Proceedings of ICILLP 2025 Symposium: Digital Governance: Inter-Firm Coopetition and Legal Frameworks for Sustainability
DOI: 10.54254/2753-7048/2025.BR26147

Potential Risks and Regulatory Paths of Data Application in Digital Courts

Minghan Liu

College of Liberal Arts and Law, Beijing University of Chemical Technology, Beijing, China
18632212618@163.com

Abstract. The application of data in current digital courts has advanced to a new stage, that is, to dig deeper into the quantity and quality of data and fully exploit the value of data. However, the lag between technology empowerment and institutional adaptation gives rise to multiple risks: technically, there are hidden dangers such as multiple paths of data leakage, algorithmic black box decision-making bias, and abuse of deepfake technology; ethically, there is a value conflict between data privacy protection and development and utilization. To address these challenges, it is necessary to build a classified and graded protection system to lay a solid security foundation, and at the same time establish an ethical review mechanism and a dynamic risk assessment system to form a governance framework that combines technical safeguards and institutional regulations.

Keywords: Digital Court, Data empowerment, Data governance

1. Introduction

Recently, Senior leaders of the judiciary have repeatedly emphasized the importance of data empowerment [1]. Data empowerment has become the core engine of deepening judicial reform. Regarding the application of digital court data, the academic community has generally recognized the strategic value of data as a new factor of production and focused on the practical paths of technology empowerment and the security risks that come with it. Although a basic framework centered on data classification and grading, security protection obligations, personal information processing norms and judicial confidentiality requirements has been established at present, its regulatory logic focuses on static data content and traditional judicial processes, making it difficult to adapt to the new risks arising from the dynamic empowerment of data in digital courts [2]. At the same time, the risks associated with data empowerment are becoming increasingly prominent. So, how can we promote the development of digital courts while achieving the co-governance of data application and protection?This article will deeply analyze the latent data risks in these dynamic application scenarios; Combining data characteristics and existing policy practices, explore the construction of a systematic risk avoidance path to maximize the value of data elements and contribute to achieving a higher level of digital justice.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7048/2025.BR26147

2. The current status of data application in digital courts

In the process of deepening the construction of digital courts, the extensive application of judicial data has become a key engine for improving the quality and efficiency of trials and governance capabilities.

First, the core prerequisite for the development and application of technology is to build an intelligent data collection and governance system. The accumulation of data in the judicial field is of extremely high value. The "206 system "uses natural language processing technology to analyze case facts and legal provisions, and combines machine learning algorithms to automatically classify and compare evidence, building an evidence data chain covering the entire process of filing, handling and reviewing cases [3]. This demonstrates the fundamental supporting role of data governance in judicial quality and efficiency. The system reduces the time for determining evidence in criminal cases by 40% by providing real-time prompts on key points of evidence review and generating correlation analysis reports. In the dimension of judicial data analysis, Internet courts in Beijing, Guangzhou, Shanghai and other places have established dual technology empowerment mechanisms: building case push models through big data mining, using machine learning algorithms to analyze the features of historical judgments, and achieving precise matching in combination with user query intentions; At the same time, a judgment deviation warning system has been developed, using deep learning models to assess the deviation between new precedents and historical data, and automatically triggering a review warning when the threshold is exceeded [4]. This "data collection - intelligent analysis - decision support" full-chain coverage has increased the accuracy of similar case push by 37% and optimized the consistency index of judgments by 29%. Shanghai courts' innovative practices are more in the dimension of social governance, by deeply analyzing judicial big data to uncover the patterns of social contradictions behind cases [5]. These practices collectively validate the fundamental enabling role of technological development in the application scenarios of judicial data. It is evident that the current judicial field has achieved full-chain coverage of technology-enabled judicial data applications, with case handling, similar case adjudication, and social governance all incorporated into the "data collection - intelligent analysis - decision support" system.

Second, platform construction promotes the development of data value. In recent years, coordinated and unified development concepts such as "all-domain digital courts", "digital court system construction", and "one network" have profoundly influenced the development direction of data application in digital courts. Digital court construction takes data development as the core driving force and maximizes the value of judicial data through the dual paths of open sharing and mining awakening. At the level of open sharing, the digital court is committed to breaking down data barriers and building a horizontally and vertically integrated data sharing platform: horizontally integrating data resources from courts with public security, procuratorial, administrative and other departments to form a cross-domain governance synergy; Vertically connect the business systems of courts at different levels to ensure information intercommunication at all nodes of the trial process [6]. Shanghai courts, relying on the "one network" construction, have built the "three major platforms" of big data platform, model platform and automatic operation platform, forming an integrated system of "five major sections" covering case handling, supervision, convenience, governance and government affairs, achieving convergence of trial resources and multi-scenario linkage; Zhejiang courts have built a working platform covering all courts in the province through the "All-Domain Digital Court", which supports the online processing of the entire process of filing, trial and enforcement, and integrates functions such as electronic case file generation, similar case push and document generation, significantly enhancing the digitalization level of judicial business

DOI: 10.54254/2753-7048/2025.BR26147

[7]. In addition, the activation of "dormant data" by digital courts has significantly enhanced the value of the data. Through data deconstruction and model building techniques, the problem of using unstructured data has been effectively solved, laying a solid foundation for deep analysis of artificial intelligence [8]. The unified and systematic development of data has successfully broken down technical barriers and brought about a fundamental shift in judicial data from 'static storage' to 'dynamic empowerment'.This' open sharing - mining and awakening 'two-way mechanism not only significantly enhances the efficiency of data usage, but more importantly, it provides strong support for the precision of intelligent trials and the scientific governance of society.

3. The potential risks of data application in digital courts

3.1. Technical risks

First, risk of data leakage. In the construction of digital courts, the technical root causes of data leakage risk present multi-level and full-chain characteristics, and its technical vulnerability runs through the entire lifecycle management of data application. At the data collection stage, there are inherent flaws in the mixed storage mode of unstructured text and biometric information. This pattern, while enhancing the richness of information, has blurred boundaries between sensitive and non-sensitive fields due to the lack of a unified data classification standard. For example, highly sensitive information such as ID numbers and bank accounts and low-sensitive information such as case facts are collected without differentiated protection, creating a "data mixup risk". Even more serious is the lack of dynamic desensitization mechanisms in the collection of biometric information. An audit by a provincial high court in 2024 found that 32% of judicial data sharing agreements did not specify field level access control, leaving biometric information such as fingerprints and facial data at risk of exposure during the collection stage. The technical contradictions in the storage process are concentrated in the conflict between the centralized architecture and the need for privacy protection. Centralized storage is easy to manage, but the single point of failure risk and the reverse breakthrough of "de-anonymization" technology create a double pressure. From a theoretical perspective, although federated learning technology can achieve data immobility and model mobility in judicial scenarios, the risk of model reverse attack has not been fully studied.

Second, data quality risk. Defects in judicial data quality are systematically eroding the foundation of fairness in digital courts. The lag in the practice of generating electronic case files on a case-by-case basis and the remnants of the traditional paper-based operation mode reflect the structural gap in the transformation of basic data; Even if electronic conversion is achieved, due to the absence of legal semantic parsing technology, a vast amount of unstructured judicial texts, such as administrative loss attribution analysis and procedural records of reconsideration, cannot be transformed into machine-resolvable standardized data units.This governance lapse leads to a double dilemma: the failure of automated capture of key fields such as administrative agencies' responses and judicial recommendations forces judges to spend more than 30 percent of their trial assistance time on manual corrections; The raw data granularity is out of balance - characterized by excessive simplification of the gist of the judgment or redundant piling of procedural nodes, which leads to the loss of a refined basis for subsequent analysis [9]. When such defective data is input into the intelligent adjudication system, the judicial principle of "same case, same judgment" is substantially undermined - the similar case push function, due to algorithmic design limitations that overly rely on surface keyword matching and ignore the relevance of legal elements, results in case similarity generally below the practical threshold of 80% Empirical research in 2024 confirmed that the rate of

DOI: 10.54254/2753-7048/2025.BR26147

sentencing deviation caused by this was 12.7 percent. What's more serious is the "information barrier" formed by the differences in data standards across systems, which makes the cross-jurisdictional reference system of similar cases technically infeasible and poses institutional obstacles to the unification of national adjudication standards.

Third, algorithmic dependency risk. In the early stages of judicial assistance technology training, guidance cases issued by the people's courts are mostly used as the basis of calculation. Even under the principle of "similar cases", the diversity of facts in judicial cases is unpredictable, and regardless of whether the basis is absolutely correct or not, the diversity and complexity of case facts cannot guarantee the absolute repetition of the established model. Instability in either of these two aspects will deal a direct blow to the correctness of the results, and judicial justice is in great danger. From the perspective of the technical operation process, the algorithmic black box problem is a potential risk, which leads to the opacity of the judicial decision-making process and the difficulty in interpreting the results, and algorithmic discrimination may produce unfair results due to training data bias. These technical problems are beyond the prediction and interpretation of human judges [10]. After the fact, human judges can, of course, distinguish the results of their judgments. As a result, the reliance of technology on the opposite, human verification, diminishes its usefulness; Excessive reliance on algorithms may also undermine the subject position of judges in judicial decision-making.The risk of technological alienation and algorithmic dependence is mainly due to the limitations of technology itself and the relentless pursuit of technological efficiency. To address this issue, it is necessary to strengthen the supervision and inspection of algorithmic models to ensure that algorithms make correct decisions, protect the interests of the people, and increase their transparency as much as possible; At the same time, it is necessary to avoid situations where technology replaces judicial judgment due to a lack of experience and technological accumulation.

3.2. Institutional risks

First, data monopoly risk. Third-party technology companies may engage in unfair competition or monopolies around judicial data in the development and application of judicial data, resulting in only a few companies having access to and monopolizing judicial data, which in turn leads to judicial data being controlled by these companies and unable to be shared and interoperable. The risk of data monopoly is mainly due to the lack of effective data governance norms and regulatory mechanisms. Therefore, in order to prevent the risk of data monopoly, it should be addressed from these aspects: first, regulate from the top-level design; second, establish relevant standards to make specific provisions for the open sharing of judicial data. There should also be a sound regulatory system and corresponding legal responsibilities to punish data monopoly behavior.

Second, data regulation dilemma. Given that there are no specific or specialized institutional norms for judicial data governance at present, there are gaps to be filled in how judicial data should be classified and graded in the context of the big data era, how data exchange and collection management should be carried out specifically, and how comprehensive security guarantees should be provided for judicial data during the litigation process. The lack of institutional norms and supporting rules such as filing for judicial data empowerment applications in platforms and courts will further lead to disorder in the development and application of judicial data.

At the same time, on the one hand, the dilemma of data rights confirmation and the fragmentation of regulation are the main problems currently faced; On the other hand, the data rights dilemma is mainly focused on the rights dilemma of judicial data, which, on the one hand, stems from day-to-day administrative management, case hearings and case enforcement work; On the other hand, it stems from the process of digitizing various forms of litigation materials such as case files, litigation

DOI: 10.54254/2753-7048/2025.BR26147

information and electronic data. Due to the large number of traditional paper-based litigation materials being digitized and new types of electronic data, it is difficult to clarify issues such as ownership, usage rights and management rights of the data, and there will be continuous disputes among courts, parties and technology providers in the process of using, sharing and protecting the data. Currently, there are discussions in the academic community about "data property rights" and "data usufruct rights". This also shows that data rights are still in a rather ambiguous state at present.

Regulatory fragmentation is mainly characterized by multiple regulatory bodies and inconsistent regulatory standards [11]. As the construction of digital courts requires the participation of various subjects such as internal regulatory departments of courts, functional departments, third-party technology enterprises, experts, citizens, etc., each subject has different focuses and starting points, and acts independently in data risk assessment, standard formulation, etc., without a unified coordination and linkage department; At the same time, each subject has different requirements for data confidentiality and data security protection standards based on its own functions and the requirements of its work. Within the same system platform, there are uncertain factors in the cross-subject flow of data with different external subjects. As a result, there is a high risk of data security, which affects the comprehensive and effective advancement of the construction of the digital court. For instance, the data standards of local courts vary, cross-provincial approval takes a long time, and judicial information data is difficult to circulate and be effective.

3.3. Ethical risks

Digital court data contains a large amount of personal and sensitive information, which, once leaked, will cause serious harm to people's privacy and the interests of the state. The risk of data leakage also reflects the contradiction between data privacy and ethics. The relationship between the use of data and the protection of privacy should be taken into account in the process of developing and using judicial data. Ethical risks in the construction of digital courts run through the data lifecycle, especially in the field of privacy protection where there is a contradiction between technological dependence and institutional lag. Take the online litigation platform as an example. There is a systemic flaw in the biometric information collection process: when courts implement facial recognition identity verification, they do not establish a dynamic desensitization mechanism, resulting in the facial feature data of the parties being stored in plaintext on the server. Cross-system data sharing raises deeper ethical dilemmas. In the data interaction between online litigation platforms and external technology companies, API interfaces generally lack field-level access control. The risk of secondary dissemination in dynamic data presentation is more concealed. What's more serious is that while blockchain evidence preservation technology ensures data immutability, the public release of the original transaction records leads third parties to reconstruct the complete online behavior trajectory of the parties through correlation analysis.

The privacy protection of digital courts is manifested in three dimensions: First, the scope of data collection has expanded from structured data to unstructured data such as biometrics and online behavior; Second, the data processing has shifted from manual review to machine learning models; Third, the demand for data sharing has shifted from inter-court to cross-departmental and cross-sectoral, according to a 2024 survey by the Supreme Court, 47 percent of courts have not signed confidentiality agreements in data sharing. These particularities call for the establishment of dynamic protection mechanisms that go beyond the traditional judicial secrecy system.

DOI: 10.54254/2753-7048/2025.BR26147

4. Regulatory pathways for data application risks in digital courts

4.1. Improve data security governance systems and full lifecycle management

Data security governance in digital courts should be based on institutional building. Jia Yu pointed out that judicial big data, as a special type of data related to public interest, must be collected, processed and provided in accordance with strict compliance requirements. Based on this, it is suggested that a four-level data classification protection mechanism be established: public data, internal data, sensitive data and core data. For public data such as desensitized judgments, the original metadata should be retained for academic research; Internal data such as trial process information should be limited to access within the court's Intranet and access logs should be retained throughout the process; Sensitive data such as biometric information should be encrypted and stored using special algorithms and transmitted through quantum key distribution networks; Core data, such as information on cases involving national security, should be subject to physical isolation and a two-person review mechanism, and the destruction must be witnessed by the technical department of the provincial court. This classification system is in line with the hierarchical protection requirements of the Data Security Law and responds to the governance needs of the particularity of judicial data.

To achieve full-process management norms, regulations on the management of the Life cycle of judicial data need to be formulated. In the data collection stage, a unified platform should be developed to integrate automatic ID number desensitization and real-time encryption functions to ensure that external data is certified by the Supreme People's Court before access; In the storage stage, a dual mode of "local backup + off-site disaster recovery" is adopted, and structured and unstructured data are managed separately; In the analysis phase, an algorithm filing review system is implemented, and key decision nodes retain manual review interfaces; In the sharing phase, a three-level approval system is established, and cross-departmental calls must be jointly signed and approved by the technical, business, and supervisory departments. Such norms embed data governance into every link of the judicial process, forming a closed-loop management system of "collect - store - analyze - share". The specific norms for data collection, storage, transmission, use and destruction can be referred to the Supreme People's Court's "Measures for the Administration of Data Security of the People's Courts".

4.2. Establish a data interaction, sharing and security technology guarantee system

Technical support, as the core support for data security in digital courts, requires the construction of a protection system covering the entire chain of "storage - flow - application". At the data storage stage, dynamic desensitization technology becomes a key line of defense. Based on Liu Chang's concept of data security protection, biometric information is irreversibly encrypted. For structured data such as ID numbers, implement a "partial hide + dynamic replacement" strategy, such as converting "330102" to "330***", which retains the regional identification function while avoiding the exposure of complete information. To verify the effect of the desensitization, an attack simulation test platform was developed to test the desensitized data through various possible attack methods. The 2024 Hangzhou Internet Court pilot showed that the platform successfully intercepted 98.7% of the simulated attacks, reducing the risk of data leakage by 67% compared to the traditional approach.

At the data flow stage, data lineage tracing technology enables full-process visualization control. The judicial data lineage tracking platform was deployed with a dual-core architecture of "metadata

DOI: 10.54254/2753-7048/2025.BR26147

engine + graph database" in line with Li Zhanguo's emphasis on lifecycle management. The visual interface supports three query modes: the case dimension shows the entire process of the evidence chain, the time dimension traces the history of data changes, and the personnel dimension analyzes the frequency of operations and abnormal behaviors. To ensure real-time performance and achieve second-level synchronization of data changes, tests by the Higher People's Court of Guangdong Province in 2024 showed that the architecture increased the efficiency of data flow monitoring by 300% and shortened the response time to abnormal events to within 15 seconds.

At the data application stage, the AI-generated content detection system builds the last line of defense. Based on the generative data evidence theory proposed by Wang Shixuan, the detection system adopts a hybrid architecture of "pre-trained model + fine-tuning". The base model was pre-trained on 500,000 historical judicial documents and 200,000 false statement samples to grasp the semantic features of judicial texts. To control the false positive rate, a dynamic threshold adjustment algorithm was developed to automatically adjust the confidence interval according to the case type. Empirical evidence from the Beijing Intellectual Property Court in 2024 showed that the system had an accuracy rate of 92.3% in identifying deepfake audio. The detection results are pushed to the judge's work platform in real time via the WebSocket protocol, and key evidence is automatically marked and retained in the review trajectory, forming a two-level verification mechanism of "AI initial screening + human review ".

4.3. Improve risk assessment, ethical adaptation and standardized governance paths

Ma Changshan proposed that digital justice needs to strike an appropriate balance between judicial initiative and judicial restraint, data business and digital justice, and digital technology and legal humanism, emphasizing the legitimacy of digital power, judicial system reform, and the determination [12] of digital justice principles. Ethical constraints are the value orientation of data governance. Build a "three rights separation" judicial data rights confirmation mechanism: clarify the right of litigation participants to be informed of the data they submit and allow them to apply to view the data usage trajectory; Establish a data usage licensing system between courts and third-party technology companies, requiring the signing of a "Judicial Data Usage Commitment Letter"; Explore a data value feedback mechanism and offer incentives such as reduced litigation fees to parties involved in data governance. Such arrangements strike a balance between the use of data and the protection of rights.

The people's courts may establish digital court ethics review committees to perform the functions of algorithm review, risk assessment and standard setting. Algorithm review requires an assessment of the explainability of the decision-making logic of the intelligent trial system, with human intervention channels reserved at key nodes; Risk assessment Regularly conduct social impact assessments of data application, with a focus on monitoring risks such as algorithmic discrimination and privacy leakage; Standards are being developed by drafting the "Ethical Guidelines for Judicial Artificial Intelligence", which define 20 core criteria including the scope of data collection and algorithmic transparency.

At the same time, the "Four Modernizations" standardization project should be advanced. Compile a dictionary of judicial data elements, unify more than 200 core concepts such as "electronic case files" and "similar case push"; Formulate the acceptance Standards for the construction of digital courts and incorporate data governance capabilities into the performance assessment of courts; Develop a unified API protocol to address the technical barriers to data intercommunication between courts and public security and procuratorial systems; Establish a judicial data security authentication system, and only systems that pass the authentication can access

DOI: 10.54254/2753-7048/2025.BR26147

the court Intranet. Such standards provide a uniform scale for technical implementation and institutional enforcement.

5. Conclusions

Although the construction of digital courts is the direction of judicial modernization, there are still certain risks and concerns in the application of data. In addition to improving the data classification and grading protection system and establishing a sound data lifecycle management system, it is also necessary to pay attention to strengthening the security guarantee of data interaction and sharing, improving the data security risk assessment and emergency response, promoting the adaptation of data ethics to judicial laws, and promoting the standardization and normalization of judicial data application systems and measures. With the acceleration of the construction of digital courts and the continuous development of technology, opportunities and risks coexist. This requires the continuous strengthening of research and improvement of supervision and regulation, so that the construction of digital courts can maintain social fairness and justice and meet the new demands of the people for fairness and justice in the process of continuously advancing judicial modernization.

References

[1] Guo Yan, Hu Mingdong. Shanghai: Focusing on New Tasks to Expand New Achievements in Digital Court Construction [N]. People's Court Daily, 2025-5-19(01).
[2] Li Huaisheng. From Static Defense to Dynamic Security: Paradigm Shift in Legal Protection of Data Security [J]. Rule of Law Studies, 2025, (01): 118-131.
[3] Jia Yu. On Digital Courts [J]. Legal Studies, 24, 46(04): 3-20.
[4] Liu Li, Liu Zhiyi. The Construction of Digital Courts in the Age of Artificial Intelligence: A Triple Approach of Data, Governance, and Intelligence [J]. Digital Rule of Law, 2024, (05): 150-167.
[5] Jia Yu. Research on the Construction Path of Digital Courts [J]. Digital Rule of Law, 2024, (03): 14-25.
[6] Ma Changshan. Digital boundary [J]. The rule of law in judicial Oriental law, 2024, (4) : 127-142. The DOI: 10.19404 / j.carol carroll nki DFFX. 20240712.005.
[7] Zhang Jiajie, Xiao Ke. Research on Judicial Big Data Boosting Digital Social Governance [J]. Digital Rule of Law, 2024, (05): 136-149.
[8] Li Zhanguo. Construction and Implementation of "All-Domain Digital Court" [J]. Chinese and Foreign Jurisve., 2022, 34(01): 5-24.
[9] Wang Shixuan. The Legal Basis of "Data-Assisted Case handling" : A Three-dimensional Perspective of Substantive Law, Evidence Law and Information Law [J]. Digital Rule of Law, 2024, (03): 54-67.
[10] Lei Lei. Theoretical Reflections on Digital Justice: Significance, Issues and Regulation [J]. Jiaotong Law, 2024, (06): 86-100.
[11] Nie Zhenhua, Yang Jielin. The Dilemma, Reinterpretation and regulation of AI-assisted sentencing [J]. Digital Rule of Law, 2024, (05): 168-180.
[12] Ma Yu. Research on Problems and Countermeasures in the Construction of Smart Courts [J]. Smart Courts Section: Innovative Experience on "Research on Problems and Countermeasures in the Construction of Smart Courts", 2024.