautious attitude and sense of distrust. Firstly, regarding the potential for evidence contamination and information distortion arising from technological empowerment, judges' review duties must expand beyond traditional "formal authenticity" to encompass "substantive authenticity" and "technological credibility." Practical cases indicate that technological forgery has evolved from crude to sophisticated, necessitating an upgrade in scrutiny methods.^[12] This requires judges to not only evaluate evidence content based on logic and experience but also to possess the technical literacy to identify technological forgeries and algorithmic hallucinations. Secondly, judges must remain vigilant against the cognitive inertia and narrowed thinking induced by technological tools. Technological empowerment can not only forge "evidence" but may also attempt to influence or even "predict" the adjudicator's thinking, inducing "automation bias." Regardless of technological advancements, the core of judicial adjudication remains value judgment and interest balancing. It is imperative to penetrate the veneer of technology, persist in substantive review, and ensure the subjectivity and irreplaceability of the judge as a human agent.^[13] Thirdly, when technological empowerment devolves into technological fraud, the judiciary must demonstrate stringent regulatory power, sanctioning the abuse of technology to uphold judicial order, fairness, and justice.

3.3 Leveraging Distrust for Procedural Control: Pacing the Judicial Process

In traditional judicial practice, judges could control the pace of proceedings (for example, employing deliberate deliberation) to facilitate mediation or the ascertainment of facts. The development of intelligent judicial technology, while providing new tools and possibilities for procedural control, also triggers a judicial strategy known as "technological conservatism" due to its inherent uncertainties. Judges do not unconditionally embrace the efficiency brought by technology; they often deliberately choose relatively more traditional, time-consuming yet controllable procedural options to offset the hidden risks of technological application. For instance, when encountering cases involving complex and specialized issues, a judge may retain skepticism towards AI-generated evidentiary opinions, preferring to revert to traditional forensic appraisal procedures coupled with the expert assistant system for cross-examination and checks and balances, even if this may reduce procedural efficiency. Fundamentally, this deliberate choice of relative inefficiency stems from a reasonable suspicion regarding the current reliability of technology. Its aim is to maximize the exposure and filtration of errors through "redundancy design" within the procedure, thereby avoiding the risk of miscarriages of justice caused by

^[12] See *Alert! AI-Generated False Judicial Precedents Have Emerged*, sourced from China Court Network, at <https://www.chinacourt.cn/article/detail/2025/10/id/9043677.shtml> (last accessed March 6, 2026).

^[13] See Wu Xiyu, *The Practical Demand for AI-Based Adjudication and Its Chinese Mission*, *Oriental Law*, No. 2, 2018.

technology.^[14] Therefore, within the context of intelligent judicature, procedural control is no longer merely simple time management; it has evolved into a composite strategic approach. It integrates technological understanding, risk assessment, and the safeguarding of fair procedures, endowing the exercise of procedural power with more complex and sophisticated strategic characteristics. This is not resistance to technology, but rather an art of exercising procedural power developed to protect the substantive fairness of adjudication when judicial authority confronts highly complex and uncertain technological systems. Through this art, judges delineate safe procedural boundaries for the judicial application of technology, ensuring that technological advancement assists judicial work without undermining the bedrock of judicial justice.

4. Conclusion: Confronting the Judicial Trust Model of "Fostering Trust with Prudence"

Within the context of the artificial intelligence era, the generative mechanism of judicial trust manifests a dialectical development characterized by the dual logics of "fostering trust" and "prudence." Traditionally, this logic was primarily reflected in interpersonal interactions at the individual judge level, where a judge would build a trustworthy image while maintaining "due suspicion" towards other parties. In the technological age, it has evolved into the judicial system's systemic response, at an institutional level, to the effects generated by technology and its social application. On one hand, it is necessary to actively construct the trustworthiness of technology to gain efficiency benefits and extend public credibility; on the other hand, institutional caution must be consistently maintained, using prudence as a shield to prevent and control technological risks and defend core judicial values. In this process, technology is no longer merely a passive and neutral tool but has become a key variable reshaping the mode of judicial trust production.

Therefore, research on judicial trust must break free from the old dichotomy of personal trust versus institutional trust, and place the "technological system" itself and its complex interactions with humans at the core of analysis. In the context of intelligence, this distrust of technology is not a negative rejection of technological progress, but a form of preventive risk governance wisdom. Future research and practice should delve into the dynamic balancing mechanism of the dual logics of "fostering trust" and "prudence" in specific judicial contexts such as case filing, trial, adjudication, and execution, and explore how to organically integrate this dialectical logic into the process of judicial intelligent development through the design of legal rules, ethical guidelines, and technical standards, ultimately achieving the synergistic advancement of "Technology for Good" and the enhancement of judicial credibility.

^[14] See Ma Guoyang, *The Dual Risks in the Application of Artificial Intelligence Evidence and Their Regulation*, Journal of Dalian University of Technology (Social Sciences), Vol. 4, 2023.