AI in the Court

Natarajan Balasubramanian, Chenyang Pan, Jian Xie, Wenjian Xu[1]

Abstract: This paper examines how the adoption of artificial intelligence (AI) technologies influences judicial performance in China. We use large language models to identify AI procurement in comprehensive court purchasing records and construct measures of court performance from more than 146 million judicial documents. Employing a staggered difference-in-differences framework, we find that AI adoption reduces appeal rates by roughly 28.9% and reversal rates by 39.6%, while substantially speeding up case processing. Classifying AI technologies into trial-specific and routine-specific applications, we find that the former drives large improvements in judicial quality while the latter primarily generates efficiency gains. These performance gains are especially pronounced in courts with complementary digital infrastructure, for more complex cases, and in less developed regions, suggesting that AI helps alleviate capacity and information constraints in resource-limited judicial settings. We further show that AI adoption generates significantly larger improvements than other digital technologies, highlighting its unique contribution to judicial decision-making. Taken together, our findings temper concerns about AI use in the public sector and demonstrate its potential to strengthen institutional performance.

Keywords: Artificial Intelligence, Judicial Performance, Courts, Fairness and Efficiency, China

JEL classification: K40, O33, O17, H83

[1] Balasubramanian: Whitman School of Management, Syracuse University, nabalasu@syr.edu. Pan: Antai College of Economics and Management, Shanghai Jiao Tong University, stevenpan.econ@gmail.com. Xie: College of Business, Southern University of Science and Technology, xiej@sustech.edu.cn. Xu: corresponding author, Antai College of Economics and Management, Shanghai Jiao Tong University, wenjianxu@sjtu.edu.cn.

Acknowledgments: We are grateful to Yige Duan, Xi Zhi Lim, Jialu Liu, Zhuoran Lu, Yuta Suzuki, Li Xu, Qinshu Xue, Fangjin Ye, Bernard Yeung, Yu Zeng, and seminar participants at the 7th AI and Strategy Consortium, the Eighth Annual Conference of Government and Economics, the Chinese University of Hong Kong, Shenzhen, Shanghai Jiao Tong University, and Xiamen University. Any remaining errors are our own. Authors contributed equally to this work and are listed alphabetically.

1. Introduction

The global diffusion of Artificial Intelligence (AI) represents a profound technological transformation, fundamentally reshaping economic activity and governance across jurisdictions (Agrawal et al., 2019). While a substantial literature documents how AI adoption in the private sector fuels productivity, innovation, and aggregate growth (e.g., Aghion et al., 2017; Acemoglu and Restrepo, 2018; Athey, 2018; Brynjolfsson et al., 2025), the role of this disruptive technology in the public sector is far less understood and significantly more contested. The application of AI in critical state functions, such as public administration and the judicial system, promises gains in efficiency and decision-making quality. Yet, these deployments simultaneously raise serious concerns over potential algorithmic bias, the amplification of existing inequalities, and the erosion of institutional accountability (Acemoglu, 2021; Kleinberg et al., 2018b; Rambachan et al., 2020). These issues are especially salient in the judicial system, where decisions directly shape fairness in society and the credibility of public institutions. This paper provides a rigorous, large-scale empirical assessment of this trade-off by examining how AI adoption in Chinese courts impacts judicial performance, specifically focusing on its influence on both the quality and efficiency of legal institutions.

The integration of AI extends across the judicial lifecycle, from evidence organization and trial assistance to decision support and document drafting, holding the potential for a fundamental overhaul of operations. Proponents argue that AI can significantly boost efficiency by automating routine tasks, alleviating administrative burdens, and enabling courts to manage massive caseloads; by handling mechanical tasks, judges are theoretically liberated to focus scarce time on substantive legal deliberation. Furthermore, advanced AI tools, such as similar-case recommendation systems, can enhance the consistency of rulings by swiftly identifying relevant precedents and applying legal provisions uniformly. However, the perils of this reliance are equally profound. The use of opaque ‘black-box’ models, inherent technological imperfections, the potential for algorithmic bias, and the risk that automated processes may ignore the complex social impacts of judicial decisions threaten to systematically distort judicial outcomes, thereby undermining the very transparency, fairness, and legitimacy of the legal institutions that underpin the rule of law. Resolving the fierce controversy over AI’s promises and risks in such a high-stakes environment is

paramount, particularly because the quality of legal institutions is a critical determinant of economic development—governing property rights, contract enforcement, and market incentives (La Porta et al., 1997; Djankov et al., 2003; Beck and Levine, 2005; Castro et al., 2009; Chemin, 2009; Beck, 2012; Chemin, 2020). Moreover, as governments globally accelerate their experimentation with AI in judicial and administrative systems, rigorous evidence is urgently needed to inform the design of effective regulations, guidelines, and safeguards that appropriately balance efficiency gains with the imperative of fairness.

We address this question in the unique and highly relevant context of Chinese courts. China is at the global forefront of AI development and has aggressively adopted these technologies in the public sector, making it an ideal setting to study their frontier effects (Beraja et al., 2023). As early as 2016, the Chinese government put the development of AI technology in its five-year (2016-2020) plan.[2] To strengthen the application, the Chinese government issued "Notice of the State Council on Issuing the New Generation Artificial Intelligence Development Plan" to encourage the AI adoption in the public sector.[3] Crucially, the institutional setting affords access to uniquely granular data on both courts' procurement and judicial documents, enabling a systematic, court-level empirical evaluation. We identify AI adoption using a novel dataset of over 51,000 comprehensive public procurement records of Chinese courts, employing large language models to pinpoint approximately 222 AI-related procurements that detail the acquired technology. To measure court performance, we exploit more than 146 million judgment documents from the China Judgments Online (CJO) database, spanning 2014–2022. Our primary measures of judicial decision quality are the appeal rate, which reflects litigants' perceived fairness of the first-instance ruling, and the reversal rate on appeal, which serves as a metric for the consistency and substantive quality of judicial outcomes. Additionally, we analyze efficiency through case processing speed, specifically the time from case filing to trial.

Using a staggered difference-in-differences specification with quarterly-level observations, we find that judicial AI adoption significantly improves court performance. Specifically, AI adoption reduces the appeal rate by around 28.9%,

[2] The full document of this five-year plan can be found here: https://www.ndrc.gov.cn/fggz/fzzlgh/gjfzgh/201603/P020191104614882474091.pdf.
[3] The policy details can be found here: http://www.scio.gov.cn/gwyzclxcfh/cfh/2017n_14540/2017n07y21r_14662/wjxgzc_14668/202208/t20220808_298141.html.

suggesting that judicial decisions become more acceptable to litigants and are thus less frequently contested. Moreover, in courts that adopt AI, the reversal rate on appeal declines by an estimated 39.6%, indicating an improvement in the quality and consistency of first-instance trials.

To test for the presence of pre-trends, we employ an event-study framework, which shows that the judicial performance of AI-adopting and non-adopting courts followed similar trajectories prior to adoption. Performance improvements emerge only after AI adoption, supporting the validity of our identification strategy. To further account for potential heterogeneity in treatment effects across time and treated units, we also estimate event-study models using recently developed heterogeneous-treatment-effect-robust estimators (De Chaisemartin and d'Haultfoeuille 2020; Borusyak, Jaravel, and Spiess 2024; Callaway and Sant'Anna 2021; Sun and Abraham 2021). The results remain consistent with our baseline findings.

We then conduct a series of robustness tests. First, we address concerns that AI adoption may coincide with concurrent reforms implemented during our sample period. we construct indicators for the timing of the judge quota reform, judicial independence reform, and "smart court" reform, and control for them in our baseline regressions. Second, we drop the year 2022, when there was a sharp reduction in published decisions. Third, we use alternative LLM-based measures to identify AI adoption. Fourth, we account for court size by weighting regressions by case and judge number separately, and by directly controlling for court size. Overall, our results are robust to all these tests.

To address endogeneity concerns, we employ an instrumental variable based on AI adoption by city-level environmental protection bureaus and tax bureaus. This IV is relevant because courts and other government departments within the same city face similar budgetary constraints, policy directives, and technology-promotion efforts. It should satisfy the exclusion restriction because city-level AI adoption patterns by environmental protection bureaus and tax bureaus are unlikely to be driven by contemporaneous shocks specific to county-level courts. AI adoption by environmental protection bureaus and tax bureaus should affect local courts' judicial outcomes only through courts' own AI adoption. The IV estimates further support our baseline findings.

This general improvement in performance is not merely a by-product of generic digitalization; instead, our analysis points to a distinct, technology-specific mechanism. By controlling for the adoption of general digital technologies, we demonstrate that AI adoption is responsible for significantly larger reductions in both appeal and reversal

rates, while broader digital tools show weaker effects. This divergence reflects the core logic of our story: conventional digital systems primarily support basic administrative functions, whereas sophisticated AI systems (e.g., those providing decision-support or precedent-matching) directly assist judges with complex, substantive tasks, thereby improving both appeal and reversal outcomes.

Beyond improvements in judicial decision quality, we also find evidence of faster case processing. The duration from case acceptance to judgment falls by 11.1%. However, we do not find statistically significant effects on judicial discretion.[4] We further examine whether AI technologies substitute for or complement labor in courts. The results show that AI adoption significantly reduces the number of judges and court clerks, suggesting a labor-substitution mechanism that may affect both decision-making roles and routine administrative tasks.

We then explore heterogeneity in the effects of judicial AI adoption along several dimensions: AI technology type, AI readiness, procurement amount, regional development, and case type. First, we find that trial-specific AI adoption consistently reduces both appeal rates and reversal rates. In contrast, routine-specific AI adoption does not generate statistically significant effects. Second, we find that the performance gains from AI adoption are larger for courts with higher AI readiness, proxied by whether having procured digital technology before, suggesting the importance of previously accumulated complementary digital capabilities. Third, AI procurements of higher contract amount lead to significantly larger reductions in both appeal and reversal rates, consistent with the idea that more substantial investments bring about greater performance improvements. Fourth, consistent with our model's prediction, we find that AI adoption yields the strongest benefits in less developed regions—courts located in provinces with lower GDP per capita, and in western China areas. Finally, AI generates substantially larger improvements for complex cases than for non-complex ones, in line with our model's results, underscoring its role in helping courts manage more intricate and information-intensive disputes.

These findings carry two important policy implications. First, the reductions in appeal and reversal rates highlight AI's potential to enhance both fairness and

[4] Following Liu et al. (2024), we define a case as involving judicial discretion if the presiding judge cites "discretionary codes" in their reasoning, and we then calculate the share of such cases at the court-quarter level. We thank the authors of Liu et al. (2024) for generously sharing their detailed list of discretionary codes with us.

consistency in judicial outcomes, demonstrating the broader value of technological innovation in the public sector and the importance of integrating AI into other areas of governance. Second, the fact that performance improvements are most pronounced in less developed regions suggests that AI deployment should be prioritized in resource-constrained courts, where the technology can help narrow regional disparities in judicial quality and improve access to justice.

Related Literature—We contribute to three strands of literature. First, we add to the growing literature on the consequences of AI adoption in the public sector (Kleinberg et al., 2018a; Beraja et al., 2023; Tirole, 2021; Yeung and Lodge, 2019; Korinek, 2024; Capraro et al., 2024; Javed and Li, 2025). Kleinberg et al. (2018a) show that machine learning can improve judges' jail-or-release decisions in the United States, while Beraja et al. (2023) find that AI adoption by local governments in China reduces the incidence of local civil unrest. To the best of our knowledge, we are the first to offer a comprehensive empirical assessment of the impact of AI adoption on judicial performance, integrating both fairness and efficiency outcomes. Our results on enhanced judicial performance complement the study by Fang et al. (2025) on the impacts of the traditional information system (Case Management/Electronic Case Files) on judicial efficiency in U.S. courts, and provide direct evidence for the literature debating the risks and concerns associated with AI adoption in the public sector, particularly the risks of algorithmic bias and amplified inequalities (Acemoglu, 2021; Kleinberg et al., 2018b; Rambachan et al., 2020). Furthermore, our evidence that performance improvements are especially pronounced in less developed regions highlights the potential for AI to strengthen institutional performance in resource-constrained environments and to promote more equitable access to justice. More broadly, we also contribute to the expanding literature on how AI shapes human decision making (e.g. Hoffman, Kahn, and Li 2018; Angelova, Dobbie, and Yang 2023; Agarwal et al. 2023; Boyacı, Canyakmaz, and De Véricourt 2024; Vaccaro, Almaatouq, and Malone 2024) and how AI affects productivity (e.g. Aghion et al., 2017; Acemoglu and Restrepo, 2018; Athey, 2018; Noy and Zhang 2023; Babina et al. 2024; Kanazawa et al. 2025; Brynjolfsson et al., 2025).

Second, we contribute to the literature on judicial decision-making and judicial performance. A large body of research shows that gender, race, political context, and other factors shape judicial behavior and performance across a wide range of settings (e.g. Shayo and Zussman, 2011; Kastellec, 2013; Ash et al., 2024; Liu et al., 2024; Ng

and He, 2017; Popova, 2012; Caplin et al., 2024; Bhuller and Sigstad, 2025; Ash, Chen, and Naidu, 2026). In addition, several studies examine the impact of technological advances in judicial systems in China. For example, Chen et al. (2025) study how information technology, specifically live broadcasting, shapes judicial decision-making, showing that it can help narrow gender disparities in litigation outcomes. Cao, Ding, and Zhou (2024) show that digital transformation, specifically the adoption of e-commerce platforms for judicial auctions, enhances the performance of Chinese courts. Our study differs from this literature by focusing on the adoption of AI, a rapidly expanding technology with distinct implications for judicial functioning. Moreover, by using appeal rates and reversal rates as measures that reflect litigants' overall satisfaction and the consistency of judicial outcomes, we depart from prior studies that concentrate mainly on judicial biases toward specific groups.

Third, we contribute to the broader literature on how technological advances shape public affairs. A large body of research shows that information technology has significantly improved outcomes in the public sector. For instance, Garicano and Heaton (2010) and Mastrobuoni (2020) show that IT adoption enhances productivity in police departments, while Athey and Stern (2000) and McCullough et al. (2016) find that IT adoption improves health care outcomes. West (2005) and Milakovich (2012) discuss how digital technologies can influence public sector performance. Elbahnasawy (2014) and Kanyam et al. (2017) examine how information technology can reduce corruption. We advance this literature by shifting the focus to the judiciary, a cornerstone of public institutions whose decisions shape the protection of property rights, the enforcement of contracts, and the incentives that underpin economic participation. By examining how frontier technologies transform this arena, our study highlights a domain where technological change can have especially far-reaching consequences for both economic and social life.

2. Background

2.1 The Chinese Judicial Systems

China operates a hierarchical judicial system comprising four distinct tiers: the Supreme People's Court (SPC) at the apex, followed by Higher Level People's Courts (HLPC), Intermediate Level People's Courts (ILPC), and Basic Level People's Courts (BLPC) at the foundation. As of December 2024, this structure encompasses 3,117

basic courts (typically one per county), 409 intermediate courts (typically one per prefecture), 32 higher courts (generally one per province), and one Supreme Court.

The Supreme People's Court serves multiple critical functions within this hierarchy. It adjudicates the most significant cases across all legal domains, formulates authoritative judicial interpretations that guide lower courts, supervises the adjudication work of all local and specialized courts, and oversees the administration of justice nationwide. This centralized oversight ensures consistency in judicial decision-making across the country's vast legal landscape.

Case allocation within this tiered system follows clear jurisdictional principles based primarily on the monetary value and societal impact of disputes. Under current standards established by the SPC, the overwhelming majority of first-instance civil cases fall within the jurisdiction of basic people's courts, with only exceptional cases of particular complexity or significance being heard directly by intermediate courts (Zhang, 2023).[5] This distribution ensures that routine legal matters are handled efficiently at the local level while reserving higher courts for cases requiring specialized expertise or having broader implications.[6]

China's four-level court system operates under a two-instance final adjudication framework, whereby cases may be heard at a maximum of two judicial levels before reaching final resolution. This system serves two primary institutional purposes. First, it ensures that the vast majority of civil cases can be resolved within the jurisdiction where the parties reside, thereby facilitating access to justice while reducing the litigation burden on individual litigants. Second, it allows higher people's courts and the Supreme People's Court to focus their limited resources on providing judicial guidance and supervision rather than adjudicating routine cases.[7] Under this framework,

[5] For example, for the civil cases, which constitute the major part of the lawsuits, the Supreme People's Court issued the Notice on Adjusting the Standards for the Jurisdiction of First-Instance Civil Cases by Higher People's Courts and Intermediate People's Courts (Fa Fa [2019] No. 14) and the Notice on Adjusting the Standards for the Jurisdiction of First-Instance Civil Cases by Intermediate People's Courts (Fa Fa [2021] No. 27). These regulations establish that higher people's courts have jurisdiction over first-instance civil cases with claim amounts exceeding 5 billion RMB or cases with significant regional influence, while intermediate people's courts handle first-instance civil cases with claim amounts exceeding 500 million RMB (when all parties' domiciles are either within or outside the accepting court's province-level administrative region). For detailed analysis, see: http://www.ilawchina.com/html/detail/476.html.

[6] Beyond this four-tier structure, the SPC has established specialized courts to address specific areas of legal complexity. These include maritime courts for shipping and ocean-related disputes, intellectual property courts for patent and trademark matters, and financial courts for banking and securities cases. This specialization allows for the development of sector-specific expertise and ensures that complex technical matters receive appropriately informed adjudication.

[7] For detailed analysis, see: https://cjs.fudan.edu.cn/info/1026/1149.htm.

parties dissatisfied with a first-instance judgment retain the right to appeal to the next higher court within a statutorily prescribed timeframe. Appeals follow the hierarchical structure, progressing from Basic People's Courts to Intermediate People's Courts, from Intermediate to Higher People's Courts, and ultimately to the Supreme People's Court when circumstances warrant such review.

The adjudication rendered by the higher court constitutes the definitive outcome for the vast majority of cases. Consequently, second-instance proceedings provide direct evidence of litigant dissatisfaction with initial first-instance judgments. The appeal rate has emerged as a critical performance indicator within the People's Court system for evaluating judicial quality and has been widely adopted in law and economics literature as a measure of court performance (e.g., Liu et al. 2024; Baye and Wright 2011).[8]

2.2 Judicial Disclosure in China

China's evolution toward judicial transparency has progressed through two distinct phases: a voluntary disclosure and experimentation period prior to 2014, followed by a mandatory disclosure regime beginning in 2014. The foundations for judicial openness were established as early as 2000, when the SPC first recommended selective disclosure of judgment information for certain case categories. This initial framework was subsequently expanded in 2007, when the SPC issued formal guidance directing all provincial higher people's courts to disclose information regarding their judicial decisions. Throughout this preliminary phase, judicial transparency gradually emerged as a central pillar of China's broader court system reform agenda.

The transition to comprehensive mandatory disclosure accelerated substantially following Zhou Qiang's appointment as Supreme People's Court President in 2013. Under his leadership, the pace of judicial openness reform increased dramatically. In November 2013, the SPC issued a landmark directive requiring all Chinese courts across all jurisdictional levels to make their decisions publicly available through a centralized online platform. This mandate established China Judgments Online (CJO, https://wenshu.court.gov.cn/), a comprehensive digital repository created and administered by the SPC. The policy became effective in January 2014, marking China's definitive entry into an era of mandatory comprehensive judicial disclosure.

[8] The SPC emphasizes that primary indicators for assessing judicial performance include the rates of case remand and judgment reversal in second-instance proceedings.

To ensure effective implementation and compliance, the Supreme People's Court introduced a quarterly reporting system in early 2014 to monitor the nationwide online disclosure of judicial decisions. This oversight mechanism reinforces supervision, inspection, and policy guidance across all levels of the court system. In parallel, higher people's courts have established their own reporting frameworks to track and promote disclosure practices within their respective jurisdictions.[9]

The evolution of disclosure activity is illustrated in Figure 1, which presents the annual number of cases published on the CJO platform, the total number of cases concluded, and the share of published cases from 2014 to 2022. The data reveals a sharp upward trajectory from 2014 through 2020, with publication volume reaching its peak in 2019. However, this trend reversed dramatically thereafter, with a substantial decline in published cases. By 2023, the number of online cases had fallen to about 7.2 million—less than 16% of the total.[10] According to SPC officials, this reduction in disclosure reflects growing institutional concerns regarding platform usability, individual rights protection, and information security risks.[11] Given the low coverage of published cases prior to 2014 and after 2022, our empirical analysis focuses on the 2014–2022 period.

Judicial documents uploaded to the CJO platform contain comprehensive case information, including case identification numbers, party names, full document content, trial judge and court personnel identities, and judgment dates. Figure A1 in the Online Appendix presents an example of judicial document. While transparency is the governing principle, certain sensitive cases may feature fully or partially redacted party names for privacy protection. All documents are mandatorily required to disclose the real names of trial judges, clerks, and other relevant court staff members. The substantive content of most judgment documents follows a standardized format that includes a concise case description, detailed reasoning for the court's decision (encompassing both factual analysis and applicable legal provisions), and the final judgment outcome. Second-instance court modifications of initial judgments typically stem from several categories of first-instance deficiencies: erroneous factual

[9] For more details, please see the report: https://paper.people.com.cn/zgjjzk/html/2024-01/30/nw.zgjjzk_20240130_2-02.htm.
[10] The total number of cases concluded in year 2023 was about 45.27 million, according an SPC report (https://www.court.gov.cn/zixun/xiangqing/427022.html).
[11] For more details, please see the report: http://www.legaldaily.com.cn/Village_ruled_by_law/content/2024-01/18/content_8951881.html.

assessments, procedural irregularities during the adjudication process, or improper application of legal statutes. Consequently, second-instance verdicts frequently contain explicit information regarding the affirmation or reversal of first-instance decisions, thereby providing direct empirical evidence of lower court decision quality and serving as a valuable metric for judicial performance assessment.

2.3 Increased Litigation Volume and Judicial Resource Constraints

China's judicial system has experienced unprecedented growth in litigation volume over the past decade, with total cases received by courts more than tripling from 14.21 million in 2013 to 45.57 million in 2023, representing an average annual growth rate of 13%. [12] This dramatic increase in caseload has occurred simultaneously with a deliberate reduction in judicial personnel through comprehensive quota system reforms in order to optimize personnel structure, restructure evaluation system, adopt scientific classification management, and precisely define functions.[13]

Before the Judge Quota Reform, the Chinese court system faced serious overstaffing and consisted of a significant number of judges lacking proper legal training or adjudication experience due to inefficient judicial management (Feng, 2015; Chen et al., 2024). In 2013, the Supreme People's Court initiated ambitious judicial reforms centered on the principles of formalization, professionalization, and specialization. A cornerstone of this reform was the establishment of a strict quota system limiting the proportion of qualified judges to 39% of the authorized complement of political-legal personnel within each court.[14]

This simultaneous expansion of caseload combined with this Judge Quota Reform has imposed severe work burden on judges. The average annual caseload per judge has increased dramatically from 65 cases in 2013 to approximately 357 cases in 2023—a fivefold increase that far exceeds sustainable workload levels.[15] This acute imbalance

[12] Data sources: Work Report of the Supreme People's Court, available at:: http://lianghui.people.com.cn/2014npc/n/2014/0310/c382480-24592263.html and https://www.gov.cn/yaowen/liebiao/202403/content_6939583.htm, respectively.
[13] For detailed analysis of the reform rationale, see the SPC's official explanation at: https://www.sdcourt.gov.cn/ytzyfy/402518/402565/1574466/index.html.
[14] The quota reform also enhanced incentives for selected judges through substantial salary increases. The Pilot Plan for the Reform of the Salary System for Judges and Prosecutors (Ren She Fa [2015] No. 111) stipulates that judges admitted to the quota system receive salaries approximately 50% higher than other civil servants in the same locality. For detailed analysis, see: http://www.socio-legal.sjtu.edu.cn/wxzy/info.aspx?itemid=3943.
[15] For detailed analysis, see: http://theory.people.com.cn/n1/2025/0426/c40531-40468622.html and https://www.gov.cn/yaowen/liebiao/202403/content_6939583.htm.

between rapidly expanding litigation demands and constrained judicial resources has emerged as a critical systemic challenge, creating strong institutional incentives for the adoption of artificial intelligence technologies to enhance judicial efficiency.

2.4 Smart Court Reforms and Judicial AI Adoption

As early as 2016, Chinese government started to embrace judicial artificial intelligence stating in the Outline of the National Informatization Development Strategy and the 13th Five-Year National Informatization Plan, both explicitly identifying "Smart Courts" construction as an integral component of the national informatization development strategy. Building upon this policy framework, the SPC issued comprehensive implementation rules for smart court construction in 2017, which formally mandated the integration of big data and artificial intelligence technologies into judicial operations.

The 2017 implementation rules specifically advocate for AI technologies including speech recognition and automated text generation to enhance the efficiency of judicial document drafting and to reduce the administrative burden of non-judicial tasks on adjudicators. This approach aligns with broader economic research demonstrating AI's comparative advantage in automating mechanical, repetitive, and routine tasks (Demirci, Hannane, and Zhu 2025; Autor 2024; Acemoglu 2025). Beyond efficiency gains, the SPC emphasized AI's governance potential, advocating for its application in monitoring and regulating judicial discretion to prevent abuse of judicial authority and corruption—persistent challenges in developing countries undergoing legal transformation while maintaining high levels of procedural formalization (Djankov et al. 2003).

The ongoing Smart Court Reform has served as a powerful catalyst for AI technology adoption across all levels of China's judicial hierarchy. This top-down reform initiative, combined with mounting litigation pressures on judges, provides a relatively exogenous source of variation in courts' AI adoption behaviors—a key advantage for empirical analysis. The reform's impact has been amplified by China's rapid advancement in AI technology development, particularly the emergence of multimodal large language models that have accelerated progress in text, image, and audio recognition technologies, as well as automated text generation capabilities.

AI integration in Chinese courts now encompasses a comprehensive suite of applications spanning legal education and public outreach, case information processing,

sentencing assistance, automated drafting and verification of judicial documents, and judicial quality evaluation systems. Figure A2 in the Online Appendix illustrates the process flow for a typical judicial decision, in the same way as Balasubramanian, Lee, and Sivadasan (2018) show procedures for patent prosecution. While substantial anecdotal evidence suggests that AI adoption can enhance both judicial quality and productivity, systematic empirical evidence and rigorous analytical frameworks have remained limited.[16] This gap in the literature motivates our comprehensive empirical investigation into the causal impacts of AI adoption on measurable court performance outcomes.

3. Theoretical Framework

To understand the role of AI in the court, we develop a task-based model of AI-assisted judicial decision-making, building on the framework of Acemoglu and Restrepo (2018) and Hampole et al. (2025). The model treats AI as a technology that raises the productivity of routine judicial tasks, freeing judges' cognitive capacity for substantive deliberation. Three testable predictions emerge: AI adoption improves average decision quality; the gain is larger for more complex cases; and the gain is larger in courts with tighter cognitive-capacity constraints. These predictions correspond directly to the main empirical findings in Sections 4 and 5.

We define the quality of a judicial decision Q as a function of the judge's ability to process "information units" I correctly. A case of complexity θ requires the successful completion of a continuum of tasks $j \in [0, \theta]$.

The total judicial output (quality) for a representative case with complexity level θ is (In this framework, "Complexity" (θ) is modeled as the length of the task continuum):

$$Q_\theta = \int_0^\theta \ln\left(y_j\right) dj$$

where y_j is the performance on task j. Following Acemoglu and Restrepo (2018), tasks are divided into two types:

1. Routine/Technical Tasks $(j \in [0, \tau])$: Precedent searching, document verification, and cross-referencing.

[16] For example, a report by the Xinye County People's Court shows that the pilot program of the 'AI Intelligent Judge Assistant' has helped improve adjudicative quality and efficiency (https://www.hncourt.gov.cn/public/detail.php?id=198918). Similar reports have been found in Shanghai and Shenzhen: http://epaper.legaldaily.com.cn/fzrb/content/20240528/Articel04003GN.htm and https://szjgdj.sz.gov.cn/home/tzqt/content/post_1411912.html, respectively.

2. Substantive/Deliberative Tasks ($j \in (\tau, \theta]$) : Evaluating witness credibility, balancing legal principles, and social impact assessment.

Following the logic of Hampole et al. (2025), AI increases the productivity of the "routine" segment. Let h_j be human effort on task j. The production for task j is:

$$y_j = \left[\left(\alpha_j A\right)^{1-\sigma} + h_j^{1-\sigma}\right]^{\frac{1}{1-\sigma}}$$

where A is the level of AI technology, α_j is the "AI-suitability" of task j. We assume α_j is high for routine tasks ($j \leq \tau$) and low for substantive tasks ($j > \boldsymbol{\tau}$), σ is the CES substitution parameter between AI and human inputs, with 1/σ equal to the elasticity of substitution.

The court faces cognitive capacity constraint: $\int_0^\theta h_j dj \leq H$. The court chooses $\{h_j\}_{j\in[0,\theta]}$ to maximize Q_θ. Let the optimal value be $Q^*(A,\ \theta, \alpha_j, \sigma)$.

Let Q*(A, θ, H) denote the court's optimized judicial quality. Applying the envelope theorem and the Leibniz integral rule to the problem above, we obtain the following three propositions; formal proofs appear in the Online Appendix.

Proposition 1 (H1). AI adoption has a positive effect on court performance: $\partial Q^*/\partial A > 0$.

AI raises the effective productivity of every task j along the continuum [0, θ], increasing y_j for all j ∈ [0, θ] and therefore increasing the integrated quality measure Q*. The gain is strictly positive for any level of AI technology A and any human capacity H.

Proposition 2 (H2). The positive impact of AI adoption is larger among complex cases: $\partial^2 Q^*/\partial A \partial \theta > 0$.

More complex cases span a wider task continuum [0, θ]. By the Leibniz rule, a marginal increase in θ adds one more AI-augmented task to the quality integral, so the cumulative benefit of AI grows with case complexity. This predicts larger reductions in appeal and reversal rates for complex cases.

Proposition 3 (H3). When AI and human effort are substitutes (σ < 1), AI adoption yields larger marginal quality gains in courts with lower human capacity H: $\partial^2 Q^*/\partial A \partial H < 0$.

When σ < 1, AI and human cognition can substitute for each other. Courts in less developed regions face tighter cognitive-capacity constraints (lower H), so AI compensates most where human resources are scarce. This predicts that performance

gains from AI adoption are largest in under-resourced courts—consistent with improvements being most pronounced in low-GDP, non-coastal, and western provinces.

The model's key outcome, Q*, is not directly observed. We proxy judicial decision quality with two observable measures: the appeal rate (the fraction of first-instance decisions that litigants choose to appeal) and the reversal rate (the fraction that appellate courts overturn). Higher Q* implies decisions that are more legally consistent, which reduces reversals, and more acceptable to litigants, which reduces appeals. The three hypotheses above therefore predict that AI adoption reduces both the appeal rate and the reversal rate, with the effects being stronger for more complex cases and for courts in less developed, lower-capacity regions.

4. Data and Measurement

4.1 Measuring Court AI Adoption

China's integration into the global trading system has created an institutional framework that enables systematic tracking of government technology adoption. Following China's accession to the World Trade Organization in 2001 and its subsequent designation as an observer to the Government Procurement Agreement (GPA), the nation initiated comprehensive reforms to liberalize its public procurement market. To comply with GPA transparency requirements, China enacted the Government Procurement Law of the People's Republic of China (GPL) in 2002, which became effective in January 2003.

Recognizing the need for more detailed procedural regulation, China substantially amended its procurement legislation in 2014, introducing comprehensive implementing rules governing procurement procedures and information disclosure requirements. Under these enhanced regulations, all government entities—including people's courts at every jurisdictional level—became mandated to publish procurement notices, bid awards, and contract information for all public procurement activities on the China Government Procurement Website (https://www.ccgp.gov.cn/). This centralized disclosure system, implemented uniformly across the judicial hierarchy beginning in 2014, provides a comprehensive and systematic data source for identifying AI-related procurement activities by individual courts, thereby enabling precise measurement of technology adoption patterns across China's judicial system.

We extract information on 2,257,414 procurement contracts issued by all levels of the Chinese government between 2014 and 2022 from the Chinese Government

Procurement Database. To identify AI technology adoption by Chinese courts, we extract comprehensive information from 51,821 procurement contracts issued by the Intermediate Level People's Court (ILPC) and the Basic Level People's Court (BLPC). The database contains detailed contract information including purchaser identification, supplier details, contract designation, comprehensive subject matter specifications, contract execution dates, and publication timestamps. Of particular analytical value is the subject matter information, which provides granular details on contract quantities, unit pricing, total contract values, and specific procurement objectives—data that enables systematic identification of AI-related procurement activities.

We employ OpenAI's latest large language model, ChatGPT-4.1, to screen these procurement contracts and select those related to AI technology. The detailed prompt used for task is provided in Table A1 of the Online Appendix. Figure A3 in the Online Appendix provides an example of such AI-related procurement contract. Following this approach, we identify 222 AI procurement contracts for our empirical analysis. Figure 2 plots the number and ratio of courts that adopt AI technology between 2014 to 2022. By 2022, approximately 120 courts—about 8% of the sample—had adopted AI.

As discussed in Section 2, AI adoption in courts is primarily driven by top-down policy reforms and rising litigation burdens, making it plausibly exogenous to local economic conditions or courts' pre-existing performance. Consistent with this view, Table A2 in the Online Appendix shows no systematic correlation between AI adoption and measures of courts' performance, size, or local economic status.

To ensure that our analysis captures economically meaningful technology investments, we conduct additional robustness checks by imposing minimum contract-size thresholds. Specifically, we require procurement contracts to exceed 50,000 or 100,000 CNY (with 100,000 CNY approximately equal to USD 14,000) for an AI purchase to qualify as substantial.

4.2. Measuring Court Performance

The scarcity of observational data has traditionally posed significant challenges for studying government organizational performance, yet China's extensive publicly available judicial data creates unprecedented opportunities for measuring court performance empirically. Our study utilizes data from over 146 million judgment

documents obtained from China Judgments Online (CJO) spanning 2014 to 2022.[17] These documents contain comprehensive case information including case names, unique case identification codes, party information (plaintiffs and defendants), substantive case details, judicial reasoning (encompassing judges' analyses and applicable legal provisions), document composition dates (typically coinciding with judgment dates), CJO publication timestamps, and complete information on court personnel (judges and clerks).[18]

Second-instance documents provide particularly valuable information for performance measurement, as they routinely reference first-instance case details including unique case codes and trial outcomes. This systematic cross-referencing enables us to track whether first-instance cases were subsequently appealed and determine consistency between first- and second-instance judgment outcomes.

We construct two primary indices to measure judicial quality and performance. The first metric is the appeal rate, calculated as the proportion of first-instance cases that are appealed to higher courts relative to the total number of first-instance cases published on CJO within each court-quarter. The second metric is the reversal rate, measured as the proportion of appealed first-instance judgments that are overturned by second-instance courts relative to the total first-instance caseload at the court-quarter level.[19] These metrics provide direct, quantifiable measures of judicial decision quality and consistency within China's hierarchical court system.

Recognizing that second-instance verdicts can vary in the extent to which they modify first-instance judgments—ranging from complete overturns to partial affirmations with selective revisions—we further decompose the reversal rate into two measures: the strict reversal rate and the non-strict reversal rate. Cases in which the second-instance document includes both reversal and affirmation of the first-instance decision are classified as non-strict reversals, while cases that refer exclusively to reversal are classified as strict reversals. We compute these measures by counting the number of first-instance cases that are strictly or partially reversed and dividing by the

[17] Given the low coverage of published cases prior to 2014 and after 2022, our empirical analysis focuses on the 2014–2022 period.

[18] Certain documents may lack case-specific information when cases are resolved through judicial conciliation or deemed unsuitable for internet publication under applicable legal provisions. These documents are collectively classified as unpublished materials in our analysis.

[19] Given that Chinese law mandates appeals of first-instance verdicts must be filed within 10-15 days of judgment, we approximate first-instance judgment dates using second-instance acceptance dates for cases where precise timing data are unavailable, when we calculate judgment delay.

total number of first-instance cases within each court–quarter, yielding the strict reversal rate and non-strict reversal rate, respectively.

Beyond these primary quality measures, we construct an indicator of judicial efficiency using the time from case acceptance to judgment. We also compute the number of judges and clerks in each court by extracting all judge and clerk names appearing in the judicial documents and counting the number of distinct names within each court–quarter. Detailed definitions of all outcome variables are presented in Table A3 of the Online Appendix.

4.3 Summary Statistics

We merge court performance data with procurement information based on unique court identifiers at the quarterly level, yielding a matched dataset covering 1,474 courts from 2014 to 2022. Figure A4 in the Online Appendix shows the annual number of cases handled by these matched courts. The primary reason for incomplete coverage is that certain provincial and county-level jurisdictions did not rigorously comply with procurement information disclosure requirements. To ensure that unmatched courts do not introduce selection bias, we conduct balance tests comparing matched and unmatched courts across key characteristics—appeal rates, reversal rates, GDP per capita, population, and public expenditure per capita. As presented in Table A4 in the Online Appendix, we find no statistically significant differences between the two groups on these dimensions, except that matched courts exhibit slightly higher GDP per capita and public expenditure per capita.

Table 1 presents descriptive statistics for our primary dependent variables, independent variables, and control variables. Each observation represents a court-quarter from 2014 to 2022. The summary statistics reveal that appeals occur in approximately 7.6% of first-instance cases in our sample. 1.4% of first-instance judicial decisions are modified by second-instance courts after appeals, with 0.84% experiencing strict reversal (complete overturning) and 0.55% experiencing non-strict reversal (selective modification). Administrative efficiency varies considerably, with an average duration of 84 days between receiving the case and make a judicial decision. Regarding AI adoption, approximately 1,218 (2.3%) of court-quarters, which is consisted by 120 courts (about 8.1% of 1,474 courts), in our sample exhibit evidence of AI technology procurement.

4.4 Empirical Specifications

We utilize a staggered difference-in-difference approach to identify the causal impact of court AI technology adoption on their performance. More precisely, we estimate:

$$y_{i,t} = \beta_0 + \beta_1 \cdot \boldsymbol{AI}_{i,t} + X_{it} \cdot \Gamma + \delta_i + \theta_t + \varepsilon_{i,t}. \quad (1)$$

where the dependent variables $y_{i,t}$ are the performance measurement of the court i in year-quarterly t, including appeal rate and reversal rate (both strict and non-strict). $\boldsymbol{AI}_{i,t}$ is a dummy indicator which takes 1 if the court i has made the first AI procurement in year-quarterly t (post-shock) and 0 otherwise (pre-shock). X_{it} denotes area-level control variables: population, GDP per capita, and public expenditure per capita. δ_i is the court fixed effect which controls the time-invariant characteristics of each court such as the court level and court location, while θ_t is the year-quarterly fixed effect which helps to account for the impacts of nationwide policy intervention and shocks. Standard errors are clustered at the court level.

Digital technology is a broader concept that encompasses AI technologies but extends well beyond them. Courts may adopt various non-AI digital technologies, such as biometric access systems, online fee payment portals, e-filing systems, and live broadcasting platforms. To examine whether AI generates additional performance gains beyond those attributable to digital technology more broadly, we construct a measure of general digital procurement following the same procedure used to identify AI adoption, and control for it in our baseline specification. This approach allows us to isolate the incremental effect of AI adoption over and above other forms of digital technology.

In order to exclude the identification of this study from being disturbed by other events such as judicial reforms prior to the AI adoption of the court, this study examines the parallel trends assumption using an event-study approach as follows:

$$y_{i,t} = \sum_{j=-6, j\neq -1}^{8} \beta_j \boldsymbol{AI}_{i,t+j} + X_{it} \cdot \Gamma + \delta_i + \theta_t + \varepsilon_{i,t}, \quad (2)$$

where $\boldsymbol{AI}_{i,t+j}$ is a dummy indicator which takes 1 if j quarters have passed since (before) the court i has made the first AI procurement and 0 otherwise.

Recent research in econometrics theory points out two-way fixed effects (TWFE) regressions deliver consistent estimates only with strong assumptions about the

homogeneity of treatment effects and may be biased when treatment effects vary over time or by adoption cohort (De Chaisemartin and D'Haultfœuille, 2020; Sun and Abraham, 2021; Callaway and Sant'Anna, 2021; Borusyak, Jaravel, and Spiess 2024). For example, courts may take time to adjust to using the AI technology, in which case its impact in the first quarter may be smaller. Alternatively, courts that adopt AI later may get access to more advanced technology, so that their treatment effects may be larger. To address this concern and show the robustness of our results, we estimate the impacts using robust estimators introduced in De Chaisemartin and D'Haultfœuille (2020), Borusyak, Jaravel, and Spiess (2024), Callaway and Sant'Anna (2021), and Sun and Abraham (2021).

5. Main Results

5.1 The Impact of AI Procurement on the Quality of Judicial Decisions

Table 2 presents the baseline estimates of how AI procurement affects the quality of courts' judicial decisions using a standard two-way fixed effects (TWFE) model. Columns (1) and (2) examine the effects on the appeal rate while controlling for court and year-quarter fixed effects, with column (2) further adding controls for population, GDP per capita, and public expenditure per capita. Columns (3) and (4) analyze the corresponding effects on the reversal rate using the same specifications.

Across all specifications, the results are highly consistent: judicial AI adoption significantly reduces both the appeal rate and the reversal rate. The magnitudes are economically meaningful. The column (2) estimate implies that AI adoption reduces the appeal rate by 2.2 percentage points, equivalent to approximately 28.9% of the sample average (7.6%). The effect on reversal outcomes is even larger: the column (4) estimate suggests that AI procurement decreases the reversal rate by 0.559 percentage points, representing about 39.6% of the mean reversal rate (1.411%). This disproportionately larger decline in reversals suggests that the improvement is not merely a mechanical byproduct of fewer appeals, but also reflects enhanced quality in first-instance judgments. Table A5 in the Online Appendix further decomposes the reversal rate into strictly reversed and non-strictly reversed cases, showing that AI adoption substantially reduces both.

To examine the validity of the parallel trends assumption and explore the dynamic pattern of treatment effects, we estimate an event-study version of the TWFE model with leads and lags relative to each court's AI procurement date, as specified in

Equation (2). To account for potential heterogeneity in treatment effects across time and treated units, we additionally report event-study estimates based on several recently proposed heterogeneous-effects robust estimators (De Chaisemartin and d'Haultfoeuille 2020; Borusyak, Jaravel, and Spiess 2024; Callaway and Sant'Anna 2021; Sun and Abraham 2021).

Figure 3(a) plots the event-study estimates for the appeal rate. Across all estimators, the coefficients in the pre-treatment periods are close to zero and show no discernible pre-trends, lending support to the parallel trends assumption. The figure also illustrates the dynamic effects of AI adoption: all four robust estimators indicate a sizable and statistically significant decline in the appeal rate beginning one quarter after adoption, with the effects stabilizing thereafter. Similarly, Figure 3(b) repeats the analysis using the reversal rate as an alternative measure of judicial decision quality. The results exhibit a similar pattern. The impact of AI procurement emerges about one quarter after adoption, with modest increases in subsequent periods.

5.2 Robustness Checks

One potential concern in our empirical strategy is that other judicial reforms implemented during the sample period may also influence court performance. If AI adoption coincides with these initiatives, our baseline results could be confounded. To mitigate this risk, we construct indicators for the timing of several key reforms and incorporate them into our baseline specifications.

We first consider the judge quota reform, which introduced a performance-based evaluation system characterized by clear and measurable assessment criteria. Under this system, judges who excel in the assessment become "quota judges," receiving increased salaries and higher promotion rates, while those who fail to meet the criteria may face reassignment to less prominent supporting roles or even dismissal. Using the provincial rollout timing provided by Chen et al. (2024), we construct the variable "*Judge quota reform*" to represent whether a court was affected by this initiative and include it in our baseline regression. Another important reform during this period is the judicial independence reform, which sought to strengthen judicial independence by reducing the influence of local governments on local courts (Liu et al., 2024; Lei & Li, 2025). This reform transferred financial and personnel authority from prefectural and county governments to provincial-level authorities. Prior research provides evidence that this reform improved judicial decision quality (Liu et al., 2024).

Lastly, "Smart courts," proposed by the Supreme People's Court (SPC) in January 2016, aim to integrate digital technologies—including electronic case filing, online litigation platforms, and data-driven decision-support systems—to enhance transparency, efficiency, and consistency in judicial processes. Since then, China has made rapid progress in implementing this agenda, and many courts' AI procurements have been undertaken explicitly to advance "smart court" construction. At the same time, a key dimension of the reform focuses on improving judicial convenience and reducing the economic costs borne by plaintiffs and defendants—for example, through online case handling, online enforcement, and online petitioning. To empirically assess whether AI procurement is the main driver of the improvement in judicial decision quality within this broader reform, we construct a measure from courts' procurement records indicating whether a court has undertaken any procurement explicitly related to "smart court" construction.

Table 3 presents the results. Columns (1) and (5) add the indicator "*Judge quota reform*"; columns (2) and (6) control for judicial independence reform; columns (3) and (7) control for the procurement related to "smart court"; while columns (4) and (8) incorporate all concurrent reforms simultaneously. After controlling for these reforms, AI procurement continues to exhibit a statistically significant effect on both the appeal and reversal rates. Although the estimated coefficients for the reversal rate are slightly smaller than those in the baseline specification, the effects remain sizable, indicating that our findings are not driven by these concurrent reforms.

Moreover, consistent with Chen et al. (2024), the results show that the judge quota reform slightly improves the quality of judicial decisions, though the impact becomes noisy when all reforms are included due to high correlation between the judge quota and judicial independence reforms. Similarly, we find that the judicial independence reform itself leads to reductions in both appeal and reversal rates, echoing the findings in Liu et al. (2024). In addition, the estimated effects of "smart court" procurement itself are negative and statistically significant. These findings suggest that "smart court" reform plays a positive role in improving the quality of judicial decisions.

To ensure that the main results are not driven by idiosyncratic shocks or chance correlations in the staggered treatment timing, we conduct a series of permutation tests (Placebo Assignments). Following the logic of Randomization Inference, we randomly reassign the treatment timing across court units, preserving the original distribution of

treatment dates while decoupling them from the actual court-level characteristics. This process is repeated 3,000 times to construct a null distribution of placebo coefficients. Figure A6 plots the resulting placebo distributions together with the true estimates. For both appeal and reversal rates, the true estimates lie in the extreme tails of the placebo distributions with two-sided empirical p-values below 0.01. These results confirm that the findings are not driven by chance or spurious correlations inherent in the staggered timing.

As noted earlier, the annual number of judicial decisions published on the China Judgments Online (CJO) platform increased steadily and remained at a high level through 2021, but then dropped sharply—from 15.02 million in 2021 to 9.30 million in 2022. This decline has been attributed to growing institutional concerns, such as the protection of individual rights and information security risks. The substantial reduction in published decisions may introduce measurement noise into our outcome variables, the appeal rate and the reversal rate, in 2022.

To address these concerns, we exclude the year 2022 from the sample. Table A6 in the Online Appendix reports the results. The estimated coefficients closely mirror those in the baseline specification, suggesting that our findings are not driven by potential measurement issues arising from the sharp decline in published decisions in 2022.

Furthermore, to address the concern that our results may be driven by the use of a particular LLM to identify AI adoption, Table A7 in the Online Appendix shows that our results are robust to alternative LLM-based measures. Specifically, we define AI adoption using DeepSeek outputs and the union of GPT and DeepSeek outputs. The results are similar to our baseline results.

A potential concern is that upper-level courts' AI adoption may be correlated with that of the focal court and may independently affect our outcome variables, particularly the reversal rate. To address this, Table A8 in the Online Appendix controls for upper-level courts' AI adoption, and the results remain robust.

Finally, to account for court size, Table A9 in the Online Appendix shows that our results are robust to weighting by court size, measured by the number of cases and the number of judges, respectively. In Table A10 of the Online Appendix we further show that the results are robust to controlling for these two variables.

5.3 Instrumenting Courts' AI Adoption

AI procurement decisions are not randomly made, but the top-down reform initiative—coupled with rising litigation pressures on judges—provides a plausibly exogenous source of variation in courts' adoption of AI technologies. While our pre-trend analyses and robustness tests support the validity of our DID design, a remaining concern is that the timing of AI adoption may be correlated with other institutional changes—for example, reductions in the number of judges or the appointment of a new court president who simultaneously implements additional reforms aimed at improving judicial quality. If AI adoption occurs alongside staff reductions, our baseline estimates may understate the true effect of AI.[20]

To address this concern, we construct an IV by exploiting AI adoption by other government departments. Specifically, we use the number of AI procurements by city-level environmental protection bureaus and tax bureaus.

This instrument satisfies the two key conditions for a valid IV. First, the relevance condition is likely to hold because courts within the same city face similar budgetary constraints, policy directives, and technology-promotion efforts from higher-level authorities. As a result, AI adoption by city-level government departments is strongly correlated with county-level courts' AI adoption decisions. This is confirmed empirically by highly significant first-stage estimates, with Kleibergen--Paap F-statistics above conventional thresholds.

Second, the exclusion restriction requires that AI adoption by environmental protection bureaus and tax bureaus affects local courts' judicial outcomes only through its effect on courts' own AI adoption. AI adoption by these city-level government departments is unlikely to be influenced by contemporaneous shocks specific to county-level courts, such as leadership changes, judge turnover, or shifts in case composition. Instead, city-level adoption patterns are likely driven primarily by broader budget allocations and technology-promotion efforts—factors that plausibly affect the focal court's outcomes only through whether the court itself procures AI. Together, these features support the validity of the instrument in isolating exogenous variation in courts' AI adoption.

Table 4 reports the IV estimates. Columns (1) and (2) present the results for the appeal rate, and columns (3) and (4) for the reversal rate, with the latter columns

[20] In Table 7, we provide evidence that AI adoption substitute court staff.

including controls. The IV estimates closely align with the baseline in terms of sign and significance but are larger in magnitude for both outcome variables. This pattern suggests that the baseline TWFE estimates may underestimate the true impact of AI adoption on judicial decision quality.

5.4 Comparison with Digital Procurement

Prior to the emergence of modern AI technologies, many courts had already undertaken various forms of digital transformation.[21] While digitalization is a broad concept encompassing a wide range of information technologies, AI represents a more advanced and powerful subset that is expected to have a distinct impact on institutional performance. An empirical question, therefore, is whether AI adoption improves judicial decision quality above and beyond the effects of traditional digital technologies. To isolate the incremental effect of AI, we include a difference-in-differences term capturing digital procurement in our specifications.

Table 5 reports the results. Columns (1) and (2) show that even after controlling for digital procurement, AI procurement continues to exert an additional and statistically significant effect in reducing the appeal rate. We also observe an effect of digital procurement, although it is smaller than the effect of AI procurement. Columns (3) and (4) present the corresponding results for the reversal rate and reveal a similar pattern: AI procurement significantly reduces reversal rates, whereas digital procurement has a more modest effect. Taken together, these findings indicate that improvements in the quality of judicial decisions are driven specifically by AI technologies rather than by general digitalization efforts.

5.5 Other Outcomes: Efficiency and Discretion

A large body of prior research documents AI's comparative advantage in automating mechanical, repetitive, and routine tasks (Demirci, Hannane, & Zhu 2025; Autor 2024; Acemoglu 2025). Anecdotal evidence from court practice aligns with this view and suggests that AI technologies can substantially enhance judicial efficiency, for example by automatically extracting key information from case files, converting courtroom audio into searchable text, preparing materials for online case publication, and streamlining routine procedural work. These capabilities reduce judges'

[21] Figure A5 in the Online Appendix plots the trend in digital technology adoption among courts.

administrative burdens and accelerate case processing. Thus, beyond improving decision quality, AI adoption may also increase courts' operational efficiency.

To evaluate this possibility, we construct a measure of court efficiency, judge delay, defined as the duration between a court's receipt of a case and the issuance of a judicial decision. In addition, we examine another outcome variable, judicial discretion. Following Liu et al. (2024), we define a case as involving judicial discretion if the presiding judge cites "discretionary codes" in their reasoning, and we then calculate the share of such cases at the court-quarter level. A reduction in discretion may improve efficiency by standardizing judicial decisions, reducing uncertainty, and lowering the costs associated with resolving similar cases.

Using the same specifications as in our baseline analysis, Table 6 reports the results. Columns (1) and (2) examine judge delay, while columns (3) and (4) focus on judicial discretion. The findings in the first two columns show that AI procurement accelerates case processing. The magnitudes are economically meaningful: the estimate in column (2) indicates that AI procurement shortens the duration from case receipt to judgment by approximately 11.1%. This effect is consistent with anecdotal evidence suggesting that AI plays a particularly important role in routine tasks. In columns (3) and (4), we find that AI adoption also slightly reduces judicial discretion, although the effect is statistically insignificant.

5.6 Labor-Substituting Effect

It is well established in the literature that AI can substitute for human labor across a wide range of tasks (e.g., Acemoglu and Restrepo, 2019; Acemoglu et al., 2022). Motivated by this insight, we examine whether AI similarly operates as a substitute for judges and court clerks in China. Because AI systems can perform routine and labor-intensive components of judicial work, their adoption may reduce the overall demand for judicial personnel. For court clerks, who are responsible for procedural, administrative, and record-keeping tasks that support court operations, a reduction following AI adoption is intuitive, as many of their tasks can be readily automated. For judges, although their positions in China have traditionally been highly stable, the Judge Quota Reform introduced in 2014 dismantled the traditional "iron rice bowl," effectively ending the system of de facto lifetime tenure for judges (Chen et al., 2024). In addition, AI adoption may influence new judicial hiring decisions, thereby affecting the number of judges employed by a court.

To investigate this possibility, we extract the judge and clerk names from each judicial document and consider different names represent different persons within the same court-quarter. We then calculate the number of judges and the number of court clerks at the court-quarter level, which are used as dependent variables to examine whether AI adoption leads to a decline in judicial staffing.[22] As shown in Table 7, we find negative and statistically significant effects for both judges and court clerks, indicating that AI adoption reduces staffing levels, consistent with a labor-substitution mechanism.

6. Heterogeneity Analyses

6.1 Different Types of AI Technology

In our baseline analysis, we consider broader AI adoption by courts. However, AI technologies can be further classified into trial-specific and routine-specific applications, depending on whether they primarily assist adjudication-related tasks or support routine administrative and operational tasks. This distinction is important because different types of AI technologies may affect court performance through different channels. Trial-specific AI technologies, such as tools for case-law retrieval and evidence review, are more directly related to judicial decision-making and therefore may have stronger effects on appeal and reversal rates. By contrast, routine-specific AI technologies may have a more limited direct effect on the quality of judicial decisions.

Table 8 reports the results. We examine the effects of trial-specific and routine-specific AI adoption separately, and then include both measures in the same specification. Trial-specific AI adoption consistently reduces both appeal rates and reversal rates, with the estimates remaining stable across specifications. Routine-specific AI adoption, by contrast, yields no statistically significant effects on either appeal or reversal rates. Interestingly, routine-specific AI adoption does reduce judgment delay significantly, by roughly 13.9%, suggesting that it primarily improves court efficiency in processing cases rather than the quality of judicial decisions. These findings indicate that the performance gains from AI adoption are mainly driven by

[22] Several caveats apply to this measure. First, judges or staff who did not participate in any cases published on the CJO platform are unobservable to us and therefore excluded from our count. Second, when the share of published cases is low, our measures may be subject to greater measurement error. Third, in the rare event that two judges or staff members share the same name within the same court and quarter, we would mistakenly treat them as a single individual.

technologies that directly support adjudication-related tasks rather than by those used primarily for routine administrative functions.

6.2 AI Readiness

The effect of AI adoption on court performance may also depend on courts' level of AI readiness. AI is not adopted in isolation; its effective use may require complementary digital infrastructure, prior experience with digital systems, and organizational capacity to integrate new technologies into existing workflows. Courts that have already adopted other digital technologies may therefore be better prepared to use AI tools effectively. Motivated by this consideration, we classify courts into high- and low-AI-readiness groups based on whether they had procured other digital technologies before their first AI procurement. Courts with prior procurement of other digital technologies are classified as high-readiness courts, while those without such prior procurement are classified as low-readiness courts.

Table 9 reports the results. Columns (1) and (3) show the effects of AI adoption for courts with low AI readiness, while columns (2) and (4) show the corresponding effects for courts with high AI readiness. The estimates indicate that AI adoption significantly reduces the appeal rate by 2.3 percentage points and the reversal rate by 0.687 percentage points in high-readiness courts. By contrast, in low-readiness courts, AI adoption reduces the appeal rate by 1.8 percentage points and the reversal rate by 0.361 percentage points, but these effects are statistically insignificant. These findings suggest that the performance gains from AI adoption are larger and more precisely estimated when courts have already accumulated complementary digital capabilities.

6.3 Heterogeneity Analyses by Monetary Value of AI Procurement

AI procurements involving different monetary values may play different roles in influencing court performance. All else equal, procurements with higher monetary values may reflect more advanced AI technologies or higher-quality equipment, and thus are more likely to generate meaningful performance improvements. In contrast, lower-value procurements may correspond to less sophisticated or limited AI tools and therefore may have weaker effects. Motivated by this consideration, we conduct a heterogeneity analysis by dividing courts into low-, medium-, and high-intensity groups based on terciles of their first AI procurement amounts.

Table 10 reports the results. As shown, AI adoption associated with high procurement amounts significantly reduces both the appeal rate and the reversal rate. By contrast, AI adoption associated with low or medium procurement amounts does not yield statistically significant effects, although the coefficients remain negative. These findings suggest that more expensive AI procurements, likely reflecting higher-quality or more capable technologies, lead to greater improvements in court performance.

6.4 Heterogeneity Analyses by Region

Our theoretical model predicts that when AI and human effort are substitutes, verified in our empirical results, AI adoption yields larger marginal quality gains in courts with lower human (resource) capacity. To examine this heterogeneity, we classify courts according to the development level of the regions in which they are located, using two commonly employed proxies. The first proxy is GDP per capita. The second distinguishes courts in western versus non-western cities, as western regions in China are, on average, less developed.

The results are presented in Table 11. As shown in Panel A of Table 11, AI adoption significantly reduces both the appeal rate and the reversal rate for courts located in below-median GDP per capita regions, whereas the coefficient for the appeal rate is statistically insignificant for courts in above-median regions. Specifically, the coefficients for the appeal rate are -0.028 versus -0.014 (below-median versus above-median), and those for the reversal rate are -0.674 versus -0.419. These differences suggest that the effects are larger in less developed regions. Panel B of Table 11 further shows that court performance improves more markedly following AI adoption in western cities, which on average are less developed.

Taken together, these findings support the expectation that AI adoption yields greater performance gains in less developed regions, where courts typically operate under lower capacity and weaker technological and institutional conditions.

6.5 Heterogeneity across Case Types

AI adoption may also generate heterogeneous effects across different types of cases. We consider two definitions of case type. First, following Chen et al. (2025), we distinguish between complex and non-complex cases by classifying cases with above-median description length as complex, since more complex cases generally require

longer and more detailed explanations. Our theoretical model shows that complex cases span a wider task continuum, AI augmentation compounds across more tasks, so the same improvement in AI productivity yields a larger total quality gain for complex cases than for simple. Second, we classify cases according to the categories defined in China's legal system–civil, criminal, and administrative cases. Given their distinct procedural requirements and informational demands, these different types of cases may be differentially affected by AI technology.

The heterogeneous effects for complex versus non-complex cases are reported in Table 12. We find that AI adoption reduces the appeal rate by 0.025 for complex cases, compared with 0.017 for non-complex cases. Similarly, AI reduces the reversal rate by 0.722 for complex cases and by 0.321 for non-complex cases. These patterns indicate that AI generates larger improvements for complex cases, consistent with our model's prediction. In other words, AI plays a particularly important role in enhancing courts' ability to manage more intricate and information-intensive cases.

The results for civil, criminal, and administrative cases are presented in Table 13. As shown, AI adoption significantly improves court performance across all three types of cases, although the magnitudes differ. This pattern may reflect differences in procedural rigidity, evidentiary standards, and judicial discretion across case types, which influence the extent to which AI tools can effectively support adjudication.

Table A11 in the Online Appendix further explores whether the effects of AI adoption differ between firm-involved (i.e., commercial) cases and non-firm cases. The results show that AI adoption reduces appeal rates in both subsamples. For reversal rates, the reduction is statistically significant for non-firm cases but not for firm-involved cases. Overall, the beneficial effects of AI adoption extend across both types of cases, though the evidence is somewhat stronger for non-firm cases. One possible explanation is that commercial cases tend to be more complex and contentious, involving higher stakes and more sophisticated legal arguments, which may limit the ability of AI tools to improve decision consistency in ways that withstand appellate scrutiny.

7. Conclusion

This study examines the role of AI technology in shaping judicial performance in China. Using a court-level panel dataset constructed from public procurement records and comprehensive judgment documents from the CJO database, we develop new

measures of AI adoption and court performance. Our results show that AI technologies substantially improve judicial performance, enhancing both decision quality and case-processing efficiency. These findings are robust to the inclusion of additional controls and to alternative identification strategies.

We also assess the incremental effects of AI relative to general digital technologies, which have received considerable attention in prior research. The evidence indicates that AI delivers significantly greater improvements in court performance than traditional digital tools, highlighting its distinctive value in judicial settings. Furthermore, through a series of heterogeneity analyses by region and by case type, we document that the effects of AI are more pronounced in less developed regions and for more complex cases, underscoring its particular importance in resource-constrained environments where institutional capacity is limited.

By providing direct micro-level evidence from courts, this study is among the first to address a central and increasingly urgent question in economics: how does AI reshape outcomes in public sectors? Our findings also open a broader research agenda concerning how AI-assisted judicial systems influence economic activity and, more generally, how modern governments will adapt to and be transformed by this emerging technology. Although our analysis focuses on courtroom applications, it lays important groundwork for future research on these deeper and more far-reaching questions.

REFERENCES

Acemoglu, D. (2021). Harms of AI. NBER Working Paper No. 29247.

Acemoglu, D. (2025). The simple macroeconomics of AI. Economic Policy, 40(121), 13–58.

Acemoglu, D., & Restrepo, P. (2018). Artificial intelligence, automation, and work. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), The Economics of Artificial Intelligence: An Agenda (pp. 197–236). University of Chicago Press.

Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. Journal of Economic Perspectives, 33(2), 3–30.

Acemoglu, D., & Restrepo, P. (2022). Tasks, automation, and the rise in US wage inequality. Econometrica, 90(5), 1973–2016.

Acemoglu, D., Autor, D., Hazell, J., et al. (2022). Artificial intelligence and jobs: Evidence from online vacancies. Journal of Labor Economics, 40(S1), S293–S340.

Agarwal, N., Moehring, A., Rajpurkar, P., & Salz, T. (2023). Combining human expertise with artificial intelligence: Experimental evidence from radiology. NBER Working Paper No. 31422.

Agrawal, A., Gans, J., & Goldfarb, A. (2018). Prediction, judgment, and complexity: A theory of decision-making and artificial intelligence. In The Economics of Artificial Intelligence: An Agenda. University of Chicago Press.

Agrawal, A., Gans, J., & Goldfarb, A. (Eds.). (2019). The Economics of Artificial Intelligence: An Agenda. University of Chicago Press.

Aghion, P., Jones, B. F., & Jones, C. I. (2017). Artificial intelligence and economic growth. NBER Working Paper No. 23928.

Angelova, V., Dobbie, W. S., & Yang, C. (2023). Algorithmic recommendations and human discretion. NBER Working Paper No. 31747.

Ash, E., Chen, D. L., & Naidu, S. (2026). Ideas have consequences: The impact of law and economics on american justice. The Quarterly Journal of Economics, 141(1), 845-887.

Ash, E., Chen, D. L., & Ornaghi, A. (2024). Gender attitudes in the judiciary: Evidence from US circuit courts. American Economic Journal: Applied Economics, 16(1), 314–350.

Assad, S., Clark, R., Ershov, D., et al. (2024). Algorithmic pricing and competition: Empirical evidence from the German retail gasoline market. Journal of Political Economy, 132(3), 723–771.

Athey, S. (2018). The impact of machine learning on economics. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), The Economics of Artificial Intelligence: An Agenda (pp. 507–547). University of Chicago Press.

Athey, S., & Stern, S. (2000). The impact of information technology on emergency health care outcomes. Unpublished manuscript.

Autor, D. (2024). Applying AI to rebuild middle class jobs. NBER Working Paper No. w32140.

Babina, T., Fedyk, A., He, A., & Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. Journal of Financial Economics, 151, 103745.

Balasubramanian, N., Lee, J., & Sivadasan, J. (2018). Deadlines, workflows, task sorting, and work quality. Management Science, 64(4), 1804-1824.

Baye, M. R., & Wright, J. D. (2011). Is antitrust too complicated for generalist judges? The impact of economic complexity and judicial training on appeals. The Journal of Law and Economics, 54(1), 1–24.

Beck, T. (2012). Legal institutions and economic development. In D. C. North, B. R. Weingast, & J. Wallis (Eds.), The Oxford Handbook of Capitalism (pp. 38–77). Oxford University Press.

Beck, T., & Levine, R. (2005). Legal institutions and financial development. In C. Ménard & M. M. Shirley (Eds.), Handbook of New Institutional Economics (pp. 251–278). Springer.

Beraja, M., Kao, A., Yang, D. Y., & Yuchtman, N. (2023). AI-tocracy. The Quarterly Journal of Economics, 138(3), 1349–1402.

Besley, T., & Persson, T. (2009). The origins of state capacity: Property rights, taxation, and politics. American Economic Review, 99(4), 1218–1244.

Bhuller, M., & Sigstad, H. (2025). Feedback and learning: The causal effects of reversals on judicial decision-making. Review of Economic Studies, 92(4), 2359-2397.

Boyacı, T., Canyakmaz, C., & de Véricourt, F. (2024). Human and machine: The impact of machine input on decision making under cognitive limitations. Management Science, 70(2), 1258–1275.

Borusyak, K., Jaravel, X., & Spiess, J. (2024). Revisiting event-study designs: Robust and efficient estimation. Review of Economic Studies, 91(6), 3253–3285.

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. The Quarterly Journal of Economics, 140(2), 889–942.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. Journal of econometrics, 225(2), 200-230.

Cao, G., Ding, H., & Zhou, L. A. (2024). Technology and legal capacity of the state: Evidence from online judicial auctions in China. Unpublished manuscript.

Caplin, A., Deming, D. J., Li, S., et al. (2024). The ABC's of who benefits from working with AI: Ability, beliefs, and calibration. NBER Working Paper No. w33021.

Caplin, A., Gomberg, A., & Sadka, J. (2024). Judging the Judges: Indexing of Complex Information Reduces Injustice (No. w32587). National Bureau of Economic Research.

Capraro, V., Lentsch, A., Acemoglu, D., et al. (2024). The impact of generative artificial intelligence on socioeconomic inequalities and policymaking. PNAS Nexus, 3(6), pgae191.

Castro, R., Clementi, G. L., & MacDonald, G. (2009). Legal institutions, sectoral heterogeneity, and economic development. The Review of Economic Studies, 76(2), 529–571.

Chemin, M. (2009). Do judiciaries matter for development? Evidence from India. Journal of Comparative Economics, 37(2), 230–250.

Chemin, M. (2020). Judicial efficiency and firm productivity: Evidence from a world database of judicial reforms. Review of Economics and Statistics, 102(1), 49–64.

Chen, H., Chen, Y., & Yang, Q. (2025). Women in the courtroom: Technology and justice. Review of Economic Studies, rdaf066.

Chen, Y., Song, Y., Wang, H., & Yang, Q. (2024). Setting the Bench Higher: The Impact of Judge Evaluation Reform on Judicial Performance in China. Available at SSRN 4925678.

Chen, Z., Chen, T., Lin, Y., et al. (2025). Building tall, falling short: An empirical assessment of Chinese skyscrapers. Journal of Urban Economics, 145, 103731.

De Chaisemartin, C., & D'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. American Economic Review, 110(9), 2964–2996.

Demirci, O., Hannane, J., & Zhu, X. (2025). Who is AI replacing? The impact of generative AI on online freelancing platforms. Management Science.

Djankov, S., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2003). Courts. The Quarterly Journal of Economics, 118(2), 453–517.

Elbahnasawy, N. G. (2014). E-government, internet adoption, and corruption: an empirical investigation. World Development, 57, 114-126.

Fang, B., Huang, S., Liu, W., & Roychowdhury, S. (2025). Frivolous or meritorious lawsuits? Judicial efficiency and corporate disclosure. Working paper.

Feng, F. (2015). Reform goals and paths of the judge quota reform (in chinese). Contemporary Law Review (in Chinese), 29(5):140–148.

Garicano, L., & Heaton, P. (2010). Information technology, organization, and productivity in the public sector: Evidence from police departments. Journal of Labor Economics, 28(1), 167–201.

Hampole, M., Papanikolaou, D., Schmidt, L. D., et al. (2025). Artificial intelligence and the labor market. NBER Working Paper No. w33509.

Henderson, J. V., Su, D., Zhang, Q., et al. (2022). Political manipulation of urban land markets: Evidence from China. Journal of Public Economics, 214, 104730.

Hoffman, M., Kahn, L. B., & Li, D. (2018). Discretion in hiring. The Quarterly Journal of Economics, 133(2), 765–800.

Javed, K., & Li, J. (2025). Bias in adjudication: Investigating the impact of artificial intelligence, media, financial and legal institutions. PLOS ONE, 20(1), e0315270.

Johnson, J. P., Rhodes, A., & Wildenbeest, M. (2023). Platform design when sellers use pricing algorithms. Econometrica, 91(5), 1841–1879.

Kanazawa, K., Kawaguchi, D., Shigeoka, H., & Watanabe, Y. (2025). AI, skill, and productivity: The case of taxi drivers. Management Science.

Kanyam, D. A., Kostandini, G., & Ferreira, S. (2017). The mobile phone revolution: have mobile phones and the internet reduced corruption in Sub-Saharan Africa?. World Development, 99, 271-284.

Kastellec, J. P. (2013). Racial diversity and judicial influence on appellate courts. American Journal of Political Science, 57(1), 167–183.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018a). Human decisions and machine predictions. The Quarterly Journal of Economics, 133(1), 237–293.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Rambachan, A. (2018b). Algorithmic fairness. AEA Papers and Proceedings, 108, 22–27.

Korinek, A. (2024). Economic policy challenges for the age of AI. NBER Working Paper No. 32980.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. W. (1997). Legal determinants of external finance. The Journal of Finance, 52(3), 1131–1150.

Lei, Z., & Li, Y. (2025). Making local courts work: The judicial recentralization reform and local protectionism in China. The Journal of Politics, 87(4),1573-1586.

Li, B., & Ponticelli, J. (2022). Going bankrupt in China. Review of Finance, 26(3), 449–486.

Liu, E., Lu, Y., Peng, W., & Wang, S. (2024). Court capture, local protectionism, and economic integration: evidence from China. Review of Economics and Statistics, forthcoming.

Mastrobuoni, G. (2020). Crime is terribly revealing: Information technology and police productivity. The Review of Economic Studies, 87(6), 2727–2753.

McCullough, J. S., Parente, S. T., & Town, R. (2016). Health information technology and patient outcomes: The role of information and labor coordination. The RAND Journal of Economics, 47(1), 207–236.

Milakovich, M. E. (2012). Digital governance: New technologies for improving public service and participation. Routledge.

Ng, K. H., & He, X. (2017). Embedded Courts: Judicial Decision-Making in China. Cambridge University Press.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187–192.

Popova, M. (2012). Politicized Justice in Emerging Democracies: Courts in Russia and Ukraine. Cambridge University Press.

Rambachan, A., Kleinberg, J., Ludwig, J., & Mullainathan, S. (2020). An economic perspective on algorithmic fairness. AEA Papers and Proceedings, 110, 91–95.

Shayo, M., & Zussman, A. (2011). Judicial ingroup bias in the shadow of terrorism. The Quarterly Journal of Economics, 126(3), 1447–1484.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of econometrics, 225(2), 175-199.

Tirole, J. (2021). Digital dystopia. American Economic Review, 111(6), 2007–2048.

Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. Nature Human Behaviour, 8(12), 2293–2303.

Wang, Y., Zhang, N., & Zhao, X. (2022). Understanding the determinants in the different government AI adoption stages: Evidence of local government chatbots in China. Social Science Computer Review, 40(2), 534–554.

West, D. M. (2005). Digital government: Technology and public sector performance. Princeton University Press.

Yeung, K., & Lodge, M. (2019). Algorithmic Regulation. Oxford University Press.

Zhang, P. (2023). Anti-corruption campaign, political connections, and court bias: Evidence from Chinese corporate lawsuits. *Journal of Public Economics*, *222*, 104861.

Zhang, Q., Yu, Z., & Kong, D. (2019). The real effect of legal institutions: Environmental courts and firm environmental protection expenditure. Journal of Environmental Economics and Management, 98, 102254.

FIGURES AND TABLES

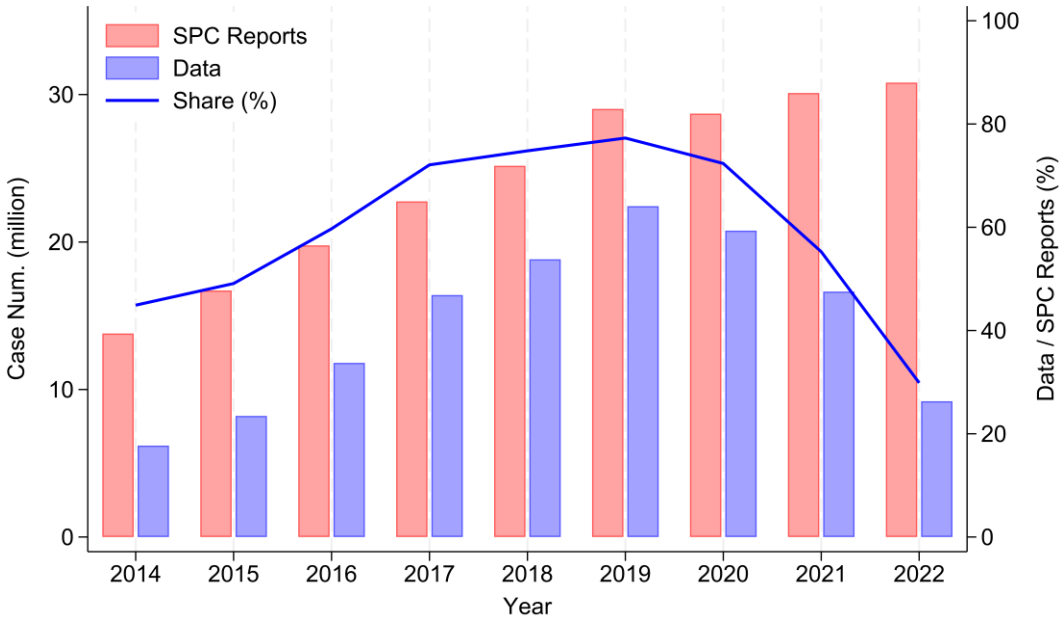


Figure 1. The Trend of the Number of Online Published Cases

This chart presents the evolution of judicial case volumes in China. The red bars show the number of cases concluded, as reported in the Supreme People's Court statistical bulletin, while the blue bars indicate the number of cases published on the CJO platform. The line graph plots the share of published cases, calculated as the number of published cases divided by the total number of cases.

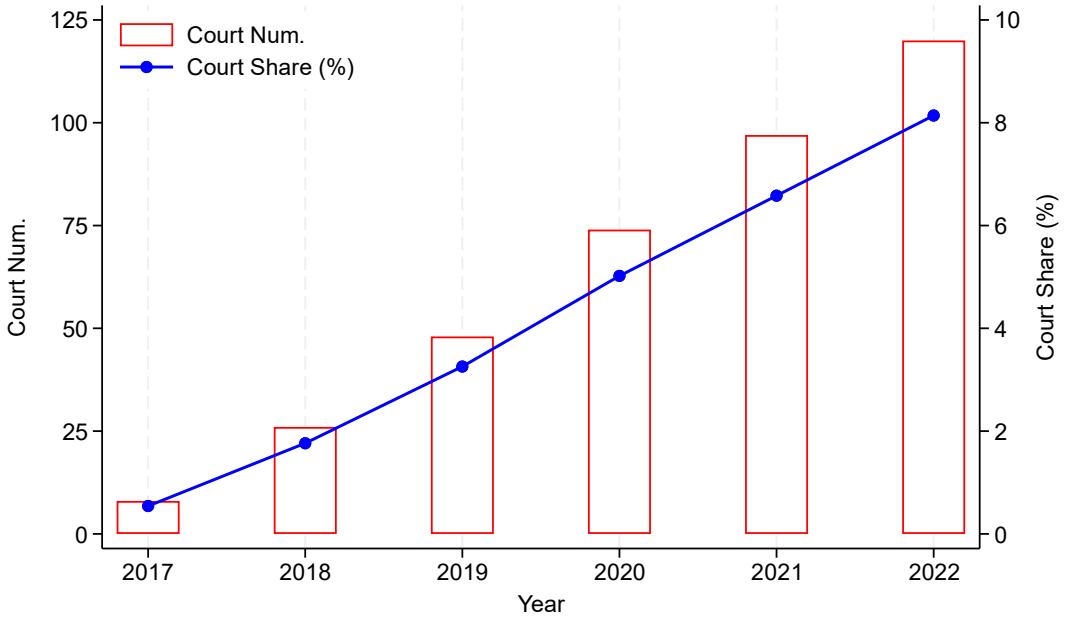


Figure 2. The Trend of AI Adoption

This chart illustrates the temporal trend of Chinese courts acquiring AI technology. The red bar chart represents the cumulative number of courts procuring AI technology, while the blue line chart shows the proportion of such courts relative to the sample courts.

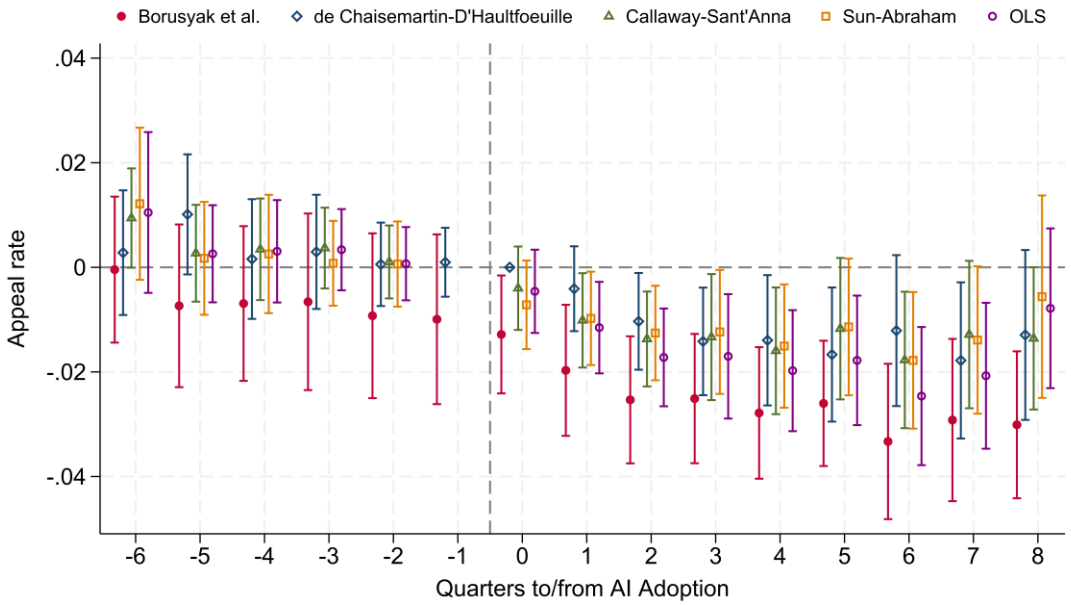


Figure 3(a): Event Study for Appeal Rate

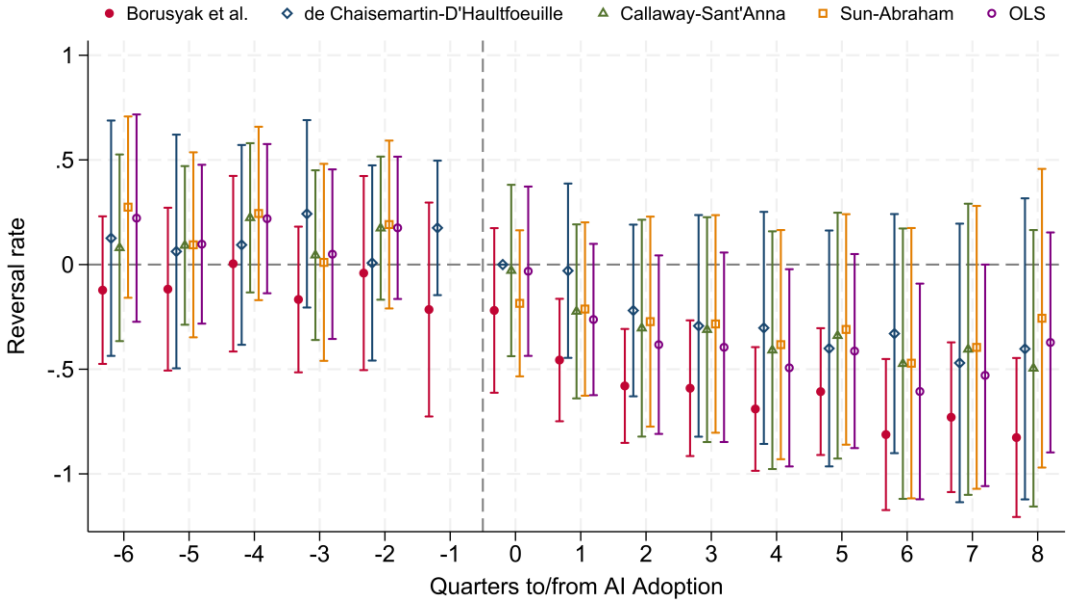


Figure 3(b): Event Study for Reversal Rate

Figure 3. Event Study Results

This figure plots the estimates and their 95% confidence intervals from TWFE-OLS (in purple circle) and four methods addressing negative weights in staggered difference-in-difference, including Borusyak et al. (2024) in red dot, De Chaisemartin and D'Haultfoeuille (2020) in blue diamond, Sun and Abraham (2021) in green triangle, and Callaway and Sant'Anna (2021) in yellow square. The omitted period is quarter −1 for estimators normalized relative to the last pre-treatment period. For Borusyak et al. (2024) and De Chaisemartin and D'Haultfoeuille (2020), the reported pre-treatment coefficients are placebo estimates rather than coefficients normalized by omitting quarter −1.

Table 1. Summary Statistics

This table provides summary statistics for characteristics at both the court level and the administrative region where the court is located. For district-level courts, we report statistics for the district or county where the court is located; for municipal intermediate courts, we report statistics for the city where the court is located. Table A3 in the Online Appendix provides the detailed definition of these characteristics.

	N	Mean	p50	p10	p90	SD
AI procurement (GPT)	44,371	0.016	0.000	0.000	0.000	0.124
AI procurement (DeepSeek)	44,371	0.019	0.000	0.000	0.000	0.137
AI procurement (Technology: Routine)	44,371	0.001	0.000	0.000	0.000	0.036
AI procurement (Technology: Trail)	44,371	0.001	0.000	0.000	0.000	0.025
AI procurement (Amount: Low)	44,371	0.006	0.000	0.000	0.000	0.079
AI procurement (Amount: Median)	44,371	0.006	0.000	0.000	0.000	0.075
AI procurement (Amount: High)	44,371	0.005	0.000	0.000	0.000	0.073
Appeal Rate	44,371	0.076	0.056	0.003	0.160	0.083
Reversal Rate	44,371	1.411	0.731	0.000	3.531	2.078
Strict Reversal Rate	44,371	0.835	0.334	0.000	2.235	1.386
Non-strict Reversal Rate	44,371	0.549	0.199	0.000	1.513	0.917
Judgment Delay (log)	43,666	4.385	4.399	3.813	4.903	0.482
Discretion (log)	44,260	-0.444	-0.343	-0.897	-0.110	0.367
Judges Num. (log)	44,370	3.443	3.466	2.565	4.317	0.726
Judges Num. (incumbent, log)	44,275	3.350	3.367	2.485	4.220	0.736
Judges Num. (newly appointed, log)	33,433	1.096	1.099	0.000	2.303	0.915
Clerks Num. (log)	44,371	3.426	3.434	2.398	4.500	0.877
Clerks Num. (incumbent, log)	44,254	3.212	3.219	2.197	4.234	0.861

Clerks Num. (newly appointed, log)	39,895	1.678	1.609	0.000	3.135	1.114
Case Number	44,371	916.081	447.000	73.000	2,097.000	1,410.630
Digital procurement	44,371	0.226	0.000	0.000	1.000	0.418
Smart court construction	44,371	0.019	0.000	0.000	0.000	0.136
Judicial independence reform	28,502	0.423	0.000	0.000	1.000	0.494
Judge quota reform	44,371	0.840	1.000	0.000	1.000	0.367
GDP Per Capita	44,371	0.463	0.348	0.175	0.833	0.395
Population	44,371	0.498	0.430	0.150	0.940	0.348
Expenditure Per Capita	44,371	0.085	0.072	0.000	0.162	0.077

Table 2. Baseline Results

This table examines the impact of AI technology on court appeal rates and reversal rates, two outcome-based indices. In columns (2) and (4), we further control for regional characteristics including regional GDP per capita, regional population, and regional public fiscal expenditure per capita. All outcome variables are winsorized at the 1% and 99% levels. Standard errors are clustered at the court level, *, ** and *** denote significance at 10%, 5% and 1% levels, respectively.

	Appeal Rate		Reversal Rate (%)	
	(1)	(2)	(3)	(4)
AI procurement	-0.022***	-0.022***	-0.565**	-0.559**
	(0.007)	(0.007)	(0.162)	(0.162)
Controls	NO	YES	NO	YES
Observations	45,204	44,371	45,204	44,371
Adjusted R-squared	0.379	0.381	0.282	0.283
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.0765	0.0764	1.411	1.411

Table 3. Robustness Test: Controlling for Other Contemporaneous Reforms

This table examines whether the estimated impact of AI adoption on court appeal rates and reversal rates is robust to controlling for other contemporaneous judicial reforms. Columns (1) and (5) additionally control for the judge quota reform. Columns (2) and (6) additionally control for the judicial independence reform. Columns (3) and (7) additionally control for smart court construction. Columns (4) and (8) include all three reform controls simultaneously. All specifications control for regional characteristics, including regional GDP per capita, regional population, and regional public fiscal expenditure per capita. All outcome variables are winsorized at the 1% and 99% levels. Standard errors are clustered at the court level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Appeal Rate				Reversal Rate (%)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AI procurement	-0.022***	-0.022***	-0.021***	-0.020***	-0.564***	-0.554***	-0.522***	-0.520***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.162)	(0.162)	(0.160)	(0.160)
Judge quota reform	-0.008*			-0.000	-0.130			-0.041
	(0.005)			(0.005)	(0.109)			(0.116)
Judicial independence reform		-0.016***		-0.016***		-0.215**		-0.190*
		(0.005)		(0.005)		(0.103)		(0.110)
Smart court construction			-0.017**	-0.016**			-0.491***	-0.483***
			(0.008)	(0.008)			(0.175)	(0.175)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	44,371	44,371	44,371	44,371	44,371	44,371	44,371	44,371
Adjusted R-squared	0.381	0.382	0.382	0.383	0.283	0.283	0.283	0.284
Court FE	YES	YES	YES	YES	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES	YES	YES	YES	YES
Mean of outcome	0.0764	0.0764	0.0764	0.0764	1.411	1.411	1.411	1.411

Table 4. IV Results

This table reports the results of the two-stage estimation using the instrumental variable constructed in Equation (3). The results of the first-stage estimation and the Kleibergen-Paap F-statistics are reported at the bottom of the table. In columns (2) and (4), we further control for regional characteristics including regional GDP per capita, regional population, and regional public fiscal expenditure per capita. All outcome variables are winsorized at the 1% and 99% levels. Standard errors are clustered at the court level, *, ** and *** denote significance at 10%, 5% and 1% levels, respectively.

	Appeal Rate		Reversal Rate (%)	
	(1)	(2)	(3)	(4)
AI procurement	-0.146**	-0.151**	-2.384*	-2.305*
	(0.063)	(0.064)	(1.342)	(1.318)
Controls	NO	YES	NO	YES
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.0765	0.0764	1.408	1.411
Observations	45,204	44,371	45,204	44,371
First Stage Coefficient	0.062***	0.063***	0.062***	0.063***
	(0.019)	(0.019)	(0.019)	(0.019)
Baseline Coefficient	-0.022***	-0.022***	-0.565**	-0.559***
KP F Stat.	10.59	10.90	10.59	10.90

Table 5. Comparison with Digital Procurement

This table controls for the procurement of digital technologies and examines the additional impact of AI technology compared to other digital technologies on appeal rates and reversal rates. In columns (2) and (4), we further control for regional characteristics including regional GDP per capita, regional population, and regional public fiscal expenditure per capita. All outcome variables are winsorized at the 1% and 99% levels. Standard errors are clustered at the court level, *, ** and *** denote significance at 10%, 5% and 1% levels, respectively.

	(1)	(2)	(3)	(4)
AI procurement	-0.018**	-0.017**	-0.494***	-0.484***
	(0.007)	(0.007)	(0.166)	(0.166)
Digital procurement	-0.010***	-0.010***	-0.143**	-0.151**
	(0.003)	(0.003)	(0.065)	(0.066)
Controls	NO	YES	NO	YES
Observations	45,204	44,371	45,204	44,371
Adjusted R-squared	0.380	0.382	0.282	0.283
COURT FE	YES	YES	YES	YES
YEAR BY QUARTER FE	YES	YES	YES	YES
Mean of outcome	0.0765	0.0764	1.408	1.411

Table 6. AI Impacts on Courts' Efficiency and Discretion

This table examines the impact of AI adoption on two measures of case-processing efficiency: judgment delay and discretion. In columns (2) and (4), we further control for regional characteristics including regional GDP per capita, regional population, and regional public fiscal expenditure per capita. All outcome variables are winsorized at the 1% and 99% levels. Standard errors are clustered at the court level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Judgment Delay (log)		Discretion (log)	
	(1)	(2)	(3)	(4)
AI procurement	-0.117***	-0.111***	-0.016	-0.014
	(0.031)	(0.031)	(0.034)	(0.034)
Controls	NO	YES	NO	YES
Observations	44,501	43,684	45,113	44,281
Adjusted R-squared	0.356	0.356	0.448	0.449
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	4.388	4.385	-0.442	-0.444

Table 7. AI Impacts on Court Personnel

This table examines the impact of AI adoption on the number of court personnel, including judges responsible for adjudicating cases and clerks who assist in case proceedings. In columns (2) and (4), we further control for regional characteristics including regional GDP per capita, regional population, and regional public fiscal expenditure per capita. All outcome variables are winsorized at the 1% and 99% levels. Standard errors are clustered at the court level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Number of Judges (log)		Number of Court Clerks (log)	
	(1)	(2)	(3)	(4)
AI procurement	-0.216***	-0.213***	-0.196***	-0.195***
	(0.055)	(0.056)	(0.073)	(0.073)
Controls	NO	YES	NO	YES
Observations	45,256	44,419	45,261	44,421
Adjusted R-squared	0.806	0.804	0.806	0.802
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	3.450	3.443	3.437	3.425

Table 8. Different Types of AI Technology

This table examines the heterogeneity in the impact of AI adoption on court performance across different types of AI technologies. We classify AI technologies into trial-specific and routine-specific applications according to whether the technology primarily assists adjudication-related tasks or supports other routine administrative and operational tasks. All specifications control for regional characteristics, including regional GDP per capita, regional population, and regional public fiscal expenditure per capita. All outcome variables are winsorized at the 1% and 99% levels. Standard errors are clustered at the court level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Appeal Rate			Reversal Rate			Judgment Delay (logs)		
	(1)	(2)	(3)	(4)	(5)	(6)	(10)	(11)	(12)
Trail-specific AI procurement	-0.022**		-0.022**	-0.591**		-0.596**	0.010		0.019
	(0.011)		(0.011)	(0.280)		(0.280)	(0.062)		(0.062)
Routine-task AI procurement		-0.005	-0.005		0.054	0.072		-0.139***	-0.139***
		(0.006)	(0.006)		(0.274)	(0.274)		(0.051)	(0.051)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	44,371	44,371	44,371	44,371	44,371	44,371	43,684	43,684	43,684
Adjusted R-squared	0.380	0.380	0.380	0.282	0.282	0.282	0.356	0.356	0.356
COURT FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
YEAR BY QUARTER FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Mean of outcome	0.0764	0.0764	0.0764	1.411	1.411	1.411	4.385	4.385	4.385

Table 9. AI Readiness

This table examines the heterogeneity in the impact of AI adoption on court performance across different levels of AI readiness. We classify courts into high- and low-AI-readiness groups based on whether they had procured other digital technologies before their first AI procurement. Courts with prior procurement of other digital technologies are classified as high-readiness courts, while those without such prior procurement are classified as low-readiness courts. All specifications control for regional characteristics, including regional GDP per capita, regional population, and regional public fiscal expenditure per capita. All outcome variables are winsorized at the 1% and 99% levels. Standard errors are clustered at the court level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Appeal Rate		Reversal Rate (%)	
	Low Readiness	High Readiness	Low Readiness	High Readiness
	(1)	(2)	(3)	(4)
AI procurement	-0.018	-0.023***	-0.361	-0.687***
	(0.011)	(0.008)	(0.259)	(0.195)
Controls	YES	YES	YES	YES
Observations	11,901	32,470	11,901	32,470
Adjusted R-squared	0.392	0.379	0.269	0.290
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.0797	0.0752	1.491	1.381

Table 10. Heterogeneity: Procurement Amount

This table examines the heterogeneity in the impact of AI adoption on court performance across different procurement amounts. We classify treated courts into low-, medium-, and high-intensity groups based on the terciles of the first AI procurement amount across all courts. All specifications control for regional characteristics, including regional GDP per capita, regional population, and regional public fiscal expenditure per capita. All outcome variables are winsorized at the 1% and 99% levels. Standard errors are clustered at the court level, *, ** and *** denote significance at 10%, 5% and 1% levels, respectively.

	Appeal Rate				Reversal Rate (%)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AI procurement (Low)	-0.018			-0.014	-0.460			-0.367
	(0.012)			(0.013)	(0.284)			(0.295)
AI procurement (Median)		-0.013		-0.012		-0.347*		-0.314
		(0.009)		(0.009)		(0.178)		(0.192)
AI procurement (High)			-0.026**	-0.024*			-0.650*	-0.593*
			(0.013)	(0.013)			(0.333)	(0.344)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	44,371	44,371	44,371	44,371	44,371	44,371	44,371	44,371
Adjusted R-squared	0.381	0.380	0.381	0.381	0.282	0.282	0.282	0.283
Court FE	YES	YES	YES	YES	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES	YES	YES	YES	YES
Mean of outcome	0.0764	0.0764	0.0764	0.0764	1.411	1.411	1.411	1.411

Table 11. Heterogeneity: Court Region Characteristics

This table presents regional heterogeneity in the impact of AI adoption on court performance. We split the sample based on whether a court's region has above- or below-median GDP per capita, and whether it is located in western China. All specifications control for regional characteristics, including regional GDP per capita, regional population, and regional public fiscal expenditure per capita. All outcome variables are winsorized at the 1% and 99% levels. Standard errors are clustered at the court level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	PANEL A: GDP Per Capita			
	Appeal Rate		Reversal Rate (%)	
	Below Median	Above Median	Below Median	Above Median
	(1)	(2)	(3)	(4)
AI procurement	-0.028***	-0.014	-0.674***	-0.419*
	(0.009)	(0.010)	(0.250)	(0.216)
Controls	YES	YES	YES	YES
Observations	23,785	19,749	23,785	19,749
Adjusted R-squared	0.374	0.390	0.265	0.305
COURT FE	YES	YES	YES	YES
YEAR BY QUARTER FE	YES	YES	YES	YES
Mean of outcome	0.0738	0.0786	1.448	1.342
	PANEL B: Western & Non-western			
	Appeal Rate		Reversal Rate (%)	
	Western	Non-western	Western	Non-western
	(1)	(2)	(3)	(4)
AI procurement	-0.028**	-0.018**	-0.860***	-0.394**
	(0.013)	(0.008)	(0.297)	(0.187)
Controls	YES	YES	YES	YES
Observations	20,152	24,219	20,152	24,219
Adjusted R-squared	0.375	0.393	0.278	0.289
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.0758	0.0769	1.532	1.310

Table 12. Heterogeneity across Case Complexity

This table studies the heterogeneity of AI technology's impact on court performance across varying case complexities. We classify cases where the descriptive text length of court documents exceeds the median sample length as complex cases. We then construct court performance metrics for both complex and non-complex cases as our new outcome variables. All specifications control for regional characteristics, including regional GDP per capita, regional population, and regional public fiscal expenditure per capita. All outcome variables are winsorized at the 1% and 99% levels. Standard errors are clustered at the court level, *, ** and *** denote significance at 10%, 5% and 1% levels, respectively.

	Appeal Rate		Reversal Rate (%)	
	Complex Case	Non-complex Case	Complex Case	Non-complex Case
	(1)	(2)	(3)	(4)
AI procurement	-0.025***	-0.017***	-0.722***	-0.321**
	(0.008)	(0.006)	(0.190)	(0.136)
Controls	YES	YES	YES	YES
Observations	44,211	43,775	44,211	43,775
Adjusted R-squared	0.362	0.286	0.264	0.190
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.105	0.0289	1.895	0.554

Table 13. Heterogeneity across Case Types

This table examines heterogeneity in the impact of AI adoption on court performance across different case types. We construct performance measures separately for civil, criminal, and administrative cases and use these as outcome variables in the regressions reported in Panels A, B, and C, respectively. In columns (2) and (4), we further control for regional characteristics including regional GDP per capita, regional population, and regional public fiscal expenditure per capita. All outcome variables are winsorized at the 1% and 99% levels. Standard errors are clustered at the court level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Appeal Rate			Reversal Rate (%)		
	Civil Cases	Criminal Cases	Administrative Cases	Civil Cases	Criminal Cases	Administrative Cases
	(1)	(2)	(3)	(4)	(5)	(6)
AI procurement	-0.021***	-0.026***	-0.080***	-0.563***	-0.362*	-1.526**
	(0.007)	(0.010)	(0.025)	(0.170)	(0.199)	(0.655)
Controls	YES	YES	YES	YES	YES	YES
Observations	44,270	40,777	21,797	44,270	40,777	21,797
Adjusted R-squared	0.346	0.328	0.234	0.266	0.170	0.072
Court FE	YES	YES	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES	YES	YES
Mean of outcome	0.0760	0.0785	0.280	1.499	0.907	3.327

APPENDIX A: Additional Figures and Tables

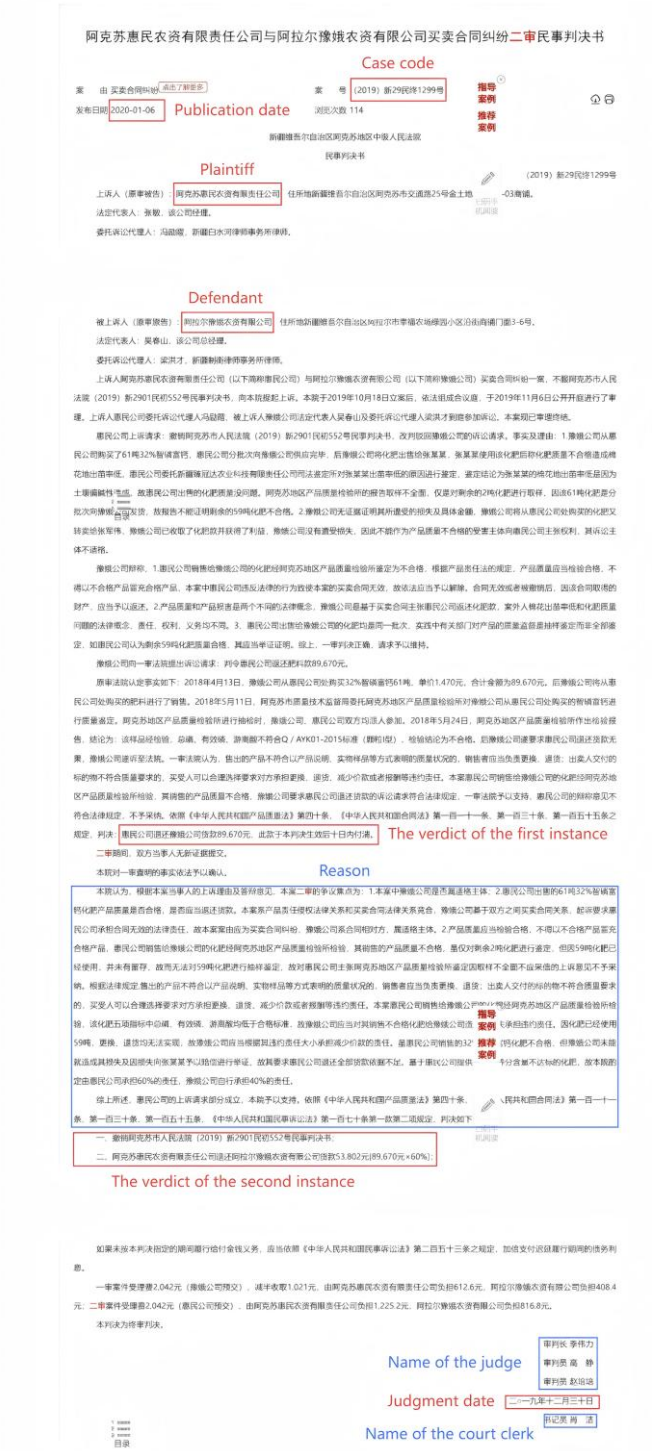


Figure A1. Judicial Document Example

This figure presents an example of a second-instance judicial document from China Judgments Online (CJO). Such documents include the case code, publication date, litigants (plaintiff and defendant, if applicable), a description of the case, the court's reasoning, the names of the presiding judges and clerks, and the judgment date. For second-instance documents, the case code of the corresponding first-instance judgment is also included, enabling us to identify the appealed case.

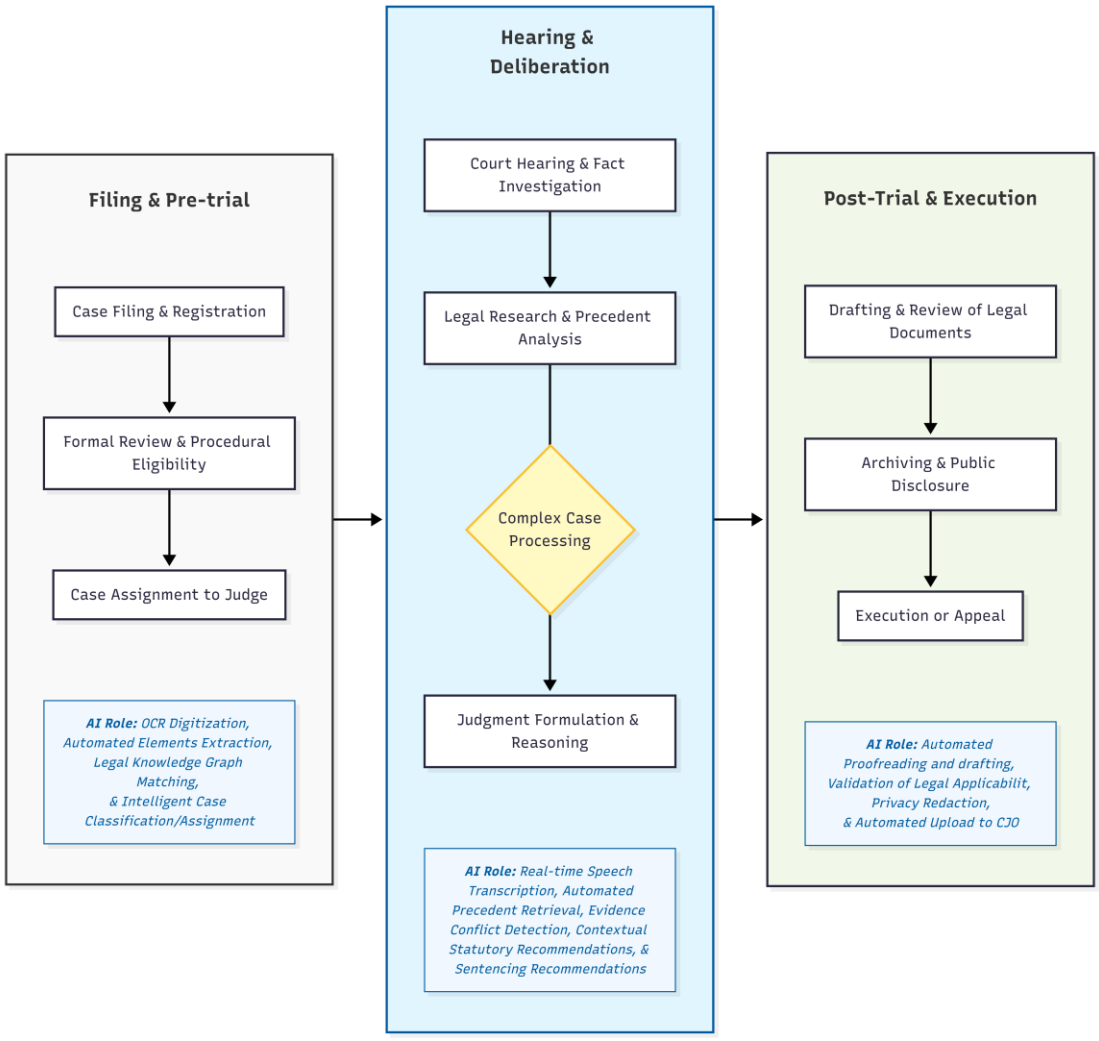


Figure A2. Process Map of Judicial Decision

This figure illustrates the typical process of a judicial decision generation as well as the role of AI in different stages.



Figure A3. Court AI Procurement Contract Example

This figure provides an example of the court AI procurement contract from Chinese Central Government Procurement (CCGP) website. Each contract includes information on the contract title, purchasing entity, supplier, main content, contract amount, and signing date. We screen contracts in which the purchaser is a court and determine whether they involve AI technology based on the contract title and main content.

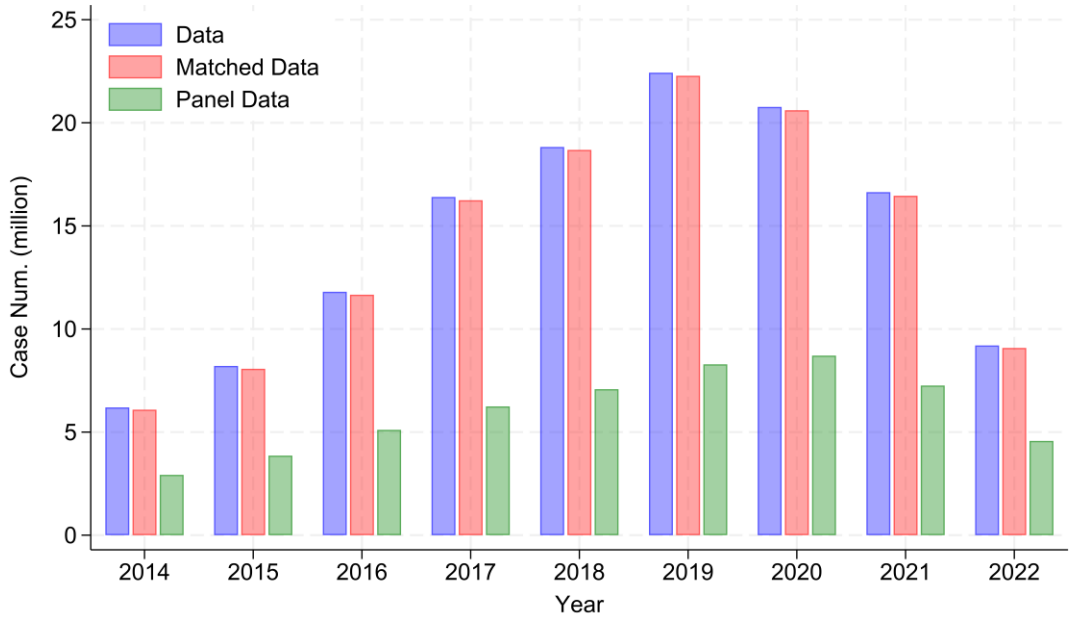


Figure A4. Sample Construction and Case Coverage

This figure shows the annual case coverage across the main stages of data construction from 2014 to 2022. "Data" denotes the raw case records collected from the original source. "Matched Data" denotes the subset of cases successfully linked to court identifiers. "Panel Data" denotes the final sample used in the empirical analysis after merging with the court-level panel and imposing the baseline sample restrictions. Case counts are reported in millions.

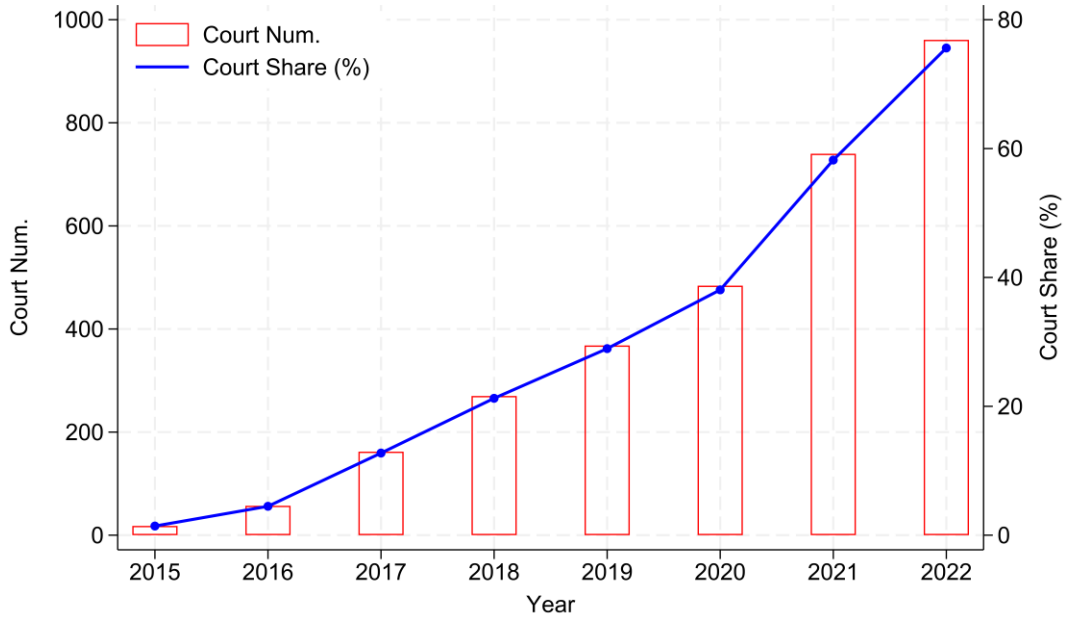


Figure A5. The Trend of Digital Adoption

This chart illustrates the temporal trend of Chinese courts acquiring digital technology (including AI). The red bar chart represents the cumulative number of courts procuring digital technology, while the blue line chart shows the proportion of such courts relative to the sample courts.

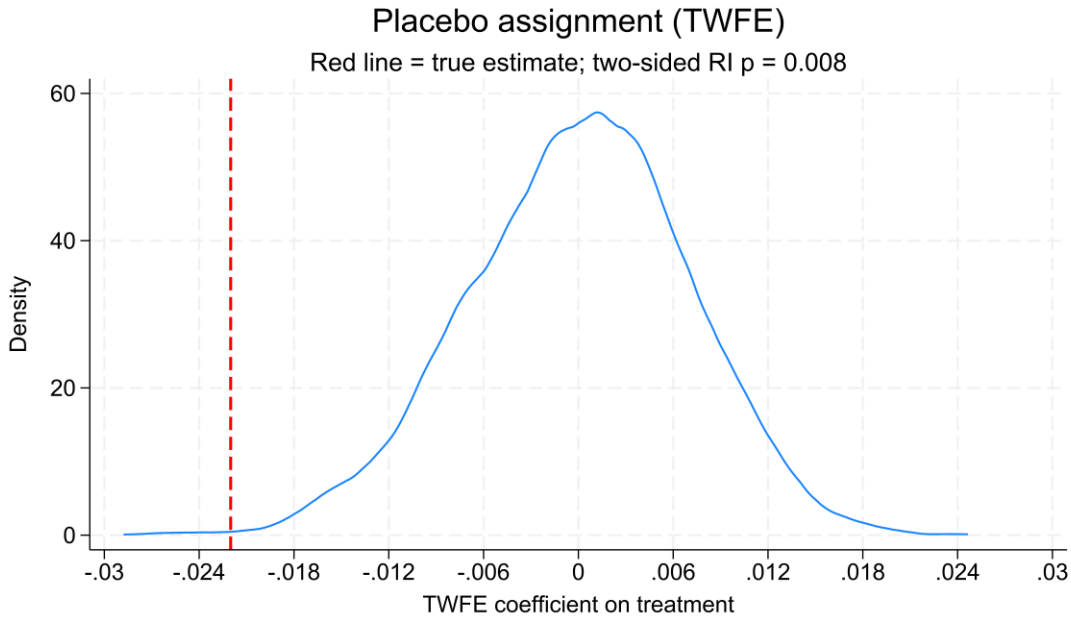


Figure A6(a): Permutation Test for Appeal Rate

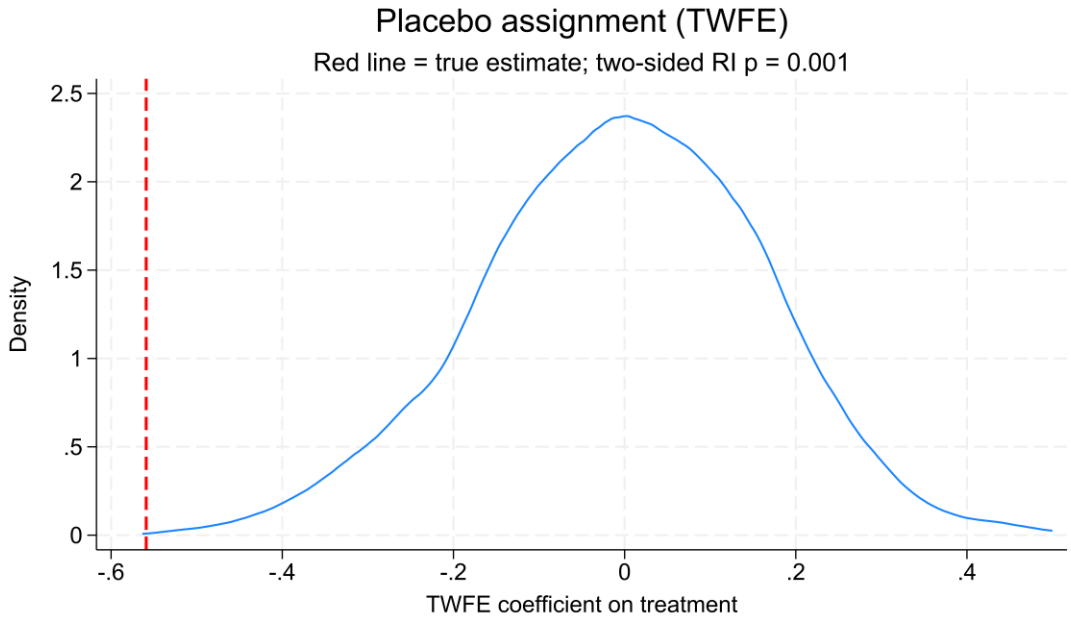


Figure A6(b): Permutation Test for Reversal Rate

Figure A6. Permutation Tests for Main Outcomes

The figure plots the distribution of 3,000 placebo coefficients generated by randomly reassigning treatment timing across courts. Panel (a) shows the results for the

Appeal Rate; Panel (b) shows the results for the Reversal Rate. The dashed red lines indicate the true point estimates. Empirical p-values are reported in the sub-titles.

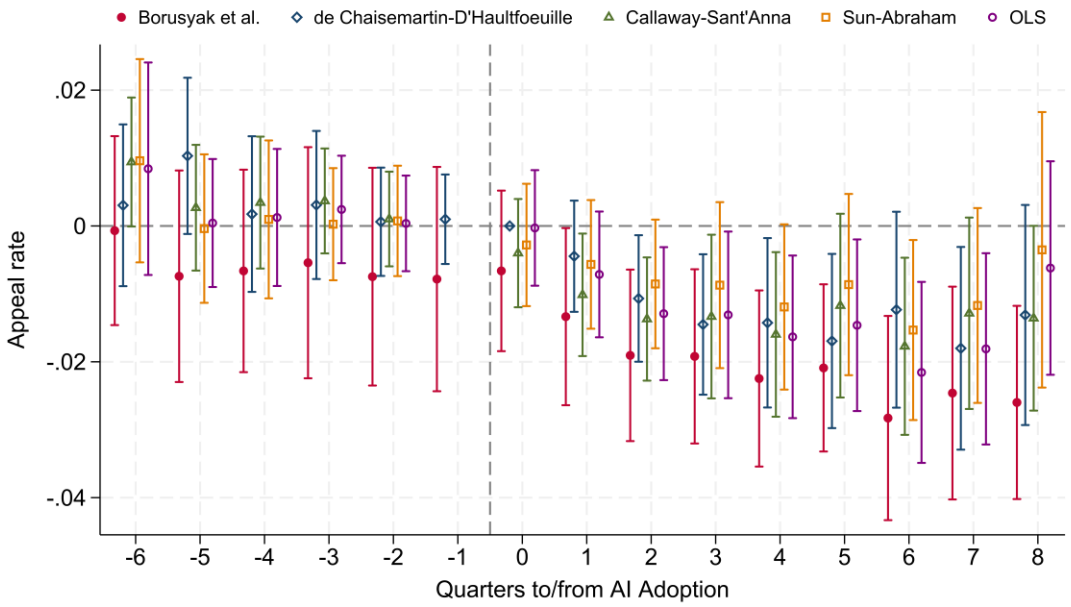


Figure A7(a): Event Study for Appeal Rate (Extra)

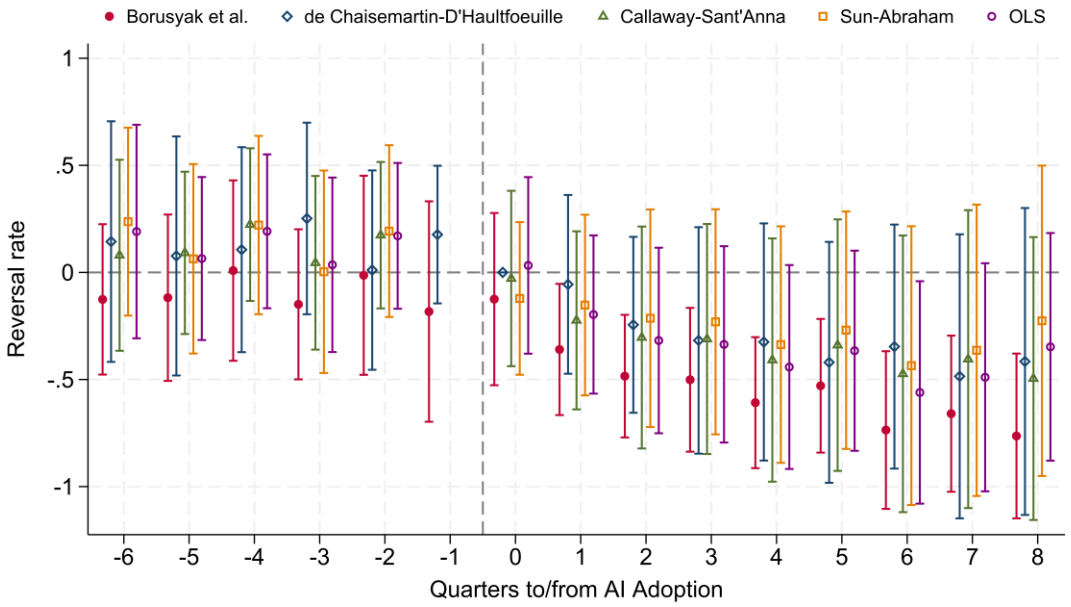


Figure A7(b): Event Study for Reversal Rate (Extra)

Figure A7. Event Study Results

This figure plots the estimates and their 95% confidence intervals from TWFE-OLS (in purple circle) and four methods addressing negative weights in staggered difference-in-difference, including Borusyak et al. (2024) in red dot, De Chaisemartin and D'Haultfoeuille (2020) in blue diamond, Sun and Abraham (2021) in green triangle, and Callaway and Sant'Anna (2021) in yellow square. All specifications control for

courts' digital procurement to account for the influence of broader digitalization. The omitted period is quarter −1 for estimators normalized relative to the last pre-treatment period. For Borusyak et al. (2024) and De Chaisemartin and D'Haultfoeuille (2020), the reported pre-treatment coefficients are placebo estimates rather than coefficients normalized by omitting quarter −1.

Table A1. LLM Prompt and Parameters to Identify AI Procurements

This table presents the prompt used for LLM-based measurement. We use a fixed version of ChatGPT 4.1, *gpt-4.1-2025-04-14*, and set the temperature to 0 to improve the replicability of the measurement outcomes. To assess robustness to the choice of LLM, we also apply the same prompt and parameter settings to the official release of DeepSeek V3.2 and use its outputs to construct an alternative measure of court AI adoption.

model="gpt-4.1-2025-04-14",
instructions="You are an expert in judicial administration, public-sector procurement, and the application of artificial intelligence in Chinese courts. Your task is to determine whether a given court procurement description involves the adoption of artificial intelligence technologies. These technologies should be used to save time, automate or partially replace labor-intensive tasks, or improve the performance or efficiency of judicial decision-making. Judicial decision-making includes case assignment, case adjudication, legal reasoning, evidence review, sentencing, and case-processing workflows handled by judges or judicial staff.",
input=f"Please carefully read the contract title, item information, specifications, model, or service requirements of the procurement project: "*INPUT TEXT*", and determine: 1. Whether the project applies artificial intelligence technologies to support, assist, or improve judicial decision-making or case-processing efficiency. If yes, answer 1; if no or if it cannot be determined, answer 0. 2. Whether the project applies digital technologies to support, assist, or improve judicial decision-making or case-processing efficiency. If yes, answer Y; if no or if it cannot be determined, answer N. Please answer only [1,Y], [1,N], [0,Y], or [0,N]. Do not provide any explanation. You must answer. I have no additional information.",
stream=False, temperature=0, timeout=5

Table A2. (Lack of) Correlation of AI Adoption with Lagged Courts' Performance and Economic Aggregates

This table reports the correlations between AI adoption and lagged measures of court performance and local economic conditions. All outcome variables are winsorized at the 1% and 99% levels. Standard errors are clustered at the court level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	AI procurement		
	Previous 1 year	Previous 2 year	Previous 3 year
	(1)	(2)	(3)
Appeal Rate	-0.030	-0.034*	-0.034
	(0.020)	(0.020)	(0.022)
Reversal Rate	-0.001	-0.000	-0.000
	(0.001)	(0.001)	(0.001)
Total Delay (from acceptance to judgment)	-0.006	-0.008	-0.009*
	(0.005)	(0.005)	(0.006)
Expenditure Per Capita	-0.063	-0.061	-0.062
	(0.047)	(0.048)	(0.050)
GDP Per Capita	0.022	0.022	0.021
	(0.016)	(0.016)	(0.016)
Population	-0.172**	-0.164**	-0.156**
	(0.082)	(0.080)	(0.077)
Population Growth Rate	0.020	0.020	0.019
	(0.022)	(0.022)	(0.022)
Case Number	-0.003	-0.007	-0.007
	(0.011)	(0.011)	(0.011)
Judge Number (log)	-0.023***	-0.022***	-0.020***
	(0.007)	(0.007)	(0.007)
Case Number Growth Rate	0.001	-0.000	-0.001
	(0.001)	(0.001)	(0.001)
Constant	0.184***	0.182***	0.172***
	(0.050)	(0.049)	(0.047)
Observations	41,311	40,109	38,887
Adjusted R-squared	0.377	0.388	0.400
COURT FE	YES	YES	YES
YEAR BY QUARTER FE	YES	YES	YES

Table A3. Variable Definitions

This table provides detailed definitions of the variables used in our analysis, including court performance and efficiency measures, regional characteristics (used as control variables), and the instrumental variable.

AI Procurement	Binary variable which takes 1 since the court signed its first AI procurement contract and 0 else.
Smart Court	Binary variable which takes 1 since the court signed its first smart court procurement contract and 0 else.
Digital Procurement	Binary variable which takes 1 since the court signed its first digital (including AI) procurement contract and 0 else.
Judicial Independence Reform	Binary variable which takes 1 since the court underwent the reform of unified management of personnel, finances and property, and 0 else.
Judge Quota Reform	Binary variable equal to 1 for periods after the court implemented the judge quota reform, which restricts the number of judges formally authorized to adjudicate cases, and 0 otherwise.
Instrumental Variable	We use the number of AI-related procurement contracts issued by the Environmental Protection Bureau in the same prefecture-level city as an alternative instrument for court AI adoption.
Appeal Rate	Number of appeal cases / Number of first-instance cases
Reversal Rate (%)	100 × Number of cases with overturned verdicts / Number of first-instance cases
Strict Reversal Rate (%)	100 × Number of cases with fully overturned verdicts / Number of first-instance cases
Non-strict Reversal Rate (%)	100 × Number of cases with patially overturned verdicts / Number of first-instance cases
Judgment Delay (days)	The time interval from case acceptance to judgment.
Discretion	Follow Liu et al. (2024), we construct the share of cases in which judges cite legal provisions that allow discretionary judgment in adjudication.
Number of Judges	The number of adjudicating judges serving in a court.
Number of Court Clerks	The number of court clerks serving in a court, who are responsible for trial records, procedural documentation, and other clerical support in adjudication.
GDP Per Capita (in 10k CNY)	GDP per capita in 10k CNY, district and county courts use the values of their respective districts, while city level courts use the values of their respective cities.
Population (in millions)	Population in millions, district and county courts use the values of their respective districts, while city level courts use the values of their respective cities.
Public Expenditure Per Capita (in 10k CNY)	Public expenditure per capita in 10k CNY, district and county courts use the values of their respective districts, while city level courts use the values of their respective cities.

Table A4. Sample Comparison

This table compares court performance and regional characteristics between courts included in our sample and those not included, using data from the first quarter of our sample period (the last quarter of 2014). For courts in our sample, we observe at least one procurement contract as well as judicial documents. For courts not in the sample, judicial documents are available, but no procurement records exist. Mean1 and Mean2 refer to the sample means for courts not in the sample and courts in the sample, respectively. P-values from t-tests of mean differences are reported. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Variables	Mean1	Mean2	Mean Diff.	P-value
Appeal Rate	0.171	0.163	0.00743	0.337
Reversal Rate	0.0332	0.0329	0.000336	0.897
Strict Reversal Rate	0.0191	0.0196	-0.000536	0.751
Non-strict Reversal Rate	0.0129	0.0125	0.000333	0.792
Judgment Delay	90.1	86.9	3.14	0.225
Number of Cases	278	291	-13.5	0.483
GDP Per Capita	37004	40389	-3385	0.0128**
Population	114	143	-29	0.024**
Public Expenditure Per Capita	5426	6122	-696	0.00244***

Table A5. Distinguish Strict Reversal and Non-strict Reversal Rates

This table decomposes reversal rate into two measures: strict and non-strict reversal rates. All outcome variables are winsorized at the 1% and 99% levels. Standard errors are clustered at the court level, *, ** and *** denote significance at 10%, 5% and 1% levels, respectively.

	Strict Reversal Rate		Non-strict Reversal Rate	
	(1)	(2)	(3)	(4)
AI procurement	-0.304***	-0.298***	-0.251***	-0.251***
	(0.097)	(0.097)	(0.068)	(0.067)
Controls	NO	YES	NO	YES
Observations	45,204	44,371	45,204	44,371
Adjusted R-squared	0.250	0.250	0.214	0.214
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.830	0.835	0.552	0.549

Table A6. Baseline Results: Excluding 2022

To eliminate the interference caused by the increased proportion of missing judicial document data in 2022, this table reports the baseline regression results for our sample after excluding 2022. All outcome variables are winsorized at the 1% and 99% levels. Standard errors are clustered at the court level, *, ** and *** denote significance at 10%, 5% and 1% levels, respectively.

	Appeal Rate		Reversal Rate (%)	
	(1)	(2)	(3)	(4)
AI procurement	-0.025***	-0.025***	-0.638***	-0.634***
	(0.008)	(0.008)	(0.190)	(0.191)
Controls	NO	YES	NO	YES
Observations	40,223	39,478	40,223	39,478
Adjusted R-squared	0.363	0.366	0.272	0.273
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.0834	0.0833	1.545	1.548

Table A7. Robustness to Alternative LLM-Based Measures of AI Adoption

This table examines the robustness of the estimated impact of AI adoption on court appeal rates and reversal rates using alternative LLM-based measures of AI adoption. Columns (1) and (3) define AI adoption based on DeepSeek outputs, while columns (2) and (4) define AI adoption based on the union of GPT and DeepSeek outputs. All specifications control for regional characteristics, including regional GDP per capita, regional population, and regional public fiscal expenditure per capita. All outcome variables are winsorized at the 1% and 99% levels. Standard errors are clustered at the court level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Appeal Rate		Reversal Rate (%)	
	(1)	(2)	(3)	(4)
AI procurement (DeepSeek)	-0.021***		-0.499***	
	(0.007)		(0.161)	
AI procurement (GPT or DeepSeek)		-0.021***		-0.498***
		(0.006)		(0.143)
Controls	YES	YES	YES	YES
Observations	44,371	44,371	44,371	44,371
Adjusted R-squared	0.381	0.381	0.283	0.283
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.0764	0.0764	1.411	1.411

Table A8. Robustness to Controlling for Upper-Level Court AI Procurement

This table examines whether the estimated impact of AI adoption on court appeal rates and reversal rates is robust to controlling for AI procurement by corresponding upper-level courts. Columns (1) and (3) report the baseline estimates, while columns (2) and (4) additionally control for whether the court's corresponding upper-level court has adopted AI technology. All specifications control for regional characteristics, including regional GDP per capita, regional population, and regional public fiscal expenditure per capita. All outcome variables are winsorized at the 1% and 99% levels. Standard errors are clustered at the court level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Appeal Rate		Reversal Rate (%)	
	(1)	(2)	(3)	(4)
AI procurement	-0.022***	-0.020***	-0.559***	-0.532***
	(0.007)	(0.007)	(0.162)	(0.163)
AI procurement (Upper-level)		-0.011**		-0.147
		(0.005)		(0.118)
Controls	YES	YES	YES	YES
Observations	44,371	44,371	44,371	44,371
Adjusted R-squared	0.381	0.382	0.283	0.283
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.0764	0.0764	1.411	1.411

Table A9. Robustness to Weighting by Court Size

This table examines the robustness of the estimated impact of AI adoption on court appeal rates and reversal rates using weighted least squares. Columns (1) and (3) weight observations by the number of cases, while columns (2) and (4) weight observations by the number of judges. All specifications control for regional characteristics, including regional GDP per capita, regional population, and regional public fiscal expenditure per capita. All outcome variables are winsorized at the 1% and 99% levels. Standard errors are clustered at the court level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Appeal Rate		Reversal Rate (%)	
	Weight: Case Num.	Weight: Judge Num.	Weight: Case Num.	Weight: Judge Num.
	(1)	(2)	(3)	(4)
AI procurement	-0.017***	-0.017***	-0.477***	-0.422***
	(0.007)	(0.006)	(0.150)	(0.130)
Controls	YES	YES	YES	YES
Observations	44,370	44,371	44,370	44,371
Adjusted R-squared	0.397	0.442	0.318	0.359
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.0764	0.0764	1.411	1.411

Table A10. Robustness Check with Additional Court-Size Controls

This table examines the robustness of the estimated impact of AI adoption on court appeal rates and reversal rates after additionally controlling for court size. We include the number of cases and the logarithm of the number of judges as additional controls. All specifications also control for regional characteristics, including regional GDP per capita, regional population, and regional public fiscal expenditure per capita. All outcome variables are winsorized at the 1% and 99% levels. Standard errors are clustered at the court level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Appeal Rate		Reversal Rate (%)	
	(1)	(2)	(3)	(4)
AI procurement	-0.019***	-0.019***	-0.471***	-0.468***
	(0.007)	(0.007)	(0.159)	(0.159)
Case Num.	0.027***	0.028***	0.599***	0.644***
	(0.006)	(0.006)	(0.122)	(0.122)
Judge Num. (log)	0.014***	0.014***	0.383***	0.381***
	(0.002)	(0.002)	(0.050)	(0.051)
Controls	YES	YES	YES	YES
Observations	45,203	44,370	45,203	44,370
Adjusted R-squared	0.383	0.385	0.286	0.287
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.0765	0.0764	1.408	1.411

Table A11. Heterogeneous Effects by Firm Involvement in Cases

This table examines whether the impact of AI adoption on court performance differs by firm involvement in cases. We separately construct outcome variables for cases involving firms and cases not involving firms, and estimate the effect of AI procurement on appeal rates and reversal rates in each subsample. Columns (1) and (3) report results for firm-involved cases, while columns (2) and (4) report results for non-firm cases. All specifications control for regional characteristics, including regional GDP per capita, regional population, and regional public fiscal expenditure per capita. All outcome variables are winsorized at the 1% and 99% levels. Standard errors are clustered at the court level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Appeal Rate		Reversal Rate (%)	
	Firm	Non-firm	Firm	Non-firm
	(1)	(2)	(3)	(4)
AI procurement	-0.015**	-0.022***	-0.319	-0.563***
	(0.008)	(0.007)	(0.209)	(0.162)
Controls	YES	YES	YES	YES
Observations	43,303	44,325	43,303	44,325
Adjusted R-squared	0.234	0.378	0.134	0.267
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.106	0.0704	2.089	1.256

APPENDIX B: Proofs of Theoretical Results

Proof of Proposition 1 (H1).

The Lagrangian is

$$\mathcal{L} = \int_0^{\theta} \ln\left(y_j\right) dj - \lambda\left(\int_0^{\theta} h_j dj - H\right).$$

FOC for h_j :

$$\frac{1}{y_j}\frac{\partial y_j}{\partial h_j} = \lambda.$$

Compute the marginal product:

$$\frac{\partial y_j}{\partial h_j} = y_j^{\sigma} h_j^{-\sigma}.$$

Thus, $\frac{1}{y_j}\frac{\partial y_j}{\partial h_j} = y_j^{\sigma-1} h_j^{-\sigma} = \lambda$ (constant across tasks), or $h_j = \lambda^{-1/\sigma} y_j^{(\sigma-1)/\sigma}$.

Next, we explore the impact of AI adoption on judicial quality. The envelope theorem gives:

$$\frac{\partial Q^*}{\partial A} = \int_0^{\theta} \frac{1}{y_j}\frac{\partial y_j}{\partial A} dj$$

Since:

$$\frac{\partial y_j}{\partial A} = y_j^{\sigma}\left(\alpha_j A\right)^{-\sigma} \alpha_j.$$

Hence,

$$\frac{\partial Q^*}{\partial A} = \int_0^{\theta} \alpha_j^{1-\sigma} A^{-\sigma} y_j^{\sigma-1} dj > 0 \qquad (1)$$

Therefore, AI adoption increases judicial quality in the equilibrium, which constitutes our first hypothesis.

H1. AI adoption has a positive effect on court performance.

Proof of Proposition 2 (H2).

To investigate whether AI has a larger impact on complex cases, we examine how the marginal benefit of AI changes with complexity θ:

$$\frac{\partial}{\partial\theta}\left(\frac{\partial Q^*}{\partial A}\right) = \frac{\partial}{\partial\theta}\left(\int_0^\theta \frac{\partial \ln\, y_j}{\partial A} dj\right)$$

By the Leibniz Integral Rule, this derivative is:

$$\frac{\partial^2 Q^*}{\partial A \partial\theta} = \frac{\partial \ln\, y_\theta}{\partial A} = \frac{1}{y_\theta}\frac{\partial y_\theta}{\partial A} = \alpha_\theta^{1-\sigma} \mathrm{A}^{-\sigma} \mathrm{y}_\theta^{\sigma-1} > 0$$

The total gain from AI is cumulative across the task continuum, which yields our second hypothesis:

H2. The positive impact of AI adoption is larger among complex cases.

Lemma 1 (supporting Proposition 3).

Lemma 1. The optimal allocation of human effort h_j^* increases with total capacity H, i.e., $\frac{\partial h_j^*}{\partial H} > 0$.

Since h_j is determined by λ, we decompose $\frac{\partial h_j^*}{\partial H}$ into two parts using the chain rule:

$$\frac{\partial h_j^*}{\partial H} = \frac{\partial h_j^*}{\partial \lambda} \cdot \frac{\partial \lambda}{\partial H}$$

Then, we need to determine the sign of each part.

Step 1. Determine the sign of $\frac{\partial h_j^*}{\partial \lambda}$.

From the first order condition, $\frac{\partial \ln\, y_j}{\partial h_j} = \lambda$, the second-order condition (SOC) for a maximum requires the objective function to be concave, meaning $\frac{\partial^2 \ln\, y_j}{\partial h_j^2} < 0$. Differentiating the FOC $\left(\frac{\partial \ln\, y_j}{\partial h_j} = \lambda\right)$ with respect to λ :

$$\frac{\partial^2 \ln y_j}{\partial h_j^2} \cdot \frac{\partial h_j^*}{\partial \lambda} = 1$$

Since $\frac{\partial^2 \ln y_j}{\partial h_j^2} < 0$, it must be that $\frac{\partial h_j^*}{\partial \lambda} < 0$. This makes intuitive sense: as the shadow price of effort (λ) rises, the effort allocated to any task ***j*** falls.

Step 2: Determine the sign of $\frac{\partial \lambda}{\partial H}$.

We differentiate the budget constraint $\int_0^\theta h_j^* \, dj = H$ with respect to H:

$$\frac{\partial}{\partial H} \int_0^\theta h_j^*\big(\lambda(H)\big) dj = \frac{\partial H}{\partial H} \Rightarrow$$

$$\underbrace{\int_0^\theta \frac{\partial h_j^*}{\partial \lambda} dj}_{\text{Negative}} \cdot \frac{\partial \lambda}{\partial H} = 1$$

The first part is negative according to step 1. For the product of a negative term and $\frac{\partial \lambda}{\partial H}$ to equal 1 (a positive number), $\frac{\partial \lambda}{\partial H}$ must be negative. This represents the diminishing marginal utility of judicial capacity: as total resources H increase, the shadow value λ of an extra unit of effort decreases.

Therefore, the optimal allocation of human effort h_j^* increases with total capacity H, i.e., $\frac{\partial h_j^*}{\partial H} = \underbrace{\frac{\partial h_j^*}{\partial \lambda}}_{\text{Negative (Step 1)}} \cdot \underbrace{\frac{\partial \lambda}{\partial H}}_{\text{Negative (Step 2)}}$ is positive.

Proof of Proposition 3 (H3).

Lastly, we explore the heterogeneous impacts across regions with different levels of constraints. In this model, less developed regions are characterized by lower human capacity$H$. We analyze the cross-partial derivative of Q^* with respect to A and H. Using Equation (1) and the Envelope Theorem derivative:

$$\frac{\partial^2 Q^*}{\partial A \partial H} = \int_0^\theta \alpha_j^{1-\sigma} A^{-\sigma} (\sigma - 1) y_j^{\sigma-2} \frac{\partial y_j}{\partial H} dj$$

From the FOC structure,

$$\frac{\partial y_j}{\partial H} = \frac{\partial y_j}{\partial h_j}\frac{\partial h_j}{\partial H} = y_j^{\sigma} h_j^{-\sigma}\frac{\partial h_j^*}{\partial H}.$$

Since $\frac{\partial h_j^*}{\partial H} > 0$ (see Lemma 1), the sign is determined by $(\sigma - 1)$.

$$\frac{\partial^2 Q^*}{\partial A \partial H}\begin{cases} < 0, & \sigma < 1 \quad \text{(AI and judges are substitutes)} \\ > 0, & \sigma > 1 \quad \text{(AI and judges are complements)} \end{cases}$$

This yields our third hypothesis:

H3. When AI and human cognition are substitutes ($\sigma < 1$), AI adoption has larger marginal effects in courts with lower human capacity (i.e., less developed regions).