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 decisions. Employing a staggered difference-in-differences framework, we find that AI adoption reduces appeal rates by roughly 15% and reversal rates by 21%, while substantially speeding up case processing. These performance gains are especially pronounced 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 compared with other digital technologies, highlighting its unique contribution to judicial decision-making. Taken together, the results 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, and seminar participants at Shanghai Jiao Tong 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., 2021b; 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). Crucially, the institutional setting affords access to uniquely granular data on both AI procurement and judicial outcomes, 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 492 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 and from judgment rendering to online publication.

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 15%, 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 21%, 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 impose minimum contract-size thresholds—requiring AI procurement amounts to exceed 50,000 or 100,000 CNY—to ensure that only substantive technology investments are classified as AI adoption. The results remain nearly identical under these stricter criteria. Second, to address concerns that AI adoption may coincide with the judicial recentralization reform implemented during our sample period, we construct an indicator capturing the timing of this reform and include it in our baseline specifications; the findings remain robust. Third, after controlling for procurement related to broader "smart court" initiatives, we continue to find that

AI technology is the key driver of improvements in judicial decision quality. Finally, we employ an instrumental variable based on the weighted average AI adoption rate of other courts within the same province and quarter. Because courts in the same province may face common budgetary constraints, peer influences, or reform priorities set by provincial judicial authorities, the instrument is plausibly relevant. At the same time, AI adoption by neighboring courts is unlikely to be affected by contemporaneous, court-specific shocks—such as leadership changes, judge turnover, or local shifts in case composition—supporting the exclusion restriction. The IV estimates further reinforce our baseline results.

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 only weak or non-significant effects. This critical 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 time between judgment issuance and online publication decreases by 23%, and the duration from case acceptance to judgment falls by 8%. We further examine whether AI technologies substitute for or complement labor in courts. The results show that AI adoption significantly reduces the number of clerks but has no discernible effect on the number of judges, suggesting that AI primarily replaces routine administrative tasks while leaving judicial decision-making roles unaffected.

We then explore heterogeneity in the effects of judicial AI adoption along several dimensions: procurement size, regional development, and case type. First, AI procurements above the median 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. Second, we find that AI adoption yields the strongest benefits in less developed regions—courts located in provinces with lower GDP per capita, in western China, and in non-coastal areas. A plausible explanation is that these courts face tighter human capital and fiscal constraints, making them more reliant on technological tools to alleviate capacity limitations. Third, AI generates substantially larger improvements for complex cases than for non-complex ones, underscoring its role in helping courts manage more intricate and information-intensive disputes. We also find sizeable performance gains in civil cases, while the effects on criminal and administrative cases are negligible.

These findings carry several important policy implications. First, the reductions in appeal and reversal rates highlight AI's potential to enhance both fairness and 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. Third, governments should institutionalize ongoing monitoring and evaluation of AI use in judicial settings, develop regulatory frameworks to mitigate risks related to bias and opacity, and ensure transparency to the public—thereby strengthening trust in legal institutions.

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 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., 2021b; 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., 2022; Ng and He, 2017; Popova, 2012; Caplin et al., 2024; Bhuller and Sigstad, 2025). 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−live broadcasting−shapes judicial decision-making, showing that it can help narrow gender disparities in litigation outcomes. Cao, Ding, and Zhou (2024) shows 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—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).[2] This distribution ensures that routine legal

[2] 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.

matters are handled efficiently at the local level while reserving higher courts for cases requiring specialized expertise or having broader implications.[3]

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.[4] Under this framework, 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. 2022; Baye and Wright 2011).[5]

[3] 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.

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

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

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 Judgements 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.

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.[6]

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 2020. 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.[7] According to SPC officials, this reduction in disclosure reflects growing institutional concerns regarding platform usability, individual rights protection, and information security risks.[8] 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. 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

[6] For more details, please see the report: https://paper.people.com.cn/zgjjzk/html/2024-01/30/nw.zgjjzk_20240130_2-02.htm.
[7] 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).
[8] For more details, please see the report: http://www.legaldaily.com.cn/Village_ruled_by_law/content/2024-01/18/content_8951881.html.

court modifications of initial judgments typically stem from several categories of first-instance deficiencies: erroneous factual 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%. [9] 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, adopt scientific classification management, and precisely define functions.[10]

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. The implementation of this quota system reached a critical milestone in 2017, when the SPC selected approximately 125,000 judges (representing 60% of the existing judicial workforce) from a pool of about 210,000 nationwide. Under the reformed system, only judges admitted to the quota possess

[9] 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.

[10] 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.

the authority to adjudicate cases, effectively removing 85,000 personnel from active judicial roles.[11]

This simultaneous expansion of caseload and contraction of judicial personnel has created severe resource constraints within the system. 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.[12] This acute imbalance 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

[11] 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.

[12] 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.

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. 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.[13] This gap in the literature motivates our comprehensive

[13] 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.

empirical investigation into the causal impacts of AI adoption on measurable court performance outcomes.

3. Data and Measurement

3.1 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 Appendix. Figure A1 in the Appendix provides an example of such AI-related procurement contract. Following this approach, we identify 492 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 180 courts—about 12% 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, Appendix Table A2 shows no systematic correlation between AI adoption and measures of courts' performance 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.

3.2. 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.[14] 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).[15]

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-

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

[15] 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.

quarter level.[16] 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 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 two indicators of judicial efficiency: the time from judgment to online publication and the time from case acceptance to judgment. Together, these measures capture different dimensions of administrative processing efficiency. 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 Appendix.

3.3 Summary Statistics

We merge court performance data with procurement information based on unique court identifiers at the quarterly level, yielding a matched dataset covering 1,511 courts from 2014 to 2022. The primary reason for incomplete coverage is

[16] 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 judgement delay and publication delay.

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 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. To obtain more reliable measures of appeal rates, reversal rates, and the number of judges and clerks at the court–year–quarter level, we exclude court–year–quarters with very few cases—specifically those below the 5th percentile of the case count distribution (fewer than 34 cases).[17]

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 9.3% of first-instance cases in our sample. 1.7% of first-instance judicial decisions are modified by second-instance courts after appeals, with 0.99% experiencing strict reversal (complete overturning) and 0.69% experiencing non-strict reversal (selective modification). Administrative efficiency varies considerably, with an average duration of 92 days between receiving the case and make a judicial decision, and an average duration of 145 days between judgment rendering and online publication. Regarding AI adoption, approximately 1405 (2.8%) of court-quarters, which is consisted by 180 courts (about 12% of 1511 courts), in our sample exhibit evidence of AI technology procurement.

[17] Our results remain robust if we do not add this restriction.

3.4 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 all AI technologies. Courts may adopt various non-AI digital technologies, such as document digitization, big data platforms, and online meeting equipment. To further examine the additional effects of AI relative to other digital technologies, we construct a measure of judicial digital procurement following the same procedure used to identify AI adoption, and we control for digital technology adoption in our baseline specification. This approach allows us to isolate and capture the incremental effect of AI technologies beyond that of 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; Goodman-Bacon, 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, the onboarding of courts that procurement 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).

4. Main Results

4.1 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 1.4 percentage points, equivalent to approximately 15.1% of the sample average (9.3%). The effect on reversal outcomes is even larger: the column (4) estimate suggests that AI procurement decreases the reversal rate by 0.366 percentage points, representing about 21.2% of the mean reversal rate (1.724%). 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 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 approximately four quarters after adoption, with 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 four quarters after adoption and reaches its peak around that time, with a modest attenuation in subsequent periods.

4.2 Robustness Checks

In the baseline analysis, we classify a procurement contract as AI-related whenever the LLM (GPT-4.1) identifies both the contract and its subject matter as pertaining to AI technology. A potential concern is that some AI-related procurements may involve only small expenditures and thus may not meaningfully influence judicial operations. To address this issue and assess the robustness of our findings, we impose alternative monetary thresholds and reclassify a procurement as AI adoption only if its contract amount exceeds 50,000 CNY or 100,000 CNY.[18]

Table 3 reports the results. Using our most stringent specification as the benchmark, we find that imposing higher thresholds yields estimates that closely mirror the baseline for both outcome variables. In fact, the estimated coefficients become slightly larger when we restrict AI procurement to contracts above 100,000 CNY, suggesting that the baseline effects are not driven by small or negligible procurements.

Another potential concern arises from the judicial recentralization reform implemented during our sample period, which sought to strengthen judicial independence by reducing the influence of local governments on local courts (Liu et al., 2022; Lei & Li, 2025). The reform transferred financial and personnel authority from prefectural and county governments to provincial-level authorities. Liu et al. (2022) provide evidence that this reform improved judicial decision quality. A concern, therefore, is that AI adoption may coincide with the reform, confounding our estimates. To address this issue, we construct an indicator capturing the timing of the reform and incorporate it into our baseline specifications.

[18] Condition on having AI procurement, 50,000 CNY and 100,000 CNY correspond to the 6th and 8th percentiles, respectively.

Table 4 presents the results. After controlling for the judicial recentralization reform, AI procurement continues to exhibit a statistically significant effect on both the appeal rate and the reversal rate. Although the estimated coefficients are slightly smaller than in the baseline specification, the effects remain sizable, indicating that our findings are not driven by the reform. Consistent with Liu et al. (2022), we also find that the reform itself leads to reductions in both the appeal rate and the reversal rate, further validating the reform's documented improvements in judicial decision quality.

Having shown that our baseline estimates are robust to alternative definitions of AI procurement and to controlling for the judicial recentralization reform, we next examine whether the effects we document are simply capturing the broader "smart court" reform rather than the specific contribution of AI.

"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 5 reports the results. After controlling for the procurement related to "smart court", AI procurement continues to exhibit a statistically significant effect on both the appeal rate and the reversal rate. By contrast, the estimated effects of

"smart court" procurement itself are negative but small and statistically insignificant. These findings suggest that within the broader "smart court" agenda, AI technology is the key component associated with improvements in the quality of judicial decisions.

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—appeal rate and reversal rate—in 2022.

In addition, the construction of both appeal and reversal rates relies on the availability of second-instance decisions, which inherently lag behind the publication of first-instance judgments. For civil cases, for example, parties have ten to fifteen days to file an appeal or a procuratorate protest after receiving the first-instance ruling, and appellate courts are required to conclude the case within three months of docketing. These procedural lags may further affect the accuracy of measured appeal and reversal rates in later years, particularly when publication volumes fluctuate substantially.

To address these concerns, we exclude the year 2022 from the sample. Table A6 in the 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.

4.3 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.[19]

To address this concern, we constructing an IV by exploiting two sources of variation: (i) pre-determined shares of law cases in each court, which represents the relative importance of the court in a specific province, and (ii) AI by other courts within the same province (excluding the focal court). Specifically, the predicted AI adoption is calculated as:

$$Z_{ct} = \sum_{j \in province} \alpha_{j-c}^{2014} * AI_{j-c,t}, \quad (3)$$

where Z_{ct} is the instrument that predicts the propensity of AI adoption in court c and year-quarter t. The term α_{j-c}^{2014} denotes the share of law cases for court j in the province (where court c is located) in 2014, excluding court c, representing the relative importance of court j in the province that court c locates. The term $AI_{j-c,t}$ is an indicator for whether such court j adopts AI in year-quarter t.

This instrument satisfies the two key conditions for a valid IV. First, the relevance condition is met because courts within the same province face similar budgetary constraints, policy directives, and technological promotion efforts from higher-level judicial authorities. As a result, the average provincial-level AI adoption rate is strongly correlated with each court's own adoption decision. This is confirmed empirically by a strong first stage, with F-statistics well above conventional thresholds. Second, the exclusion restriction requires that aggregate provincial AI adoption affects a focal court's judicial outcomes only through its

[19] In Table 9, we provide evidence that AI adoption substitute court staff.

effect on the court's own AI adoption. The instrument is constructed using pre-determined 2014 case shares and excludes the focal court's own procurement, making it unlikely to be influenced by contemporaneous, court-specific shocks such as leadership changes, judge turnover, or shifts in case composition. Provincial adoption patterns are driven primarily by broader budget allocations, reform tempo, and peer effects across courts—factors that plausibly influence the focal court's outcomes only via their effect on whether the court itself procures AI. Together, these features support the validity of the instrument in isolating exogenous variation in courts' AI adoption.

Table 6 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 including controls. The IV estimates closely align with the baseline in terms of sign and significance but are substantially larger in magnitude for both outcome variables. This pattern suggests that the baseline TWFE estimates may indeed underestimate the true impact of AI adoption on judicial decision quality.

4.4 Comparison with Digital Procurement

Prior to the emergence of modern AI technologies, many courts had already undertaken various forms of digital transformation. 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 7 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. In contrast, the effect

of digital procurement is smaller and becomes only marginally significant once controls are included. 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 does not exhibit a meaningful 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.

4.4 Anther Outcome: Efficiency

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' 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 two measures of court efficiency: (i) publication delay, defined as the number of days between judgment issuance and online publication, and (ii) judge delay, defined as the duration between a court's receipt of a case and the issuance of a judicial decision.

Using the same specifications as in our baseline analysis, Table 8 reports the results. Columns (1) and (2) examine publication delay, while columns (3) and (4) focus on judge delay. Across all specifications, the findings consistently show that AI procurement accelerates both online publication and case processing. The magnitudes are economically meaningful. The estimate in column (2) indicates that

AI adoption reduces the time between judgment rendering and online publication by 33.6 days, equivalent to 23.2% of the mean (145 days). The effect on judge delay, while smaller, remains sizable: column (4) shows that AI procurement shortens the duration from case receipt to judgment by 7.4 days, or 8.0% of the mean. This relative difference in magnitudes is consistent with anecdotal accounts suggesting that AI plays a particularly important role in tasks such as converting courtroom audio to searchable text and preparing materials for online publication.

4.5 The Other Outcome: 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 (Wang 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 use the number of judges and the number of court clerks as dependent variables and examine whether AI adoption leads to a decline in judicial staffing.[20] As shown in Table 9, we find a negative but statistically insignificant effect for judges, and a negative and statistically

[20] Because the coverage of published cases drops substantially in 2022—making our measures of judges and clerks noisier—we exclude 2022 from this analysis.

significant effect, at the 5% level (after adding controls), for court clerks, indicating that AI adoption reduces staffing levels, consistent with a labor-substitution mechanism. In terms of magnitude, the coefficients of -2.686 for court clerks imply that, following AI adoption, the number of court clerks decreases by about 2.686, which is about 6.59% of the mean level. Overall, these findings provide further evidence of AI's labor-substituting effects, especially in a less skilled profession.

5. Heterogeneity Analyses

5.1 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 based on whether the monetary value of an AI procurement is above or below the median among all AI procurements.

Table 10 reports the results. As shown, AI adoption associated with above-median procurement amounts significantly reduces both the appeal rate and the reversal rate. By contrast, AI adoption associated with below-median 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.

5.2 Heterogeneity Analyses by Region

AI adoption may generate heterogeneous effects across regions with different levels of development. In less developed areas, courts typically face shortages of

human capital—judges have, on average, lower educational attainment and weaker professional support—and operate under lower baseline efficiency with more severe resource and institutional constraints. In such environments, AI technologies can substitute for a wide range of routine and procedural tasks, substantially alleviating existing capacity constraints. As a result, the marginal productivity of AI is likely to be higher in low-capacity, low-technology settings, giving rise to stronger improvements in judicial performance relative to more developed regions.

To examine this heterogeneity, we classify courts according to the development level of the regions in which they are located, using three 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 third differentiates coastal from non-coastal cities, where non-coastal areas tend to be less developed than their coastal counterparts.

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 coefficients are statistically insignificant for courts in above-median regions. Specifically, the coefficients for the appeal rate are -0.021 versus -0.008 (below- median versus above-median), and those for the reversal rate are -0.608 versus -0.203, indicating substantial differences in magnitude. 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. Likewise, Panel C of Table 11 reveals larger reductions in appeal and reversal rates for courts in non-coastal cities.

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.

5.3 Heterogeneity Analyses by Case Type

AI adoption may also generate heterogeneous effects across different types of cases. We consider two definitions of case type. First, 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. Because AI can automate routine and time-consuming tasks, thereby allowing judges to allocate more attention to substantive legal reasoning, its benefits are more likely to be pronounced for complex cases, whereas simpler cases can often be handled effectively even without AI. 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 0.016 for complex cases, compared with 0.008 for non-complex cases. Similarly, AI reduces the reversal rate by 0.457 for complex cases and by 0.096 for non-complex cases. These patterns indicate that AI generates larger improvements for complex cases, consistent with the idea that automation of routine tasks frees judicial time and attention for more demanding legal reasoning. 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 for civil cases, whereas we find no clear evidence of performance improvements for criminal or administrative cases. 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.

6. 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., & 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.

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., Yang, D. Y., & Yuchtman, N. (2023). Data-intensive innovation and the state: Evidence from AI firms in China. The Review of Economic Studies, 90(4), 1701–1723.

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

Beraja, M., & Zorzi, N. (2025). Inefficient automation. Review of Economic Studies, 92(1), 69–96.

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, 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.

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. (2023). 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. (2022). Judicial independence, local protectionism, and economic integration: Evidence from China. NBER Working Paper No. 30432.

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, H., Chen, Y., Yang, Q., & Song, Y. (2024). Setting the bench higher: The impact of judge evaluation reform on judicial performance in China. SSRN Working Paper No. 4925678.

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, 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

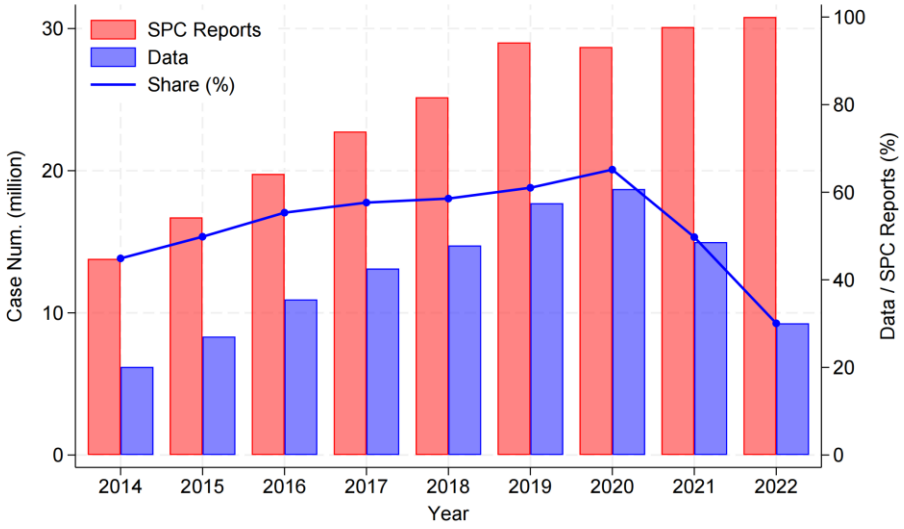


Figure 1. The Trend of the Number of Online Published Cases

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.

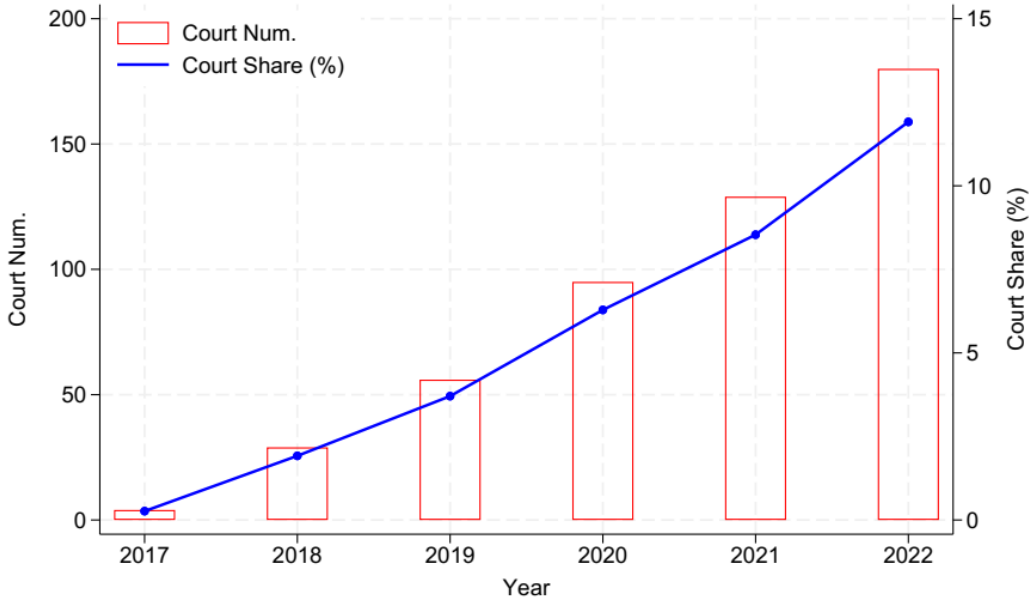


Figure 2. The Trend of AI Adoption

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.

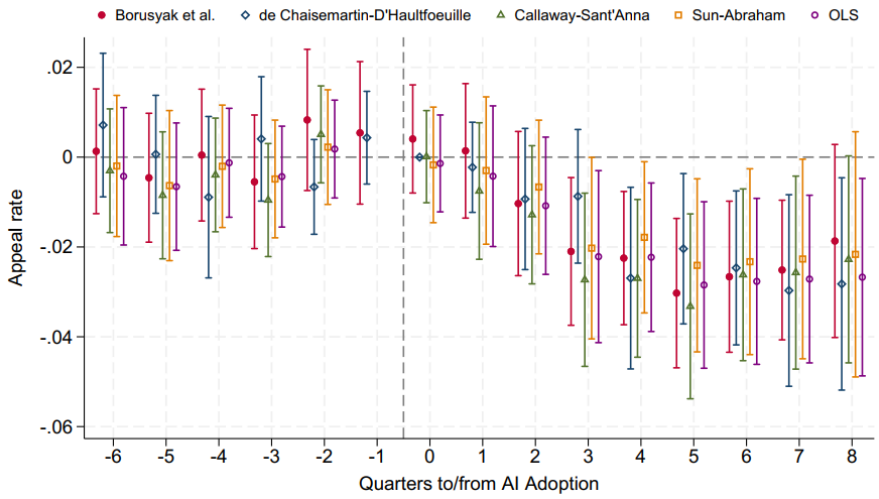


Figure 3(a): Event Study for Appeal Rate

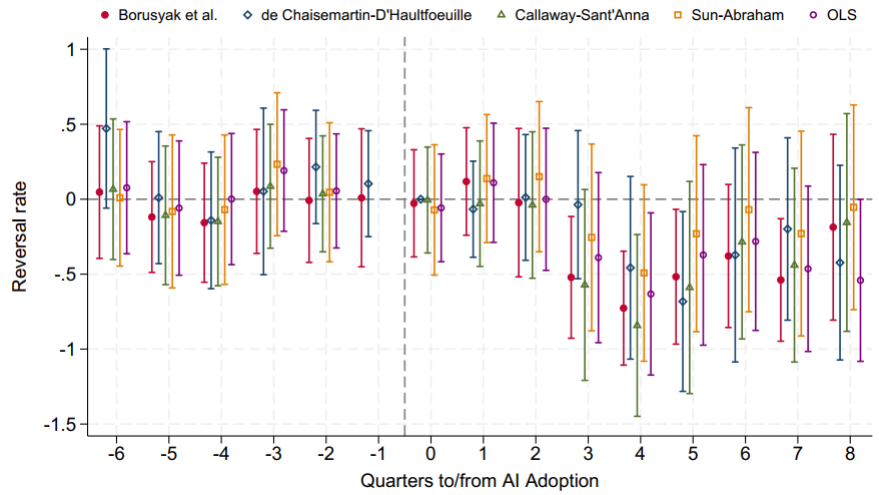


Figure 3(b): Event Study for Reversal Rate

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. (2021) in red dot, de Chaisemartin and D'Haultfoeuille (2020) in blue diamond, San and Abraham (2020) in green triangle, and Callaway and Sant'Anna (2020) in yellow square.

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 Appendix provides the detailed definition of these characteristics.

	N	Mean	p50	p10	p90	SD
AI Procurement	50,060	0.028	0.000	0.000	0.000	0.165
AI procurement (above 50k CNY)	50,060	0.027	0.000	0.000	0.000	0.163
AI procurement (above 100k CNY)	50,060	0.027	0.000	0.000	0.000	0.162
Smart Court	50,060	0.026	0.000	0.000	0.000	0.159
Digital Procurement	50,060	0.257	0.000	0.000	1.000	0.437
Uniform Reform	50,060	0.467	0.000	0.000	1.000	0.499
Instrumental Variable	48,182	0.037	0.000	0.000	0.167	0.080
Appeal Rate	50,060	0.093	0.061	0.004	0.212	0.109
Reversal Rate (%)	50,060	1.724	0.794	0.000	4.294	2.811
Strict Reversal Rate	50,060	0.989	0.355	0.000	2.586	1.779
Non-strict Reversal Rate	50,060	0.690	0.211	0.000	1.818	1.312
Publication Delay	50,060	145.042	106.824	44.719	271.303	134.318
Judgment Delay	49,714	92.997	82.749	46.714	140.389	55.885
Number of Judges	45,078	46.945	37.000	17.000	88.000	37.010
Number of Court Clerk	45,072	40.731	30.000	13.000	76.000	37.505
GDP Per Capita (in 10k CNY)	50,060	0.471	0.359	0.180	0.864	0.386
Population (in millions)	50,060	0.868	0.470	0.173	1.724	1.312
Public Expenditure Per Capita (in 10k CNY)	50,060	0.079	0.067	0.000	0.152	0.072

Table 2. Baseline Results

This table examines the impact of AI technology on court appeal rates and reversal rates, two outcome-based indices. 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	-0.016***	-0.014**	-0.369**	-0.366**
	(0.006)	(0.006)	(0.146)	(0.145)
Controls	NO	YES	NO	YES
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.095	0.093	1.761	1.724
Observations	51995	50060	51995	50060
Adjusted R-squared	0.478	0.479	0.338	0.338

Table 3. Robustness Test: Restricting High Amount of AI Procurement

This table further restricts the contract value for AI technology procurement and examines the impact of AI technology procurement contracts exceeding RMB 50,000 and RMB 100,000 on court appeal rates and 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.

	Appeal Rate		Reversal Rate (%)	
	(1)	(2)	(3)	(4)
AI procurement (above 50k CNY)	-0.015**		-0.362**	
	(0.006)		(0.147)	
AI procurement (above 100k CNY)		-0.015**		-0.389***
		(0.006)		(0.148)
Controls	YES	YES	YES	YES
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.093	0.093	1.724	1.724
Observations	50,060	50,060	50,060	50,060
Adjusted R-squared	0.479	0.479	0.338	0.338

Table 4. Robustness Test: Controlling for Judicial Independence Reform

This study additionally controls for the reform of centralized management of personnel, finances, and assets in Chinese courts, and investigates the impact of AI technology on court appeal rates and 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.

	Appeal Rate		Reversal Rate (%)	
	(1)	(2)	(3)	(4)
AI procurement	-0.016***	-0.014**	-0.364**	-0.362**
	(0.006)	(0.006)	(0.145)	(0.145)
Uniform reform	-0.018***	-0.016***	-0.330***	-0.286***
	(0.004)	(0.004)	(0.101)	(0.103)
Controls	NO	YES	NO	YES
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.095	0.093	1.761	1.724
Observations	51,995	50,060	51,995	50,060
Adjusted R-squared	0.479	0.480	0.338	0.339

Table 5. Robustness Test: Controlling for Smart Court

This table additionally controls for the court's "Smart Court" procurement projects and examines the impact of AI technology on court appeal rates and reversal rates after eliminating the interference from "Smart Court" procurement. 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	-0.015**	-0.014**	-0.336**	-0.334**
	(0.006)	(0.006)	(0.146)	(0.146)
Smart court	-0.005	-0.005	-0.173	-0.171
	(0.006)	(0.006)	(0.170)	(0.164)
Controls	NO	YES	NO	YES
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.093	0.095	1.724	1.761
Observations	50,060	50,060	50,060	50,060
Adjusted R-squared	0.479	0.479	0.338	0.339

Table 6. 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. 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	-0.057**	-0.055**	-1.340**	-1.452**
	(0.025)	(0.026)	(0.669)	(0.671)
Controls	NO	YES	NO	YES
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.095	0.093	1.761	1.724
Observations	50,050	48,182	50,050	48,182
First Stage Coefficient	0.550***	0.547***	0.550***	0.547***
	(0.065)	(0.065)	(0.065)	(0.065)
KP F Stat.	70.51	70.77	70.51	70.77

Table 7. 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. 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	-0.015**	-0.012*	-0.380**	-0.369**
	(0.006)	(0.006)	(0.148)	(0.147)
Digital procurement	-0.004	-0.005*	0.027	0.008
	(0.003)	(0.003)	(0.069)	(0.068)
Controls	NO	YES	NO	YES
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.095	0.093	1.761	1.724
Observations	51,995	50,060	51,995	50,060
Adjusted R-squared	0.478	0.479	0.338	0.338

Table 8. AI Impacts on Courts' Efficiency

This table examines the impact of AI adoption on two measures of case-processing efficiency: publication delay and judgment delay. 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.

	Publication Delay (days)		Judgment Delay (days)	
	(1)	(2)	(3)	(4)
AI procurement	-36.175***	-33.565***	-7.573***	-7.390***
	(12.340)	(12.428)	(2.567)	(2.615)
Controls	NO	YES	NO	YES
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	145.123	145.042	93.474	92.997
Observations	52,177	50,231	51,818	49,875
Adjusted R-squared	0.406	0.412	0.231	0.231

Table 9. 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. Because the coverage of published cases drops substantially in 2022—making our measures of judges and clerks noisier—we exclude 2022 from this analysis. 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		Number of Court Clerks	
	(1)	(2)	(3)	(4)
AI procurement	-0.867	-1.140	-2.744*	-2.686**
	(1.841)	(1.754)	(1.584)	(1.306)
Controls	NO	YES	NO	YES
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	48.449	46.945	41.819	40.731
Observations	46,917	45,078	46,911	45,072
Adjusted R-squared	0.902	0.903	0.919	0.918

Table 10. Heterogeneity: Procurement Amount

This table examines the heterogeneity in the impact of AI adoption on court performance across different procurement amounts. Using the median AI procurement amount across all courts, we construct two indicators distinguishing whether a court's first AI procurement exceeds or falls below the median. 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 (Amount	-0.019**	-0.017*	-0.522**	-0.510**
Above Median)	(0.009)	(0.009)	(0.216)	(0.214)
AI procurement (Amount	-0.008	-0.007	-0.131	-0.141
Below Median)	(0.008)	(0.008)	(0.182)	(0.184)
Controls	NO	YES	NO	YES
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.093	0.093	1.724	1.724
Observations	51,995	50,060	51,995	50,060
Adjusted R-squared	0.478	0.479	0.338	0.339

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, whether it is located in western China, and whether it is located in a coastal area. 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)	(5)	(6)	(7)	(8)
AI procurement	-0.024***	-0.021**	-0.010	-0.008	-0.646***	-0.608***	-0.186	-0.203
	(0.009)	(0.009)	(0.009)	(0.008)	(0.220)	(0.212)	(0.194)	(0.192)
Controls	NO	YES	NO	YES	NO	YES	NO	YES
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	0.099	0.099	0.093	0.089	1.998	1.992	1.721	1.629
Observations	25,864	25,617	25,187	24,443	25,864	25,617	25,187	24,443
Adjusted R-squared	0.467	0.462	0.488	0.492	0.341	0.336	0.337	0.342
	PANEL B: Western & Non-western							
	Appeal Rate				Reversal Rate (%)			
	Western		Non-western		Western		Non-western	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AI procurement	-0.024**	-0.022**	-0.011	-0.010	-0.500**	-0.504**	-0.310*	-0.322*
	(0.011)	(0.011)	(0.007)	(0.007)	(0.246)	(0.244)	(0.177)	(0.173)
Controls	NO	YES	NO	YES	NO	YES	NO	YES
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	0.099	0.097	0.094	0.092	2.090	2.041	3.196	3.115
Observations	22,553	21,931	29,442	28,129	22,553	21,931	29,442	28,129
Adjusted R-squared	0.494	0.496	0.464	0.467	0.341	0.338	0.332	0.339
	PANEL C: Coastal & Non-coastal							
	Appeal Rate				Reversal Rate (%)			
	Non-coastal		Coastal		Non-coastal		Coastal	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AI procurement	-0.020**	-0.018**	-0.011	-0.012	-0.404**	-0.425**	-0.393*	-0.418**
	(0.009)	(0.009)	(0.008)	(0.008)	(0.204)	(0.205)	(0.205)	(0.198)
Controls	NO	YES	NO	YES	NO	YES	NO	YES
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	0.096	0.094	0.096	0.094	1.965	1.911	1.691	1.659
Observations	31,394	30,564	20,601	19,496	31,394	30,564	20,601	19,496
Adjusted R-squared	0.502	0.502	0.436	0.445	0.346	0.344	0.319	0.33

Table 12. Heterogeneity: 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 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)	(5)	(6)	(7)	(8)
AI procurement	-0.018**	-0.016**	-0.009	-0.008	-0.408**	-0.457**	0.035	0.096
	(0.007)	(0.007)	(0.007)	(0.007)	(0.181)	(0.178)	(0.239)	(0.238)
Controls	NO	YES	NO	YES	NO	YES	NO	YES
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	0.127	0.124	0.045	0.044	0.024	0.023	0.010	0.010
Observations	51,833	49,907	51,323	49,436	51,833	49,907	51,323	49,436
Adjusted R-squared	0.437	0.436	0.354	0.356	0.242	0.247	0.155	0.157

Table 13. Heterogeneity: 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. 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: Civil Cases			
	Appeal Rate		Reversal Rate (%)	
	(1)	(2)	(3)	(4)
AI procurement	-0.017***	-0.017***	-0.549**	-0.534**
	(0.006)	(0.006)	(0.239)	(0.239)
Controls	NO	YES	NO	YES
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.094	0.092	0.021	0.021
Observations	51,868	49,947	51,868	49,947
Adjusted R-squared	0.411	0.416	0.256	0.259
	PANEL B: Criminal Cases			
	Appeal Rate		Reversal Rate (%)	
	(1)	(2)	(3)	(4)
AI procurement	-0.013	-0.009	-0.303	-0.244
	(0.010)	(0.011)	(0.236)	(0.249)
Controls	NO	YES	NO	YES
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.100	0.098	0.014	0.013
Observations	47,507	45,824	47,507	45,824
Adjusted R-squared	0.373	0.376	0.155	0.16
	PANEL C: Administrative Cases			
	Appeal Rate		Reversal Rate (%)	
	(1)	(2)	(3)	(4)
AI procurement	-0.042**	-0.034	-0.980	-0.915
	(0.021)	(0.023)	(0.612)	(0.629)
Controls	NO	YES	NO	YES

Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.290	0.291	0.038	0.038
Observations	28,156	26,746	28,156	26,746
Adjusted R-squared	0.238	0.238	0.073	0.073

APPENDIX: Additional Figures and Tables



类案检索智能辅助指导服务

2022-12-21 公告来源：中国政府采购网【打印】

一、合同编号：[redacted]

二、合同名称：类案检索智能辅助指导服务

Contract Name: Intelligent guidance service for similar case retrieval

三、项目编号：[redacted]

四、项目名称：类案检索智能辅助指导服务

五、合同主体

采购人（甲方）：齐齐哈尔市中级人民法院

Purchaser

地 址：[redacted]

联系方式：[redacted]

供应商（乙方）：聚法科技（长春）有限公司

Suppiler

地 址：[redacted]

联系方式：[redacted]

六、合同主要信息

主要标的名称：类案检索智能辅助服务

Main Content: Intelligent retrieval and auxiliary service for similar cases

规格型号（或服务要求）：详见合同

主要标的数量：1.0000套

主要标的单价：396889.0000元

合同金额：39.688900万元

Contract Amount

履约期限、地点等简要信息：[redacted]

采购方式：竞争性磋商

七、合同签订日期：2022-12-16

Sign Date

八、合同公告日期：2022-12-21

htgs.ccgp.gov.cn/GS8/contractpublish/detail/2c8382aa848afb0b018533263d1217c5?contractSign=0 1/2

2025/11/13 20:24 类案检索智能辅助指导服务

九、其他补充事宜：

附件：

免责声明：本页面提供的政府采购合同是按照《中华人民共和国政府采购法实施条例》的要求由采购人发布的，中国政府采购网对其内容概不负责，亦不承担任何法律责任。

Figure A1. 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.



阿克苏惠民农资有限责任公司与阿拉尔豫娥农资有限公司买卖合同纠纷二审民事判决书

Case code

案　由 买卖合同纠纷　点击了解更多　　案　号 （2019）新29民终1299号　指导案例　推荐案例

发布日期 2020-01-06　Publication date　浏览次数 114

新疆维吾尔自治区阿克苏地区中级人民法院

民事判决书

Plaintiff

（2019）新29民终1299号

上诉人（原审被告）：阿克苏惠民农资有限责任公司，住所地新疆维吾尔自治区阿克苏市交通路25号金土地-03商铺。

法定代表人：张毅，该公司经理。

委托诉讼代理人：冯国盛，新疆白水河律师事务所律师。

Defendant

被上诉人（原审原告）：阿拉尔豫娥农资有限公司，住所地新疆维吾尔自治区阿拉尔市幸福农场绿园小区沿街商铺门面3-6号。

法定代表人：吴春山，该公司总经理。

委托诉讼代理人：梁洪才，新疆制衡律师事务所律师。

上诉人阿克苏惠民农资有限责任公司（以下简称惠民公司）与阿拉尔豫娥农资有限公司（以下简称豫娥公司）买卖合同纠纷一案，不服阿克苏市人民法院（2019）新2901民初552号民事判决书，向本院提起上诉。本院于2019年10月18日立案后，依法组成合议庭，于2019年11月6日公开开庭进行了审理。上诉人惠民公司委托诉讼代理人冯国盛、被上诉人豫娥公司法定代表人吴春山及委托诉讼代理人梁洪才到庭参加诉讼。本案现已审理终结。

惠民公司上诉请求：撤销阿克苏市人民法院（2019）新2901民初552号民事判决书，改判驳回豫娥公司的诉讼请求。事实及理由：1.豫娥公司从惠民公司购买了61吨32%智诺富钙。惠民公司分批次向豫娥公司供应完毕，后豫娥公司将化肥出售给张某某，张某某使用该化肥后称化肥质量不合格造成棉花地出苗率低。惠民公司委托新疆臻冠达农业科技有限责任公司司法鉴定所对张某某出苗率低的原因进行鉴定。鉴定结论为张某某的棉花地出苗率低是因为土壤偏碱性造成。故惠民公司出售的化肥质量没问题。阿克苏地区产品质量检验所的报告取样不全面，仅是对剩余的2吨化肥进行取样，因此61吨化肥是分批次向豫娥公司购买货，故报告不能证明剩余的59吨化肥不合格。2.豫娥公司无证据证明其所遭受的损失及具体金额，豫娥公司将从惠民公司处购买的化肥又转卖给张军伟，豫娥公司已收取了化肥款并获得了利益，豫娥公司没有遭受损失，因此不能作为产品质量不合格的受害主体向惠民公司主张权利，其诉讼主体不适格。

豫娥公司辩称，1.惠民公司销售给豫娥公司的化肥经阿克苏地区产品质量检验所鉴定为不合格，根据产品责任法的规定，产品质量应当检验合格，不得以不合格产品冒充合格产品，本案中惠民公司违反法律的行为致使本案的买卖合同无效，故依法应当予以解除。合同无效或者被撤销后，因该合同取得的财产，应当予以返还。2.产品质量和产品损害是两个不同的法律概念，豫娥公司是基于买卖合同主张惠民公司返还化肥款，案外人棉花出苗率低和化肥质量问题的法律概念、责任、权利、义务均不同。3、惠民公司出售给豫娥公司的化肥均是同一批次，实践中有关部门对产品的质量监督是抽样鉴定而非全部鉴定，如惠民公司认为剩余59吨化肥质量合格，其应当举证证明。综上，一审判决正确，请求予以维持。

豫娥公司向一审法院提出诉讼请求：判令惠民公司返还肥料款89,670元。

原审法院认定事实如下：2018年4月13日，豫娥公司从惠民公司处购买32%智诺富钙61吨，单价1,470元，合计金额为89,670元。后豫娥公司将从惠民公司处购买的肥料进行了销售。2018年5月11日，阿克苏市质量技术监督局委托阿克苏地区产品质量检验所对豫娥公司从惠民公司处购买的智诺富钙进行质量鉴定。阿克苏地区产品质量检验所进行抽检时，豫娥公司、惠民公司双方均派人参加。2018年5月24日，阿克苏地区产品质量检验所作出检验报告，结论为：该样品经检验，总磷、有效磷、游离酸不符合Q / AYK01-2015标准（颗粒型），检验结论为不合格。后豫娥公司多要求惠民公司退还货款无果，豫娥公司遂诉至法院。一审法院认为，售出的产品不符合以产品说明、实物样品等方式表明的质量状况的，销售者应当负责更换、退货；出卖人交付的标的物不符合质量要求的，买受人可以合理选择要求对方承担更换、退货、减少价款或者报酬等违约责任。本案惠民公司销售给豫娥公司的化肥经阿克苏地区产品质量检验所检验，其销售的产品质量不合格，豫娥公司要求惠民公司退还货款的诉讼请求符合法律规定，一审法院予以支持，惠民公司的辩称意见不符合法律规定，不予采纳。依照《中华人民共和国产品质量法》第四十条、《中华人民共和国合同法》第一百一十一条、第一百三十条、第一百五十五条之规定，判决：惠民公司退还豫娥公司货款89,670元，此款于本判决生效后十日内付清。The verdict of the first instance

二审期间，双方当事人无新证据提交。

本院对一审查明的事实依法予以确认。Reason

本院认为，根据本案当事人的上诉理由及答辩意见，本案二审的争议焦点为：1.本案中豫娥公司是否属适格主体；2.惠民公司出售的61吨32%智诺富钙化肥产品质量是否合格，是否应当返还货款。本案系产品责任侵权法律关系和买卖合同法律关系竞合，豫娥公司基于双方之间买卖合同关系，起诉要求惠民公司承担合同无效的法律责任，故本案案由应为买卖合同纠纷，豫娥公司系合同相对方，属适格主体。2.产品质量应当检验合格，不得以不合格产品冒充合格产品，惠民公司销售给豫娥公司的化肥经阿克苏地区产品质量检验所检验，其销售的产品质量不合格，量仅对剩余2吨化肥进行鉴定，但因59吨化肥已经使用，并未有留存，故而无法对59吨化肥进行抽样鉴定，故对惠民公司主张阿克苏地区产品质量检验所鉴定因取样不全面不应采信的上诉意见不予采纳。根据法律规定,售出的产品不符合以产品说明、实物样品等方式表明的质量状况的，销售者应当负责更换、退货；出卖人交付的标的物不符合质量要求的，买受人可以合理选择要求对方承担更换、退货、减少价款或者报酬等违约责任。本案惠民公司销售给豫娥公司的化肥经阿克苏地区产品质量检验所检验，该化肥五项指标中总磷、有效磷、游离酸均低于合格标准，故豫娥公司应当对其销售不合格化肥给豫娥公司造成的损失承担违约责任。因化肥已经使用59吨，更换、退货均无法实现，故豫娥公司应当根据其违约责任大小承担减少价款的责任。虽惠民公司销售的32%智诺富钙（钙化肥不合格，但豫娥公司未能就造成其损失及因损失向张某某予以赔偿进行举证，故其要求惠民公司退还全部货款依据不足。基于惠民公司提供的化肥、版本脱胎于分含量不达标的化肥，版本脱胎定由惠民公司承担60%的责任，豫娥公司自行承担40%的责任。

综上所述，惠民公司的上诉请求部分成立，本院予以支持，依照《中华人民共和国产品质量法》第四十条、《中华人民共和国合同法》第一百一十一条、第一百三十条、第一百五十五条、《中华人民共和国民事诉讼法》第一百七十条第一款第二项规定，判决如下

一、撤销阿克苏市人民法院（2019）新2901民初552号民事判决书；

二、阿克苏惠民农资有限责任公司退还阿拉尔豫娥农资有限公司货款53,802元[89,670元×60%]；

The verdict of the second instance

如果未按本判决指定的期间履行给付金钱义务，应当依照《中华人民共和国民事诉讼法》第二百五十三条之规定，加倍支付迟延履行期间的债务利息。

一审案件受理费2,042元（豫娥公司预交），减半收取1,021元，由阿克苏惠民农资有限责任公司负担612.6元，阿拉尔豫娥农资有限公司负担408.4元。二审案件受理费2,042元（惠民公司预交），由阿克苏惠民农资有限责任公司负担1,225.2元，阿拉尔豫娥农资有限公司负担816.8元。

本判决为终审判决。

Name of the judge　审判长 李伟力　审判员 高　静　审判员 赵培强

Judgment date　二〇一九年十二月三十日

Name of the court clerk　书记员 肖　洁

Figure A2. 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.

Table A1. LLM Prompt to Identify AI Procurements

This table presents the prompt we input to the LLM. We use a fixed version of ChatGPT 4.1 (gpt-4.1-2025-04-14), and we set the temperature at 0.1 (a low level) to enhance the replicability of our outcomes.

model="gpt-4.1-2025-04-14",
instructions="Suppose you are an expert in reviewing Chinese government procurement contracts",
input="Please read the purchase item "{*INPUT TEXT*}" carefully and judge: 1. whether the project involves AI technology or AI equipment, if yes, answer 1, if no or if unable to judge, answer 0; 2. whether the project involves digitization, information technology or digital equipment. If yes, answer Y, if no or unable to judge, answer N. Please only answer [1,Y] or [1,N] or [0,Y] or [0,N], without explanation. You must answer. I have no specific information",
stream=False, temperature=0.1, 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 Adoption					
	(1)	(2)	(3)	(4)	(5)	(6)
	Previous 1 year		Previous 2 year		Previous 3 year	
Appeal Rate	-0.029	-0.028	-0.020	-0.021	-0.029	-0.031
	(0.019)	(0.022)	(0.019)	(0.023)	(0.020)	(0.023)
Reversal Rate	-0.000	-0.001	-0.000	-0.001	0.000	-0.000
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)
Public Expenditure Per Capita		-0.069		-0.066		-0.052
		(0.106)		(0.109)		(0.105)
Population Growth Rate		0.001		0.002		0.003
		(0.008)		(0.008)		(0.007)
GDP Growth Rate		0.006		0.004		0.001
		(0.008)		(0.009)		(0.010)
GDP Per Capita		0.035*		0.036*		0.035*
		(0.019)		(0.020)		(0.020)
Court FE	YES	YES	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES	YES	YES
Mean of outcome	0.028	0.028	0.028	0.028	0.028	0.028
Observations	50,028	46,111	48,622	44,761	47,201	43,428
Adjusted R-squared	0.376	0.384	0.387	0.394	0.398	0.403

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.
AI procurement (above 50k CNY)	Binary variable which takes 1 since the court signed its first AI procurement contract (amount above 50k CNY) and 0 else.
AI procurement (above 100k CNY)	Binary variable which takes 1 since the court signed its first AI procurement contract (amount above 100k CNY) 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.
Uniform Reform	Binary variable which takes 1 since the court underwent the reform of unified management of personnel, finances and property, and 0 else.
Instrumental Variable	$\sum_{j \in province} \alpha_{j-c}^{2014} * AI_{j-c,t}$. The term α_{j-c}^{2014} denotes the share of the number of law cases for court j in the province (where court c is located) in 2014, excluding court c. The term $AI_{j-c,t}$ is an indicator for whether such court j adopts AI in year-quarter t.
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
Publication Delay (days)	The time interval from the judgment to the online publication of the document in the case.
Judgment Delay (days)	The time interval from case acceptance to judgment.
Number of Judges	Number of judicial personnel in the court.
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	0.154	0.151	0.003	0.579
Strict Reversal Rate (%)	1.525	1.545	-0.020	0.818
Non-strict Reversal Rate (%)	1.016	1.034	-0.018	0.765
Reversal Rate (%)	2.642	2.655	-0.013	0.920
GDP Per Capita	0.372	0.407	-0.0340	0.015**
Population	0.925	1.013	-0.0890	0.155
Public Expenditure Per Capita	0.0570	0.0630	-0.00600	0.016**

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	-0.198**	-0.182**	-0.173***	-0.175***
	(0.088)	(0.088)	(0.065)	(0.063)
Controls	NO	YES	NO	YES
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	1.085	0.989	0.771	0.69
Observations	51,995	50,060	51,995	50,060
Adjusted R-squared	0.269	0.268	0.247	0.253

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	-0.020***	-0.017**	-0.358**	-0.342**
	(0.007)	(0.007)	(0.170)	(0.169)
Controls	NO	YES	NO	YES
Court FE	YES	YES	YES	YES
Year by quarter FE	YES	YES	YES	YES
Mean of outcome	0.103	0.101	1.936	1.897
Observations	46,870	45,036	46,870	45,036
Adjusted R-squared	0.472	0.475	0.337	0.340