Smarter Courts, Easier Life for Judges? Evidence from China

Hai Ding Jiaqi Liu Li-An Zhou Yuling Zhu *

Abstract: This paper examines whether artificial intelligence (AI) alleviates capacity constraints in the public sector or instead induces additional demand that offsets efficiency gains. We study the staggered adoption of intelligent voice transcription (IVT) technology across courts in China, which automates trial recording and reduces reliance on clerical labor. Using comprehensive administrative data, we show that AI adoption significantly improves trial efficiency. These gains arise from more streamlined in-court proceedings and improved allocation of clerical resources. However, these efficiency gains do not reduce judicial workload. Courts respond by lowering filing thresholds and processing more cases, and over time this demand response offsets the initial gains, leading to a new equilibrium with higher caseloads but limited long-run reductions in adjudication duration. Our findings highlight a demand-induced mechanism through which technological progress in capacity-constrained public systems may not translate into sustained reductions in workload.

Keywords: AI-Assisted Courts, Trial Efficiency, Judicial Effectiveness

JEL Classification: J22, K41, P35

*Ding: Guanghua School of Management, Peking University (Email: dinghai2025@126.com); Liu: PBC School of Finance, Tsinghua University (Email: jqliu@mail.tsinghua.edu.cn); Lou; Zhou: Guanghua School of Management, Peking University (Email: zhoula@gsm.pku.edu.cn); Zhu, School of Public Administration and Policy, Shanghai University of Finance and Economics (Email: zhuyuling@sufe.edu.cn). We appreciate the helpful comments from... The authors are listed in alphabetical order and contribute equally to the article. .. acknowledges the support of the National Natural Science Foundation of China.

1 Introduction

Judicial systems around the world operate under persistent capacity constraints. In many settings, courts face a chronic imbalance between the inflow of cases and available adjudicative resources, resulting in delays, backlogs, and limited access to justice. While traditional reforms—such as increasing judicial staffing or simplifying procedures—have had only limited success, recent advances in artificial intelligence (AI) have generated growing optimism about the potential to transform public sector productivity. By automating routine tasks and reducing administrative frictions, AI technologies promise to significantly expand effective capacity in judicial processes.

Yet a fundamental question remains unresolved: does technological progress actually relieve workload pressures in capacity-constrained public systems, or does it instead induce additional demand that absorbs these gains? When unmet demand is substantial, improvements in processing capacity may lead to endogenous expansions in usage—for example, by lowering barriers to entry or increasing the willingness of individuals to seek formal dispute resolution. In such settings, the ultimate effect of AI is not obvious ex ante: efficiency gains may translate into faster processing, or they may instead result in a higher volume of cases with little change in equilibrium congestion. Despite its importance, there is limited causal evidence on how AI affects not only efficiency but also equilibrium outcomes in public service provision.

This paper studies this question in the context of China's judicial system, which has experienced rapid adoption of AI-assisted technologies in recent years. We exploit the staggered rollout of intelligent voice transcription (IVT) systems across courts, which automate trial recording and, in later stages, partially substitute for clerical labor. Combining detailed data on judicial cases, trial proceedings, and technology deployment, we implement a difference-in-differences design to estimate the causal effects of AI adoption on court performance.

This paper exploits the staggered adoption of the Intelligent Voice Transcription (IVT) reform and the subsequent Trial Without Clerk (TWC) reform across courts in China to quantify the effects of AI technologies in judicial settings, an area that remains relatively understudied. Using large-scale judicial data—including judgment documents, trial livestream records, and internal information on AI system deployment—we examine the impact of AI-assisted reforms on trial efficiency and court staff workload. We find that these reforms significantly improve trial efficiency. Specifically, the adoption of intelligent voice transcription reduces adjudication duration by approximately 12.10%, while the second-stage reform leads to an additional reduction of about 10.77%. We further identify two key mechanisms underlying these effects: optimizing court proceedings and optimizing time allocation across cases. The former is reflected in reductions in both hearing duration and the number of hearings,

while the latter is evidenced by spillover effects on cases under the ordinary procedure and on the preparation of verdict documents.

These empirical findings raise three additional questions: Does the reform enhance access to justice? Are the efficiency gains achieved at the cost of increased individual workload? And do they come at the expense of adjudication quality? We conduct further analyses to assess these potential trade-offs. First, using the ratio of rejected filings as a proxy for filing thresholds, we find that the reform lowers entry barriers and improves access to justice. Second, although the number of cases handled per capita increases significantly after the reform, there is no evidence of a rise in extreme overtime work. Third, we examine adjudication quality from two perspectives: documentation quality and judgment quality. For documentation quality, using measures such as the length of litigation records, the number of cited legal articles, and generative-AI-based evaluation scores, we find that the reform leads to more complete and accurate real-time trial records. For judgment quality, using appeal rates and the probability of reversal on appeal as proxies, we find no significant impact.

We move beyond average efficiency effects to examine whether the reform genuinely alleviates the burden on judges and clerks over time, or instead induces work intensification through expanded case processing. The dynamic evidence suggests that AI-assisted technologies improve productivity in the short run. However, in judicial systems characterized by rising case inflows and binding capacity constraints, these gains are endogenously absorbed by increased caseloads, leading to a busier equilibrium. As a result, short-term improvements do not translate into sustained long-term reductions in adjudication duration or persistent improvements in quality.

Our paper extends the literature on artificial intelligence (AI) applications in the judicial sector. With the rapid advancement and increasing adoption of AI and machine learning (ML) techniques, a growing body of economic research has begun exploring their applications in judicial processes. For instance, studies have examined AI-assisted adjudication (Kleinberg et al., 2018; Ludwig & Mullainathan, 2024; Nissan, 2017) and algorithm-driven predictive policing (Jabri, 2021; Mastrobuoni, 2020). While these studies primarily focus on the direct impact of technology on operational efficiency and service delivery, they have paid less attention to how technology adoption influences the behavior of its users. Our paper, by examining the underlying behavioral mechanisms, offers a systematic analysis of how judicial personnel adapt to the changes induced by new technologies under "demand pressure".

Our paper contributes to the burgeoning literature on the impact of artificial intelligence (AI) on labor productivity. Recent studies have established that generative AI can significantly boost task-level efficiency—often acting as an equalizing technology that benefits low-skill workers in professional writing (Noy & Zhang, 2023), customer support (Brynjolfsson et al., 2025), and creative tasks (Chen &

Chan, 2024). Emerging evidence suggests a persistent "rebound" effect where efficiency gains are offset by expanded labor supply: at the intensive margin, AI exposure has been linked to an "extended workday" paradox as firms leverage heightened monitoring to extract more effort (Jiang et al., 2025); at the extensive margin, AI lowers the shadow cost of entry, triggering a demand-side surge in institutional access (Shah & Levy, 2026). Our study bridges these disparate strands by shifting the focus from private agent optimization to an endogenous supply-side mechanism within capacity-constrained public systems, and identify a mechanism where the institution itself leverages the capacity liberated from clerical tasks to lower filing thresholds and process more cases. By documenting this institutional manifestation of the "Jevons Paradox" (Jevons, 1865), we provide a unique contribution in showing how technological progress in the public sector may lead to a new equilibrium of higher throughput but stagnant adjudication duration.

Our paper also adds to the literature on how to improve judicial efficiency thus mitigating case overloads in courts. Excessive caseloads relative to judicial capacity create persistent backlogs, long adjudication delays, and inefficient allocation of resources, particularly in developing countries (Bosio, 2023; Dakolias, 1999).[1] Prior studies have evaluated institutional reforms such as delegating tasks to clerks and judicial assistants, limiting oral arguments, expanding alternative dispute resolution (Robel, 1990), increasing the number of judges, raising judicial compensation, and establishing judicial councils (Assegaf, 2019; Bełdowski et al., 2020; Blank & van Heezik, 2020; Dias & Gomes, 2018; El Bialy, 2016; Nanwani, 2016; Rosenn, 2000). Evidence suggests that these measures have achieved limited success in reducing case duration or clearing backlogs. By contrast, technological adoption offers a potentially more effective solution, yet systematic evidence on its impact remains scarce. Our study addresses this gap by evaluating how advanced technologies can improve judicial processing and enhance court capacity. More broadly, the paper contributes to the literature on technology in the public sector, which shows that digital tools can strengthen state capacity, improve project implementation (Banerjee et al., 2023; Muralidharan et al., 2016), deter fraud (Assunção et al., 2023; Doleac, 2017; Guriev et al., 2021; Jin & Lee, 2014), and enhance fiscal and administrative efficiency, especially in developing countries (Cohen, 2024; Dzansi et al., 2022; Greenstone et al., 2022).

Moreover, our paper contributes to a growing literature showing that efficiency improvements can trigger behavioral responses that offset their initial gains. A central insight in this literature is that reductions in the marginal cost of an activity—whether driven by technological progress or capacity expansion—induce additional demand, thereby attenuating improvements in per-unit performance. This mechanism has been documented in a range of settings, including transportation, where road

[1] For example, judges in Chile handle more than 5,000 cases annually on average, while courts in Ecuador dispose of only about 40 percent of filed cases—implying that even if no new cases were accepted, it would take nearly a decade to clear the existing backlog (Bosio, 2023).

expansions induce proportional increases in traffic (Duranton & Turner, 2011), consistent with classic congestion theories (Vickrey, 1969). Related evidence from task-intensive environments further shows that improvements in processing capacity are often accompanied by expansions in workload, limiting observable productivity gains (Brynjolfsson, 1993; Brynjolfsson & Hitt, 1996). Our paper extends this line of research to judicial systems, a setting characterized by high-skill labor, institutional constraints, and endogenous case inflows. We show that digital technologies not only reduce per-case processing time but also induce additional case inflows, leading to a rebound in average processing time. By documenting this demand-induced adjustment in a professionalized public sector context, our findings provide new evidence on how technological progress reshapes workload and performance when demand is endogenous.

The remainder of this paper is organized as follows: Section 2 reviews the institutional background. Section 3 describes the data used in our empirical analyses. Section 4 presents our econometric strategies. Section 5 shows main empirical results and a series of robustness checks. Section 6 further discusses the mechanisms from the perspective of in-court frictions and personnel reallocation. Section 7 goes beyond the evidence on broader impact of the reform on workload as well as its dynamic effects. Section 8 concludes the paper.

2 Institutional Background

2.1 Court Hearing and Trail Transcripts

Court hearings serve as the institutional core of the adjudication process, providing the primary forum where legal arguments are contested and evidence is presented. Within this framework, trial transcripts constitute the definitive, official written record of proceedings. They not only ensure procedural legality but also preserve the verbatim statements of litigants, witness testimonies, and judicial interactions. Consequently, these transcripts form the foundational baseline for judicial decision-making, case quality evaluation, and subsequent appellate or retrial reviews. Because trial transcripts feed directly into these downstream stages of adjudication, any recording inaccuracies or verification delays can propagate through the entire case life-cycle. This sequential dependency renders the trial recording process a potentially binding constraint and a major source of operational friction within court systems.

Across divergent legal traditions, the generation and management of these vital records rely fundamentally on specialized judicial support personnel. In civil law jurisdictions, such as China,[2] these

[2]Contemporary legal systems are commonly classified into two major families: civil law and common law. Civil law prevails in continental Europe and much of Latin America, Africa, and Asia, whereas common law characterizes the United Kingdom and most of its former colonies. The two differ primarily in the allocation of judicial authority, with civil-law systems relying on state-appointed professional judges and codified statutes and common-law systems relying more on lay juries and judicial independence (Glaeser & Shleifer, 2002).

responsibilities are consolidated within a unified clerkship role, where government-appointed professional clerks manage the entire workflow—ranging from courtroom organization and real-time recordkeeping to file compilation and archiving.[3] Conversely, common law jurisdictions disaggregate these trial-support functions among multiple specialized auxiliary positions.[4] For instance, the task of verbatim recording is typically assigned to dedicated court reporters rather than a generalized clerk. Despite these differences in institutional design, the heavy reliance on human capital for trial transcription remains an invariant feature; producing and verifying these records requires substantial investments of labor and public funds, representing a significant institutional cost in modern court operations.

2.2 AI Transcription Reform in China

Under the traditional model of judicial production, trial recording in Chinese courts relied exclusively on the human capital of court clerks, who manually captured and summarized courtroom proceedings. This labor-intensive workflow encompassed extensive pre-trial preparation, real-time shorthand transcription during hearings, and protracted post-hearing editing and finalization. Given the constrained duration of hearings and the high density of information exchanged, human clerks faced an inherent physical bottleneck, making omissions and inaccuracies in verbatim recording structurally unavoidable. As a result, the post-hearing confirmation of trial transcripts by litigants became a major institutional checkpoint fraught with friction. When disputes arose over the recorded content — particularly when verification could not be completed immediately in court — judges, clerks, and litigants were forced to engage in costly ex-post reconstructions of the proceedings. Crucially, these transcription deficiencies and verification delays propagated downstream, frequently triggering additional hearings and generating systemic procedural backlogs that significantly expanded the duration of the entire adjudication cycle, thereby imposing substantial transaction costs on the court system.

To alleviate these operational bottlenecks and transaction costs, Chinese courts initiated a progressive reform of trial recording practices in the mid-2010s as part of the national "smart court" mandate.[5] Specifically, beginning in 2016, the Supreme People's Court promoted AI-enabled transcription

[3] In China, the role and responsibilities of court clerks are defined by law: *the Organic Law of the People's Courts* assigns clerks to trial recordkeeping and other adjudicative support tasks, and *the Measures for the Administration of Court Clerks* further specify their duties in pretrial preparation, courtroom organization, and case file management.

[4] In the U.S. federal court system, trial-support functions comparable to those performed by clerks in civil law jurisdictions are disaggregated among multiple specialized positions, including law clerks, courtroom deputy clerks, and court reporters. Pursuant to the Federal Court Reporting Program (28 U.S.C. § 753), statutory mandates require each court session to be recorded verbatim using methods elected by the district judge, such as shorthand, stenotype, stenomask, or electronic sound recording equipment. Court reporters are primarily responsible for preparing and certifying these written transcripts upon the request of a party or order of the court, subject to regulated fee schedules and delivery timelines. For further details, see the Administrative Office of the U.S. Courts, "Federal Court Reporting Program," archived at https://web.archive.org/web/20150905061701/http://www.uscourts.gov/services-forms/federal-court-reporting-program (accessed May 16, 2026).

[5] The construction of "smart courts" originates from China's broader judicial reforms and national informatization strategy initiated after 2012. In 2016, the Supreme People's Court formally proposed the goal of building "smart courts", which was subsequently incorporated into the *National Informatization Development Strategy Outline*.

systems to generate real-time text during hearings,[6] and subsequent regulations formally permitted audio-video recordings to partially substitute for written transcripts in certain cases.[7] This top-down policy framework paved the way for a major technological transition, which unfolded sequentially across local courts in two distinct stages: progressing from an initial phase of human-AI collaboration (where AI assisted existing clerks) to a second phase of complete institutional substitution, where AI fully replaced human clerks in the trial recording process.

Stage I: Intelligent Voice-Assisted Trial Recording (IVT Reform). Launched in 2016, the first stage of the reform introduced this AI-enabled transcription system to the courtroom. Under this stage, court clerks were still required to be physically present in the courtroom throughout hearings. However, instead of typing everything manually from scratch, they shifted to supervising and correcting the text generated by the AI system. This hybrid approach helped clerks finalize transcripts more quickly and reduced the need for extensive post-hearing verification. The roll-out of this stage followed a staggered timeline across courts, moving from early pilots in provinces like Jiangsu and Zhejiang in 2016–2017 to a broader expansion in 2018 as adoption accelerated.

Stage II: Intelligent Voice Substitution (Trial Without Clerk, TWC Reform). Building on the technology established in the first stage, the second stage went a step further by entirely removing court clerks from the courtroom in selected cases. First introduced in Shanghai in 2020 and later expanded, this model relies completely on integrated audio-video hardware and the automated AI speech recognition system to capture and record proceedings. Supported by digital authentication technologies such as time-stamping and blockchain, the automated text and audio-video records together serve as the official court record. The key difference from the first stage is that the TWC reform completely eliminates the input of human clerical labor during hearings. By cutting out the traditional human-managed transcription and verification loops, this stage achieves a much larger reduction in time delays and further compresses the overall case-processing timeline.

Taken together, these reforms divide the production of trial transcripts into three distinct modes:

- **Record Mode 1 (Traditional Manual Recording):** Clerks record everything manually, requiring extensive and time-consuming post-hearing verification.
- **Record Mode 2 (AI-Assisted Recording):** Clerks remain in court to supervise and correct AI-generated drafts, which reduces verification costs but still requires full clerical labor.

[6] Unlike conventional dictation software, these specialized judicial AI systems are explicitly trained on vast libraries of legal data to accurately recognize technical legal terminology, filter courtroom noise, and use voice biometrics to automatically distinguish and label the speech of different participants (such as judges, plaintiffs, and defendants) in real time.

[7] In 2016, the Supreme People's Court called for reforming trial recording methods in its *Opinions on Further Promoting the Diversion of Complex and Simple Cases and Optimizing the Allocation of Judicial Resources*, encouraging the use of audio-video recording to simplify or replace manual transcription. Article 8 of *the Provisions on the Audio and Video Recording of Court Trials* by People's Courts (2017) further allows audio-video recordings to substitute for written transcripts in civil cases under summary procedures.

- **Record Mode 3 (AI-Based Substitution):** Clerks are entirely removed from the courtroom, and trial records are finalized and confirmed in real time by the AI system.

The staggered roll-out of these reforms across different courts and over time creates substantial variation in trial recording practices. This empirical setting provides a valuable opportunity to analyze how moving from partial technological assistance (Stage I) to the complete substitution of human labor (Stage II) affects court efficiency and related outcomes.

3 Data and Variables

3.1 Key Outcome Variables

Judicial Case Documents. Since January 2014, the State Council of China has mandated that all local courts upload judicial dec isions to a publicly accessible platform known as *China Judgments Online*. As of March 2025, the platform hosts more than 152 million judgment documents. Our identification strategy exploits the staggered rollout of the Intelligent Voice Transcription (IVT) Reform after 2017 and, in particular, the introduction of its second-stage component—IVT without clerks—in Shanghai in April 2020. To estimate the stage-specific impacts of the reform, we compile the universe of judgment documents for cases heard in Shanghai courts from 2017 to 2022. To construct a credible counterfactual, we collected cases during the same period from courts in Hangzhou, Ningbo, and Suzhou. Courts in these jurisdictions did not adopt the second-stage reform during our sample period, allowing them to serve as a clean control group. These locations were selected based on their close similarity to Shanghai along key socioeconomic and judicial dimensions, including regional GDP, judicial civility indices, and population size, thus improving the comparability of pre-treatment trends.

Using large-scale text analysis techniques, we extract rich case-level information from judgment documents, including the identities of plaintiffs and defendants, the presiding judges and clerks, courts, case characteristics, verdict outcomes, and precise timestamps for filing, adjudication, and online publication. These data enable us to construct a set of outcome variables relevant for measuring shifts in judicial productivity and behavior, such as indicators of case filing thresholds, appeal rates, reversal rates, and judge turnover. In addition, the structure of the documents allows us to quantify textual and procedural features, including the length of specific sections and the number of statutory provisions cited. In addition, we apply generative AI techniques and large language models to evaluate multiple dimensions of judgment writing quality, including scores for expression, structure, fact presentation, and format.

China Court Trial Online. Data on hearing duration are from *China Court Trial Online*, an official livestreaming platform established by Chinese courts to enhance judicial transparency by broadcasting courtroom proceedings. Since its launch, the platform has accumulated a large archive of trial videos,

providing valuable real-time insights into courtroom proceedings. We collected approximately 20 million livestream records from 2014 to 2024 from this platform. We extracted metadata embedded in the video files—including the start and end timestamps—to compute the duration of each trial. We then link these trial records to their corresponding judgment documents using unique case identifiers. This linkage enables us to obtain a precise, case-level measure of hearing duration.

3.2 Key Independent Variables

Data on Deployment of AI Equipment in Courts. We obtained detailed data on the deployment timeline and specific implementation of intelligent speech systems in specific courts from *Company K*, who has played a key role in supporting the development of "Intelligent Courts" by providing a range of technological solutions. In partnership with Company K, many courts have integrated intelligent speech technology into trial proceedings, reforming the traditional transcription process. This technology enables the automatic real-time transcription of spoken content from judges, plaintiffs, defendants, and other trial participants, converting speech into text during court sessions.

We also cross-verify these private data with court-related procurement data from government purchasing records. Using these proprietary data, we can systematically identify whether a court has implemented IVT reforms and determine the exact timing of these reforms. This allows us to clearly distinguish between pre-reform and post-reform periods, enabling us to construct more accurate reform dummy variables in the empirical models and conduct a more precise analysis of their impact.

3.3 Additional Variables

Data on Nighttime Light. Nighttime light remote sensing data measure the brightness of the Earth's surface at night using satellite-based observations and have been widely used as proxies for economic activity, urbanization, and population density. In this paper, we leverage nighttime light data to construct a proxy for overtime work at courts.

Specifically, we use court address information to match geographic coordinates via the Amap API, apply address correction, and extract nighttime light intensity for the grid where the court is located as well as surrounding grids. We compare the nighttime light intensity of the court's grid with the median intensity of its eight neighboring grids. If the court's grid exhibits a higher light intensity than the surrounding median, we classify it as experiencing overtime work during that period. We rely on light intensity measured at 1:00 a.m., which allows this measure to capture extreme late-night activity. This measure provides a high-frequency, externally observable proxy for overtime work that is not subject to administrative reporting bias. It also enables us to examine whether the adoption of AI-assisted reforms alleviates the workload of court personnel.

Descriptive statistics for main variables used in the empirical specification are summarized in Table

1.

Table 1. *Summary Statistics*

VARIABLES	Observation	Mean
Case Level		
IVT Reform (=1)	1,604,391	0.792
IVT Reform with Clerk (=1)	1,604,391	0.712
IVT Reform without Clerk (=1)	1,604,391	0.080
Adjudication Duration (logged)	578,463	4.264
Hearing Duration (logged)	246,669	3.043
Number of Hearings	595,750	1.341
Post-Hearing Share	465,277	0.343
Post-Hearing Duration (Days)	465,277	47.07
Pre-Hearing Duration (Days)	465,277	51.99
Cases under summary procedure (=1)	1,604,391	0.637
Appealed (=1)	164,623	0.002
Uphold the Decision (=1)	146,563	0.952
Words of Litigation Record	637,468	174.7
Number of Cited Article	164,623	8.669
Text Score of Expression	637,201	4.371
Text Score of Fact	637,201	4.501
Text Score of Structure	637,201	4.563
Text Score of Format	637,201	4.520
Court-Month Level		
Filing Threshold	3,763	0.085
Number of cases (All, logged)	3,829	6.521
Number of cases (Hard for litigation, logged)	3,829	2.598
Rate of Work Overtime	3,385	0.429
Number of judges quitting	3,829	0.493
Number of clerks quitting	3,829	1.024

Notes: All case-level summary statistics are calculated from the subset of court documents corresponding to judgment cases.

4 Econometric Strategy

To estimate the causal impact of the IVT reform on judicial outcomes, we exploit the staggered roll-out of the reform across courts and implement a difference-in-differences (DID) design. The baseline specification is:

$$Y_{ict} = \alpha + \beta IVT_{ct} + X_i'\gamma + \theta_c + \mu_t + \varepsilon_{ict} \tag{1}$$

where i, c, and t denote case, court, and judgment year-month, respectively. Y_{ic} represents case-level judicial outcomes. Our primary outcome is *Adjudication Duration*, defined as the number of days between the filing date and the final judgment date of case i, expressed in natural logarithms. IVT_{ct} is a dummy variable equal to one if court c has implemented the IVT reform by time t, and zero otherwise. The coefficient of interest is β. X_i' is a vector of case-level control variables, including case type (e.g.,

civil or criminal), trial procedure (e.g., first-instance or second-instance), and case complexity, proxied by whether the case is adjudicated under the summary procedure or the ordinary procedure. θ_c and μ_t denote court fixed effects and year-month fixed effects, respectively. ε_{ict} is the error term. Robust standard errors are clustered at the court-by-year-month level.

	Value of *IVT Reform with Clerk (=1)*	**Value of *IVT Reform without Clerk (=1)***
Record Mode 1	0	0
Record Mode 2	1	0
Record Mode 3	0	1

Figure 1. *Values of Independent Variables Under Three Modes*

Furthermore, the IVT reform was implemented in two stages: Stage 1, *IVT with a clerk*, and Stage 2, *IVT without a clerk*, also known as the TWC reform. These two stages correspond to two distinct mechanisms through which AI adoption may affect judicial performance. We estimate the following specification:

$$Y_{ict} = \alpha + \lambda_1 IVT_with_clerk_{ct} + \lambda_2 IVT_without_clerk_{ct} + X_i'\gamma + \theta_c + \mu_t + \varepsilon ict. \tag{2}$$

where $IVT_with_clerk_{ct}$ equals one if court c has implemented Stage 1, but not Stage 2, by time t, and zero otherwise. $IVT_without_clerk_{ct}$ equals one if court c has implemented Stage 2 by time t, and zero otherwise. The omitted category is the pre-reform mode, in which trial records are produced entirely by clerks without AI assistance. All other variables are defined as in Equation (1). Figure 1 maps these two indicators onto the three trial-recording modes: *Record Mode 1*, the pre-reform manual mode without AI assistance; *Record Mode 2*, the AI-assisted mode in which clerks remain in the courtroom to monitor and verify real-time transcripts; and *Record Mode 3*, the AI-substitution mode in which IVT replaces clerks in courtroom transcription. This specification estimates the effects of the two reform stages separately: λ_1 captures the effect of Stage 1, *IVT with a clerk*, while λ_2 captures the effect of Stage 2, *IVT without a clerk*, both relative to the pre-reform manual-recording mode.

5 Impact of IVT Reform on Trial Efficiency

5.1 Baseline Results: Reducing Adjudication Duration

Table 2 presents the results from estimating Equations (1) and (2). Columns 1-2 report estimates without case-level controls, including only court and year-month fixed effects. Columns 3-4 additionally control for case characteristics, which capture systematic differences in case complexity that may influence adjudication duration. Columns 1 and 3 estimate the average effect of the IVT reform based on Equation (1), while Columns 2 and 4 decompose the reform into two stages based on Equation (2). These different specifications consistently indicate that the IVT reform substantially reduces adjudica-

tion duration. The coefficients of *IVT Reform* in Columns 1 and 3 are -0.129 and -0.125, both statistically significant at the 1% level and quantitatively stable across specifications. The two-stage decomposition in Columns 2 and 4 further shows that both stages of the reform contribute to the observed efficiency gains. Of particular interest is *IVT Reform without Clerk*, which captures the effect of fully substituting AI for clerical labor in courtroom transcription. The coefficient on this variable is -0.114 in Column 2 and -0.108 in Column 4, both significant at the 1% level.

Table 2. *Impact of IVT Reform on Trial Efficiency*

	Adjudication Duration (logged)			
	(1)	(2)	(3)	(4)
IVT Reform (=1)	-0.129***		-0.125***	
	(0.024)		(0.025)	
IVT Reform with Clerk (=1)		-0.129***		-0.125***
		(0.024)		(0.025)
IVT Reform without Clerk (=1)		-0.114***		-0.108***
		(0.036)		(0.036)
Year-Month fixed effects	YES	YES	YES	YES
Court fixed effects	YES	YES	YES	YES
Case Characteristic Controls	NO	NO	YES	YES
Observations	578,463	578,463	578,463	578,463
R-squared	0.226	0.226	0.252	0.252
p-value: IVT Reform with Clerk (=1) = IVT Reform without Clerk (=1)	-0.015(0.577)			

Notes: All estimations are conducted using the subsample of court documents identified as judgments. Verdict documents are not included in the present analysis; they will be incorporated in subsequent empirical exercises. Robust standard errors adjusted for clustering at the court-by-year-month level are given in parentheses. ***, **, and * indicate statistical significance at 1%, 5% and 10%, respectively.

The economic magnitude of the effect is also sizable. Taking Column 3, which incorporates the full set of controls, as our preferred specification, the estimate implies that the IVT reform reduces adjudication duration by approximately 11.75% ($1 - e^{-0.125}$), or a reduction of about 13-15 days per case.[8] In terms of labor savings, the aggregate time savings are equivalent to the workload of approximately 12 full-time judges per court per year.[9] In terms of financial savings, the reform generates annual labor cost savings of at least 1.2 million RMB per court after netting out equipment costs, roughly two to three times the annualized procurement cost of the IVT system.[10] Accounting for the broader social benefits associated with improved judicial efficiency would further amplify these gains.

[8] The average adjudication duration is approximately 130 days for Shanghai courts and 115 days for courts in neighboring cities, so an 11.75% reduction translates into roughly 13-15 days per case.

[9] This calculation is based on a median annual caseload of around 116 cases per judge and an average of 100 judges per court.

[10] According to internal data, fully deploying an IVT system, including equipment installation and annual maintenance, costs roughly 2 to 3 million RMB, with an annualized cost of 400,000 to 600,000 RMB assuming a five-year depreciation period. Even under a conservative estimate of 150,000 RMB per judge per year (noting that our sample courts are located in coastal regions with relatively higher salaries), the reform generates annual labor savings of at least 1.2 million RMB per court after netting out equipment costs ($= 150{,}000 \times 12 - 600{,}000$).

5.2 Event Study and Robustness Checks

To interpret β causally, the parallel trends assumption requires that, absent the IVT reform, courts adopting the reform at different times would have followed similar trends. We assess this assumption and examine dynamic treatment effects using the following event-study specification:

$$Y_{ict} = \alpha + \sum_{k=-4, k\neq -1}^{11} \beta_k \cdot D_{c,k} + X_i'\gamma + \theta_c + \mu_t + \varepsilon ict, \tag{3}$$

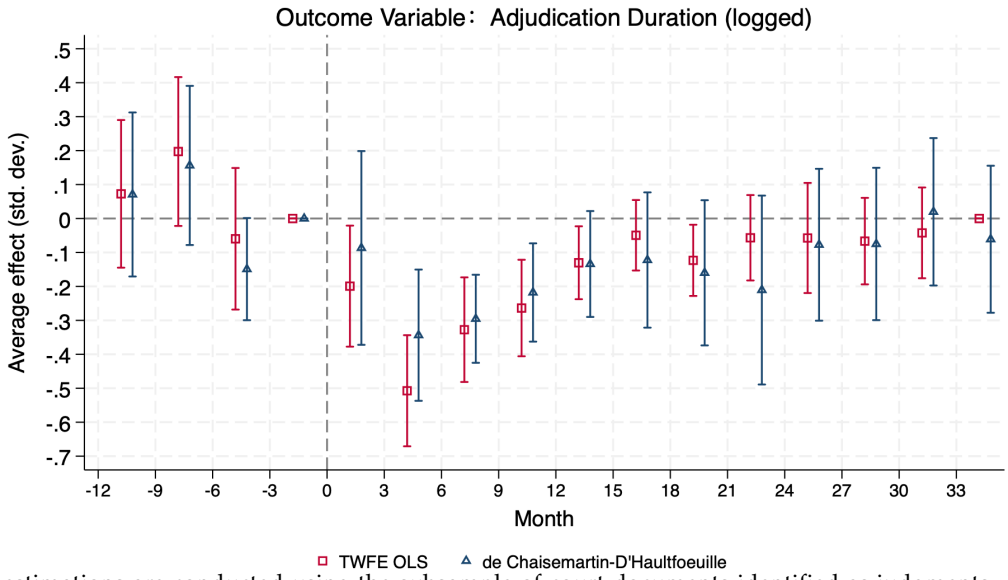
where $D_{c,k}$ equals one if court c is in event period k relative to the implementation of the IVT reform, and zero otherwise. Each event period corresponds to three months, with $k = -1$ (months -3 to -1) omitted as the reference period. The endpoint indicators bin periods four or more periods before the reform and eleven or more periods after the reform. X_i' denotes the same set of case-level controls as in the baseline specification, while θ_c and μ_t denote court fixed effects and year-month fixed effects, respectively. The coefficients β_k for $k < 0$ capture pre-treatment dynamics, and those for $k \geq 0$ trace the evolution of treatment effects after the reform. To address treatment-effect heterogeneity in staggered DID settings, we also estimate event-study models using heterogeneity-robust estimators.

Figure 2 reveals a clear dynamic pattern: adjudication duration declines sharply immediately after the IVT reform, but begins to rebound after about six months, with the effect fading and becoming statistically insignificant after roughly twelve months. We investigate this initial decline and subsequent rebound in adjudication duration in greater detail in a later section. By contrast, Figure 3 indicates that hearing duration declines persistently and remains statistically significant throughout the post-reform period. This contrast highlights heterogeneous dynamic responses across different stages of case processing.

We conduct two additional robustness checks to address potential confounding factors. First, since the COVID-19 pandemic substantially disrupted court operations in Shanghai (the pilot region for the *IVT Reform without Clerk*) during the first half of 2022, we re-estimate the models excluding cases adjudicated in 2022, and further excluding all post-2019 observations. Second, to rule out the concern that the estimated effects merely reflect compositional changes in judicial personnel rather than behavioral responses to the reform, we identify the presiding judge and clerk for each case and include individual fixed effects in all regressions. As reported in Appendix C, the estimated effects remain consistent with the baseline results under both exercises, supporting the robustness of our findings.

5.3 Does Efficiency Come at the Cost of Quality?

The preceding analyses show that the IVT reform improves trial efficiency. A natural question is whether these gains come at the expense of judicial quality, which we examine below. We measure judicial quality along two complementary dimensions: *ex-post review* indicators (appeal and reversal



Notes: All estimations are conducted using the subsample of court documents identified as judgments. This figure overlays the event-study plots constructed using two different estimators: a dynamic version of the TWFE model (in red with square markers); and De Chaisemartin and d'Haultfoeuille (2020) (in blue with triangle markers). The bars represent 95 percent confidence intervals. The following figures for dynamic impact are similar.

Figure 2. *Dynamic Impact of IVT Reform on Adjudication Duration*

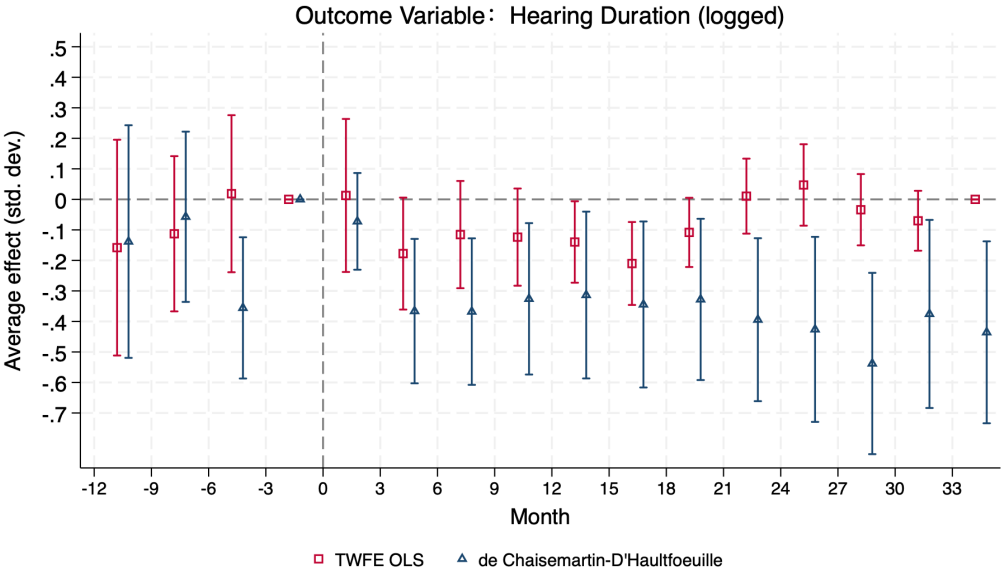


Figure 3. *Dynamic Impact of IVT Reform on Hearing Duration*

rates), which capture how judgments are perceived by litigants and appellate courts, and *textual quality* indicators (litigation-record length and statutory citations), which capture the depth of trial records and legal reasoning.

We begin with the ex-post review dimension. *Appealed (=1)* is a dummy variable equal to one if a case is appealed and zero otherwise; *Uphold the Decision (=1)* is a dummy variable equal to one if an appealed case is upheld without modification and zero otherwise. The appeal rate reflects litigants' dissatisfaction with the first-instance ruling, since well-reasoned and procedurally sound judgments

are less likely to be challenged; the reversal rate reflects the appellate court's assessment, since decisions upheld without modification signal no material errors in fact-finding or legal application. Table 3 reports the results. As shown in Columns (1)–(4), the coefficients on the reform variables are small in magnitude and statistically insignificant for both dependent variables, indicating that while the reform accelerates trial processing, it does not compromise the substantive correctness of judgments.

Table 3. *Impact of IVT Reform on Judgement Quality*

	Appealed (=1)		Uphold the Decision (=1)	
	(1)	(2)	(3)	(4)
IVT Reform (=1)	0.000		0.003	
	(0.001)		(0.004)	
IVT Reform with Clerk (=1)		0.000		0.003
		(0.001)		(0.004)
IVT Reform without Clerk (=1)		0.000		0.003
		(0.001)		(0.004)
Year-Month fixed effects	YES	YES	YES	YES
Court fixed effects	YES	YES	YES	YES
Case Characteristic Controls	YES	YES	YES	YES
Observations	164,623	164,623	145,496	145,496
R-squared	0.002	0.002	0.015	0.015

Notes: Robust standard errors adjusted for clustering at the court-by-year-month level are given in parentheses. ***, **, and * indicate statistical significance at 1%, 5% and 10%, respectively.

We turn next to the textual quality dimension, following Chen et al. (2026) and Liu et al. (forthcoming). *Words of Litigation Record* measures the total word count of the litigation-record section of each judgment document, capturing the richness and completeness of the trial record; *# of Cited Articles* counts the number of statutory provisions cited in each judgment, reflecting the precision with which the legal basis for the ruling is articulated. Table 4 reports the results. Columns (1)–(2) show that litigation records become longer after the reform, suggesting that AI-assisted transcription captures arguments, evidence, and procedural details more fully and thereby improves the richness and completeness of trial records. Columns (3)–(4) show a significant increase in statutory citations, consistent with two complementary channels: more complete trial records facilitate a clearer articulation of the legal basis for decisions, and clerks devote greater attention to legal reasoning once relieved of the mechanical burden of manual transcription.

Beyond these results, we further assess the reform's effect on documentation quality and find improvements in substantive dimensions such as linguistic clarity and factual accuracy (see Appendix D for details). Taken together, the evidence indicates that the IVT reform improves trial efficiency without compromising judicial quality: court personnel process cases more quickly while maintaining—and along some dimensions even enhancing—the substantive quality of judicial decisions.

Table 4. *Impact of IVT Reform on Documentation Quality*

	Words of Litigation Record		Number of Cited Articles	
	(1)	(2)	(3)	(4)
IVT Reform (=1)	5.031***		0.563***	
	(1.285)		(0.157)	
IVT Reform with Clerk (=1)		5.024***		0.552***
		(1.286)		(0.157)
IVT Reform without Clerk (=1)		6.283***		1.063***
		(1.582)		(0.266)
Year-Month fixed effects	YES	YES	YES	YES
Court fixed effects	YES	YES	YES	YES
Judge fixed effects	YES	YES	YES	YES
Clerk fixed effects	YES	YES	YES	YES
Case Characteristic Controls	YES	YES	YES	YES
Observations	631,874	631,874	163,796	163,796
R-squared	0.344	0.344	0.211	0.212

Notes: All estimations are conducted using the subsample of court documents identified as judgments. Robust standard errors adjusted for clustering at the court-by-year-month level are given in parentheses. ***, **, and * indicate statistical significance at 1%, 5% and 10%, respectively. !!Coefficients for Case Characteristic Controls in Columns 3-4 are omitted.

6 Mechanisms: From In-Court Frictions to Personnel Reallocation

We unpack two channels through which the IVT reform improves efficiency. The first is a *direct effect* within the courtroom: real-time AI transcription reduces in-court frictions such as pauses for note-taking and disputes over transcript accuracy. The second is a *spillover effect* beyond the courtroom: by removing clerks from in-court duties, the second-stage reform frees personnel capacity for reallocation to other case-processing tasks. These two channels jointly explain why both stages deliver comparable gains: Stage 1 primarily activates the in-court channel, while Stage 2 additionally activates personnel reallocation.

6.1 Direct Effect: Reducing In-Court Frictions

By replacing manual note-taking with real-time, high-accuracy speech-to-text transcription, the IVT system may streamline in-court proceedings along several dimensions. Under the traditional system, courtroom proceedings are constrained by clerks' manual transcription speed: judges may pause questioning to allow clerks to complete their notes, parties may be asked to repeat statements when segments are missed, and disputes during post-hearing verification of the record may necessitate further clarification. These frictions not only prolong individual hearings but also force cases to be split into multiple sessions when manual note-taking fails to keep pace or when transcription inaccuracies require supplementary proceedings. By generating accurate transcripts automatically and in real time, the IVT system alleviates both sources of friction. We therefore examine this direct effect along two

complementary dimensions: the duration of each hearing and the number of hearings per case.

We first examine the impact of the IVT reform on hearing duration. We replace the dependent variable with *Hearing Duration*, measured as the duration of court hearing videos obtained from the official livestreaming platform, which provides a clean measure of in-court processing time. Columns (1)-(2) of Table 5 report the results. The reform shortens hearing duration by approximately 6.20% ($1 - e^{-0.064}$), and the two-stage decomposition shows that both stages contribute significantly, with the coefficient on *IVT Reform without Clerk* estimated at -0.085 and significant at the 1% level.

We next turn to the number of hearings per case. Columns (3)-(4) of Table 5 show that the reform significantly reduces the number of hearings per case, with a particularly pronounced effect under *IVT Reform without Clerk*: the coefficient is -0.272 in Column (4), substantially larger in magnitude than the -0.056 estimated for *IVT Reform with Clerk*. This pattern suggests that the efficiency gains from AI adoption depend critically on how human labor is reorganized around the technology. While the first stage merely accelerates recording speed, the second stage represents a deeper level of automation that eliminates residual coordination frictions and allows hearings to proceed more continuously.

Beyond these two direct in-court outcomes, we further trace where the resulting efficiency gains accrue within the case lifecycle by decomposing total adjudication duration into pre- and post-hearing components. We define *Post-Hearing Duration* as the number of days from hearing to judgment, which captures all case processing that occurs after the hearing concludes. Columns (5)-(6) of Table 5 report the results. The reform significantly reduces post-hearing duration, with a particularly large effect under *IVT Reform without Clerk*. This pattern indicates that much of the efficiency gain documented above is realized after the hearing stage, consistent with the view that real-time, high-accuracy transcripts eliminate downstream frictions, such as verification disputes and supplementary clarifications, that would otherwise prolong post-hearing procedures.

Taken together, these results indicate that the IVT reform enhances trial efficiency through two reinforcing channels, shorter hearings and fewer repeated sessions, with the bulk of the resulting time savings materializing in the post-hearing stage.

6.2 Spillover Effect: Personnel Reallocation Beyond the Courtroom

We next examine whether the IVT reform generates spillover effects beyond the courtroom by reallocating clerical capacity across case-processing tasks. The logic is that the second-stage reform removes clerks from in-court recording duties in eligible cases, thereby freeing clerical time that can be redirected to other tasks and other cases. We examine this capacity-reallocation mechanism along two margins: spillovers to cases under the ordinary procedure and spillovers to verdict documents.

Spillovers to cases under the ordinary procedure. Under the rules governing the reform, the clerk-free mode is mainly applied to cases under the summary procedure, while most cases under the ordinary

Table 5. *Streamlining Hearing Duration and Alleviating Hearing Frictions*

	Hearing Duration (logged)		Number of Hearings		Post-Hearing Duration (logged)	
	(1)	(2)	(3)	(4)	(5)	(6)
IVT Reform (=1)	-0.064**		-0.056***		-0.285***	
	(0.027)		(0.019)		(0.060)	
IVT Reform with Clerk (=1)		-0.064**		-0.056***		-0.278***
		(0.027)		(0.019)		(0.060)
IVT Reform without Clerk (=1)		-0.085**		-0.272***		-0.631***
		(0.038)		(0.029)		(0.094)
Year-Month fixed effects	YES	YES	YES	YES	YES	YES
Court fixed effects	YES	YES	YES	YES	YES	YES
Case Characteristic Controls	YES	YES	YES	YES	YES	YES
Observations	246,668	246,668	595,750	595,750	465,277	465,277
R-squared	0.111	0.111	0.064	0.066	0.192	0.193

Notes: The estimations are conducted using the subsample of court documents identified as judgments. *Post-Hearing Duration (logged)* represents the total case duration occurring after the hearing, measured as the number of days from hearing to judgment (logged). Robust standard errors adjusted for clustering at the court-by-year-month level are given in parentheses. ***, **, and * indicate statistical significance at 1%, 5% and 10%, respectively.

procedure still require clerks to participate in court hearings. This institutional feature provides a useful test of the reallocation mechanism. If the second-stage reform merely improves the efficiency of directly treated summary cases, we should observe limited effects on ordinary-procedure cases. By contrast, if removing clerks from summary-procedure hearings frees clerical capacity, this capacity can be reallocated to more complex ordinary-procedure cases, which require greater out-of-court clerical input, such as organizing evidentiary materials, preparing case files, and assisting with legal drafting.

Table 6 reports the results. Column (1) shows that the IVT reform significantly reduces adjudication duration for cases under the summary procedure. Column (2) examines cases under the ordinary procedure. Although these cases are generally not directly subject to the clerk-free mode, the coefficient on *IVT Reform without Clerk* is negative and statistically significant, whereas the coefficient on *IVT Reform with Clerk* is small and insignificant. This pattern is consistent with a spillover effect: once clerks are released from in-court recording duties in summary cases, courts can reallocate clerical time toward ordinary-procedure cases, thereby improving the processing efficiency of more complex cases. Taken together, the results in Table 6 suggest that the second-stage reform improves judicial efficiency not only by changing how directly treated cases are recorded, but also by reallocating clerical capacity across cases.

Spillovers to verdict documents. We next examine whether the reform also affects court documents whose preparation is primarily conducted outside the hearing. In the baseline analysis, we focus on judgment documents, which are closely linked to in-court proceedings. In practice, however, verdict documents constitute another major category of court documents and account for a volume compara-

Table 6. *Spillover Effects on Cases under Ordinary Procedure*

	Adjudication Duration (logged)	
	(1) Cases under summary procedure	(2) Cases under ordinary procedure
IVT Reform with Clerk (=1)	-0.137***	-0.030
	(0.031)	(0.043)
IVT Reform without Clerk (=1)	-0.156***	-0.185***
	(0.040)	(0.067)
Observations	412,953	294,134
R-squared	0.241	0.226
Year-Month FE	YES	YES
Court FE	YES	YES
Case Characteristic Controls	YES	YES

Notes: The estimation in Column (1) is based on cases under the summary procedure. The estimation in Column (2) is based on cases under the ordinary procedure. Robust standard errors adjusted for clustering at the court-by-year-month level are given in parentheses. ***, **, and * indicate statistical significance at 1%, 5% and 10%, respectively.

ble to judgments. Because the preparation of verdict documents relies heavily on out-of-court clerical and administrative work, they provide another setting in which to test whether the reform reallocates clerical capacity beyond courtroom transcription.

Table 7 reports the results. Columns (1) and (2) use the subsample of verdict documents. The overall IVT indicator in Column (1) is not statistically significant, but the stage-specific estimates in Column (2) reveal a clear pattern: *IVT Reform with Clerk* has no significant effect, whereas *IVT Reform without Clerk* significantly reduces the processing time of verdict documents. This pattern is consistent with the capacity-reallocation mechanism. The first-stage reform primarily improves in-court transcription while clerks remain tied to hearings, whereas the second-stage reform releases clerks from in-court duties and allows more time to be allocated to out-of-court document preparation.

Overall, these results provide complementary evidence for the spillover effect of the reform. By freeing clerks from in-court transcription duties in eligible cases, the second-stage reform allows courts to reallocate clerical capacity to more complex cases and out-of-court document preparation, thereby improving judicial efficiency beyond the courtroom.

7 Broader Impact: Does the IVT Reform Reduce Judicial Workload?

7.1 Efficiency Gains or Workload Intensification?

Does AI-assisted technology truly make judicial work easier, or does it simply enable courts to process more cases at a faster pace, potentially intensifying the workload of judicial personnel? Building on the preceding analyses, we move beyond average efficiency effects and examine whether the IVT reform genuinely alleviates the burden on judges and clerks over time, or instead leads to work intensification through expanded case processing. To address this question, we study the dynamic impacts of the

Table 7. *Spillover Effects on Verdict Documents*

	Adjudication Duration (logged)	
	Verdict documents	
	(1)	(2)
IVT Reform (=1)	-0.013	
	(0.072)	
IVT Reform with Clerk (=1)		0.002
		(0.070)
IVT Reform without Clerk (=1)		-0.371***
		(0.100)
Year-Month fixed effects	YES	YES
Court fixed effects	YES	YES
Case Characteristic Controls	YES	YES
Observations	127,831	127,831
R-squared	0.242	0.243

Notes: All estimations are conducted using the subsample of court documents identified as verdict documents. Robust standard errors adjusted for clustering at the court-by-year-month level are given in parentheses. ***, **, and * indicate statistical significance at 1%, 5% and 10%, respectively.

IVT reform along three dimensions: (i) trial efficiency, (ii) workload, and (iii) judgment quality. This dynamic perspective allows us to trace how the effects of the reform evolve in the short and medium term, and to assess whether initial gains are sustained, amplified, or offset by subsequent adjustments in court operations.

Dynamic Impact of the IVT Reform on Trial Efficiency. Tables 2 and 5 document that the IVT reform significantly reduces both adjudication duration and hearing duration on average. A dynamic analysis, however, reveals a sharp contrast in how these effects evolve over time. As shown in Figure 2, adjudication duration declines substantially in the months immediately following the reform, reaching its lowest point roughly six months after implementation. Thereafter, the effect gradually attenuates and becomes statistically indistinguishable from zero approximately one and a half years post-reform. By contrast, Figure 3 shows that the reduction in hearing duration persists throughout the entire post-reform period.

This divergence is informative. The sustained decline in hearing duration is consistent with the mechanical nature of AI-assisted transcription: by continuously streamlining in-court proceedings, the technology permanently reduces the time required to conduct hearings. The transient effect on adjudication duration, in contrast, points to broader system-level adjustments. In China's judicial system, which is characterized by chronic excess demand and a persistent shortage of judges, efficiency gains at the case level do not necessarily translate into lasting reductions in overall processing time. Consistent with this interpretation, the initial efficiency gains generated by the IVT reform appear to

be gradually offset by increased caseload pressure and congestion effects. Although Table **??** shows no increase in the *Rate of Work Overtime*, this measure captures extreme overtime rather than marginal increases in regular working hours and thus may understate the growth in underlying workload. Taken together, the observed "U-shaped" dynamic pattern of adjudication duration suggests that while the IVT reform initially improves efficiency, the resulting expansion in case throughput ultimately absorbs these gains as the system adjusts to binding capacity constraints.

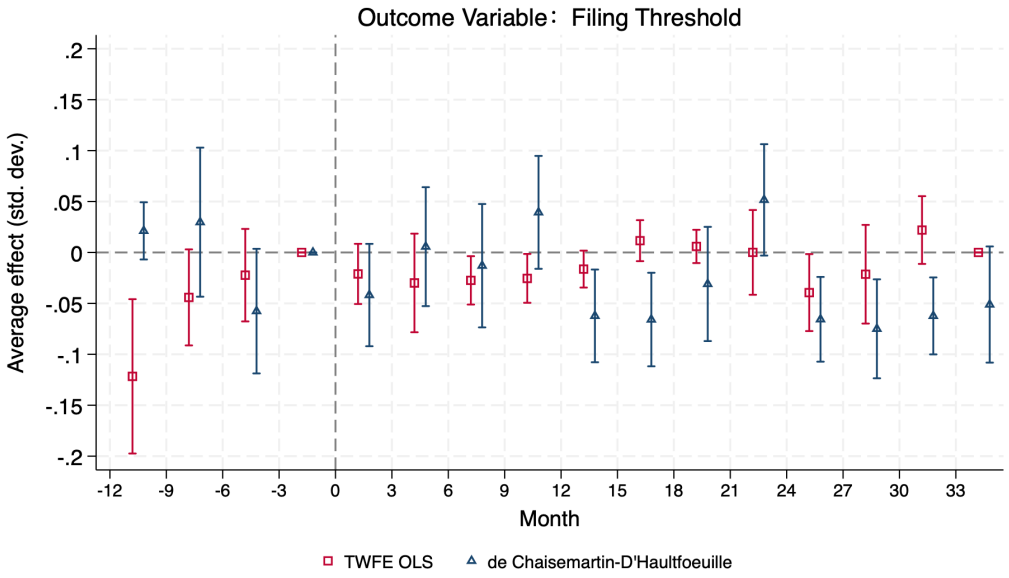
Dynamic impact of the IVT Reform on Workload. Figures 4 and 5 provide further support for the interpretation above. At the court level, Panel (a) of Figure 4 shows that the post-reform reduction in filing thresholds is highly persistent. This pattern is consistent with a "floodgate effect," whereby cases that were previously screened out are increasingly admitted once administrative frictions are lowered, echoing the results in Tables 8 and **??**. Panel (b) of Figure 4 traces the evolution of *work overtime* and reveals a U-shaped pattern that closely mirrors the dynamics of adjudication duration. The reduction in overtime is most pronounced around 9–12 months after the reform, but then gradually dissipates and eventually reverses, turning positive roughly two years later. While Column (5) of Table **??** suggests that extreme overtime does nor increase on average, the dynamic evidence indicates that judicial workload continues to build up over time, particularly in the medium run. At the individual level, Figure 5 documents a steady increase in the number of cases handled per person per month, further reinforcing the floodgate interpretation. In the short run, rising caseloads coincide with falling adjudication duration, reflecting genuine efficiency gains from AI assistance. Over time, however, continued growth in case volume generates congestion and diseconomies of scale, gradually offsetting these gains and pushing adjudication duration back toward its pre-reform level.

Dynamic impact of the IVT Reform on Judgment Quality. Figure 6 presents the dynamic effects of the IVT reform on judgment quality and reveals a similarly transient pattern. The decline in appeal rates observed shortly after the reform gradually attenuates over time, suggesting that initial improvements in case outcomes are not fully sustained as judicial workloads continue to expand.

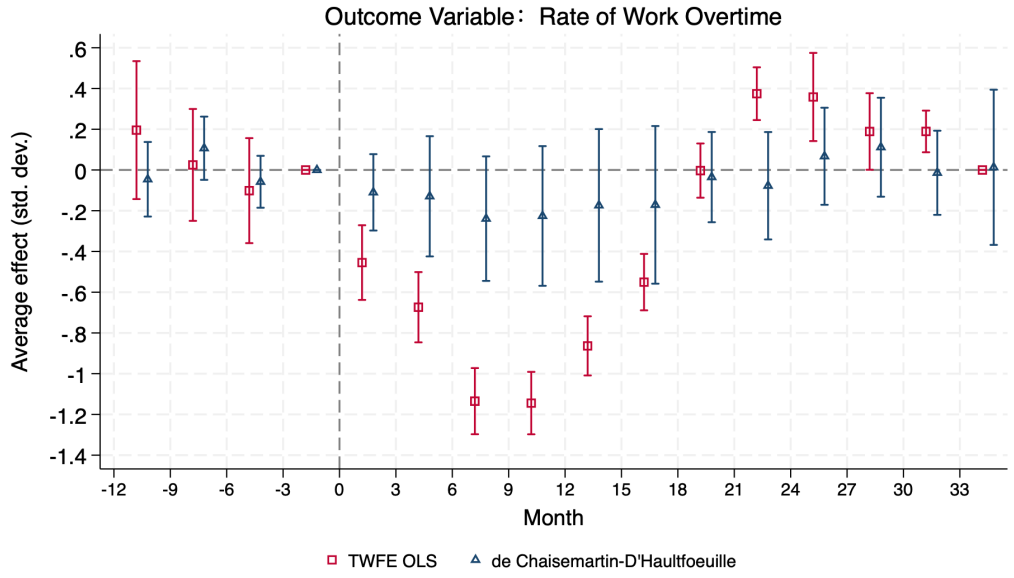
Taken together, these dynamics reveal an equilibrium adjustment: AI-assisted technology raises productivity in the short run, but in judicial systems with rising case inflows and binding capacity constraints, these gains are endogenously absorbed through increased caseloads, resulting in a *busier* equilibrium. As a consequence, short-run improvements in efficiency and quality do not translate into permanent reductions in adjudication duration or sustained improvements in judgment quality in isolation.

7.2 Does IVT Reform Promote Access to Justice?

In recent years, Chinese courts have faced growing caseload pressure amid limited judicial capacity. To cope with overload, courts may tighten filing standards, effectively raising the threshold for case



(a) Filing Threshold

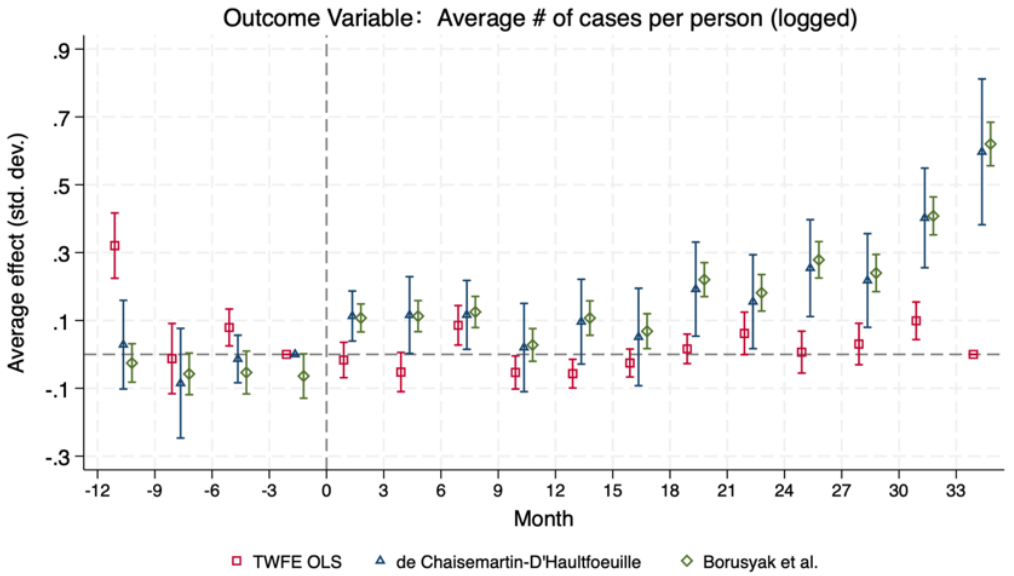


(b) Work Overtime

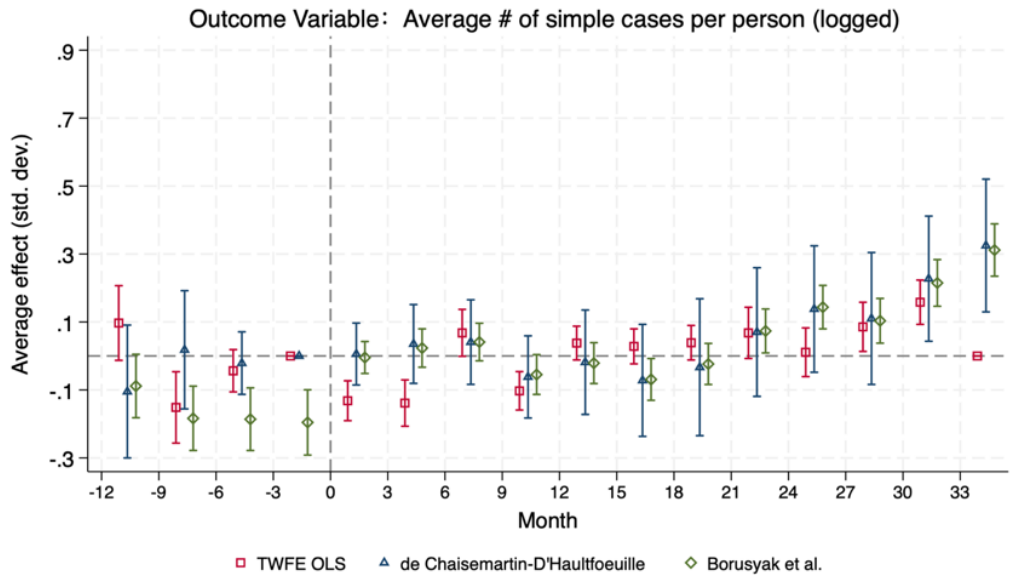
Figure 4. *Dynamic Impact of IVT Reform on Workload (Court Level)*

acceptance and excluding marginal cases from the formal judicial process. Such practices, while relieving short-run judicial workload pressure, can undermine access to justice by preventing disputes from being adjudicated through formal legal channels. If the IVT reform expands courts' processing capacity, a natural implication is that courts can relax filing constraints and accept more cases. In this sense, the reform may promote access to justice by lowering institutional barriers at the entry stage of the judicial process.

To examine this empirically, we construct a court-month–level measure of the case filing threshold that captures the restrictiveness of case acceptance. The variable, *Filing Threshold*, is defined as the ratio of rejected filings to the total number of cases filed in a given court-month, with higher values



(a) *Average # of cases per person (All cases)*



(b) *Average # of cases per person (Simple cases)*

Figure 5. *Dynamic Impact of IVT Reform on Workload (Individual Level)*

indicating stricter filing standards. Table 8 reports the results. As shown in Column (1), the IVT reform significantly lowers the filing threshold, indicating that courts become more willing or able to accept cases after the reform. Column (2) shows that the effect of IVT Reform without Clerk is roughly twice as large as that of IVT Reform with Clerk, suggesting that the second-stage reform more strongly expands case intake. This pattern is consistent with the baseline results and the time-reallocation mechanism: by releasing clerks from in-court duties, the second stage increases available administrative capacity, which in turn permits courts to lower filing thresholds and improve access to justice.

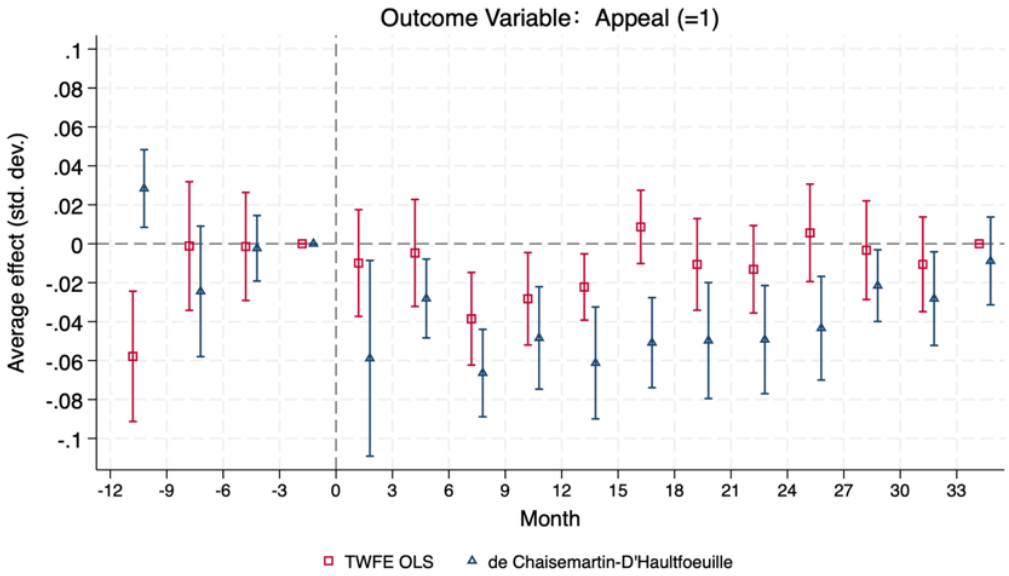


Figure 6. *Dynamic Impact of IVT Reform on Judgment Quality*

Table 8. *Impact of IVT Reform on Threshold for Case Acceptance*

	Filing Threshold	
	(1)	(2)
IVT Reform (=1)	-0.052***	
	(0.018)	
IVT Reform with Clerk (=1)		-0.052***
		(0.018)
IVT Reform without Clerk (=1)		-0.102**
		(0.044)
Year-Month fixed effects	YES	YES
Court fixed effects	YES	YES
Observations	3,763	3,763
R-squared	0.713	0.717

Notes: The estimations are based on court-month level data. Robust standard errors adjusted for clustering at the court-by-year-month level are given in parentheses. ***, **, and * indicate statistical significance at 1%, 5% and 10%, respectively. We construct the indicator of the case filing threshold to capture the restrictiveness of case acceptance at the court-month level. *Filing Threshold* is measured as the ratio of rejection rulings to total cases for each court in a given month. The analyses focus on civil cases, as criminal cases are directly filed by prosecutors and seldom involve rejection, withdrawal, or mediation.

8 Conclusion

With the rapid advancement of AI technologies, their applications in the field of smart judiciary have expanded considerably. Effectively integrating AI across diverse judicial scenarios serves to improve efficiency and quality, bolster fairness and justice, and optimize the allocation of judicial resources.

In this paper, we provide quantitative analyses of the effects of Intelligent Voice Transcription (IVT) and Trial Without Clerk (TWC) reforms on trial efficiency and judicial effectiveness. Our findings reveal that... These findings highlight not only the positive influence of AI technologies in judicial

settings but also the role of human-machine collaboration in improving efficiency.

More importantly, our work offer actionable insights for policymakers, suggesting that AI-assisted reforms can alleviate the "Case Overload and Judge Shortage" dilemma without compromising judicial quality. We propose several policy recommendations to further enhance judicial efficiency and optimize the implementation of AI-assisted reforms in courts:

Declaration of generative AI and AI-assisted technologies in the writing process.

Statement: During the preparation of this work, the authors used Chat-GPT in order to improve the readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

Assegaf, R. (2019). The supreme court: Reformasi, independence and the failure to ensure legal certainty. In T. Lindsey & S. Butt (Eds.), *The politics of court reform: Judicial change and legal culture in indonesia* (pp. 143–168). Routledge.

Assunção, J., Gandour, C., & Rocha, R. (2023). Deter-ing deforestation in the amazon: Environmental monitoring and law enforcement. *American Economic Journal: Applied Economics*, *15*(2), 125–156.

Banerjee, A., Hanna, R., Olken, B., Satriawan, E., & Sumarto, S. (2023). Electronic food vouchers: Evidence from an at-scale experiment in indonesia. *American Economic Review*, *113*(2), 514–547.

Bełdowski, J., Dabrosz, Ł., & Wojciechowski, W. (2020). Judges and court performance: A case study of district commercial courts in Poland. *European Journal of Law and Economics*, *50*(1), 171–201.

Blank, J. L. T., & van Heezik, A. (2020). Policy reforms and productivity change in the judiciary system: A cost function approach applied to time series of the Dutch judiciary system between 1980 and 2016. *International Transactions in Operational Research*, *27*(4), 2083–2104.

Bosio, E. (2023). *A survey of judicial effectiveness: The last quarter century of empirical evidence* (Policy Research Working Paper No. 10501). World Bank.

Brynjolfsson, E. (1993). The productivity paradox of information technology. *Communications of the ACM*, *36*(12), 66–77.

Brynjolfsson, E., & Hitt, L. M. (1996). Paradox lost? firm-level evidence on the returns to information systems spending. *Management Science*, *42*(4), 541–558.

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

Chen, H., Chen, Y., & Yang, Q. (2026). Women in the courtroom: Technology and justice. *Review of Economic Studies*, *93*(3), 1574–1601.

Chen, Z., & Chan, J. (2024). Large language model in creative work: The role of collaboration modality and user expertise. *Management Science*, *70*(12), 9101–9117.

Cohen, I. (2024). *Journal of Public Economics*, *236*, 105154.

Dakolias, M. (1999). Court performance around the world: A comparative perspective. *Yale Human Rights and Development Law Journal*, *2*, 87–142.

Dias, J., & Gomes, C. (2018). Judicial reforms "under pressure": The new map/organisation of the Portuguese judicial system. *Utrecht Law Review*, *14*(1), 68–85.

Doleac, J. (2017). The effects of dna databases on crime. *American Economic Journal: Applied Economics*, *9*(1), 165–201.

Duranton, G., & Turner, M. A. (2011). The fundamental law of road congestion: Evidence from us cities. *American Economic Review*, *101*(6), 2616–2652.

Dzansi, J., Jensen, A., Lagakos, D., & Telli, H. (2022). *Technology and local state capacity: Evidence from ghana*. National Bureau of Economic Research.

El Bialy, N. (2016). The 2007 judicial reform and court performance in Egypt. *Review of Law and Economics*, *12*(1), 69–98.

Glaeser, E. L., & Shleifer, A. (2002). Legal origins. *The Quarterly Journal of Economics*, *117*(4), 1193–1229.

Greenstone, M., He, G., Jia, R., & Liu, T. (2022). Can technology solve the principal-agent problem? evidence from china's war on air pollution. *American Economic Review: Insights*, *4*(1), 54–70.

Guriev, S., Melnikov, N., & Zhuravskaya, E. (2021). 3g internet and confidence in government. *Quarterly Journal of Economics*, *136*(4), 2533–2613.

Jabri, R. (2021). Algorithmic policing [Available at SSRN 4275083].

Jevons, W. S. (1865). *The coal question; an inquiry concerning the progress of the nation, and the probable exhaustion of our coal-mines*. Macmillan; Co.

Jiang, W., Park, J., Xiao, R. J., & Zhang, S. (2025). *Ai and the extended workday: Productivity, contracting efficiency, and distribution of rents* (Working Paper No. 34223). National Bureau of Economic Research.

Jin, G., & Lee, J. (2014). Inspection technology, detection, and compliance: Evidence from florida restaurant inspections. *RAND Journal of Economics*, *45*(4), 885–917.

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

Liu, E., Lu, Y., Peng, W., & Wang, S. (Forthcoming). Judicial Independence, Local Protectionism, and Economic Integration: Evidence from China. *Review of Economics and Statistics*.

Ludwig, J., & Mullainathan, S. (2024). Machine learning as a tool for hypothesis generation. *Quarterly Journal of Economics*, *139*(2), 751–827.

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

Muralidharan, K., Niehaus, P., & Sukhtankar, S. (2016). Building state capacity: Evidence from biometric smartcards in india. *American Economic Review*, *106*(10), 2895–2929.

Nanwani, S. (2016). Multilateral development banks and law and development projects in asia: Experiences and directions. *IIUM Law Journal*, *24*(1), 1–22.

Nissan, E. (2017). Digital technologies and artificial intelligence's present and foreseeable impact on lawyering, judging, policing and law enforcement. *AI & Society*, *32*, 441–464.

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

Robel, L. K. (1990). Caseload and judging: Judicial adaptations to caseload. *Brigham Young University Law Review*, *1990*(1), 3–60.

Rosenn, K. S. (2000). Judicial review in Brazil: Developments under the 1988 constitution. *Southwestern Journal of Law and Trade in the Americas*, *7*(3), 351–378.

Shah, A. V., & Levy, J. Y. (2026). *Access to justice in the age of ai: Evidence from u.s. federal courts* (Working Paper). Massachusetts Institute of Technology.

Vickrey, W. S. (1969). Congestion theory and transport investment. *American Economic Review*, *59*(2), 251–260.

Online Appendix

Smarter Courts, Easier Life for Judges? Evidence from China

by

Hai Ding, Jiaqi Liu, Li'an Zhou, and Yuling Zhu

Appendix A Individual-Level Results

Table A1. *Individual-Level Results*

	Hearing Duration (logged, average)	Adjudication Duration (logged, average)	Appeal Rate
	(1)	(2)	(3)
IVT Reform with Clerk (=1)	-0.083***	-0.077***	-0.007***
	(0.021)	(0.016)	(0.001)
IVT Reform without Clerk (=1)	-0.103***	-0.084***	-0.002
	(0.027)	(0.028)	(0.002)
Year-Month fixed effects	YES	YES	YES
Court fixed effects	YES	YES	YES
Observations	144,281	254,357	502,309
R-squared	0.097	0.100	0.128

Notes: The estimations are derived from civil and criminal cases aggregated at the individual-month level. The dependent variables, *Trial Duration* and *Adjudication Duration*, are measured as the monthly averages. Robust standard errors adjusted for clustering at the individual level are given in parentheses. ***, **, and * indicate statistical significance at 1%, 5% and 10%, respectively.

Appendix B Heterogeneity

Appendix B.1 Heterogeneity Based on Courts' Upstream Pressure: Pre-Reform Caseload

Table B1. *Heterogeneity Based on Courts' Upstream Pressure: Pre-Reform Caseload (2 stages)*

	# of cases (logged)		Night Light Intensity (logged)		Rate of Work Overtime	
	(1) High Caseload	(2) Low Caseload	(3) High Caseload	(4) Low Caseload	(5) High Caseload	(6) Low Caseload
IVT Reform with Clerk (=1)	0.707**	0.280*	-0.102	-0.028	-0.157**	0.021
	(0.285)	(0.161)	(0.062)	(0.042)	(0.062)	(0.051)
IVT Reform without Clerk (=1)	1.194***	0.358	-0.133*	-0.097	-0.206***	-0.042
	(0.294)	(0.290)	(0.072)	(0.064)	(0.056)	(0.059)
Year-Month fixed effects	YES	YES	YES	YES	YES	YES
Court fixed effects	YES	YES	YES	YES	YES	YES
Observations	2,047	1,782	1,757	1,628	1,757	1,628
R-squared	0.653	0.687	0.873	0.908	0.177	0.140

Notes: The estimations are based on court-month level data. Robust standard errors adjusted for clustering at the individual level are given in parentheses. ***, **, and * indicate statistical significance at 1%, 5% and 10%, respectively. *Pre-Reform Caseload* refers to the number of cases handled by the court in 2016. Courts with a caseload above the sample median are classified as *High Caseload*, while those with a caseload at or below the median are classified as *Low Caseload*. Results are similar when *Pre-Reform Caseload* is defined based on 2015 case volumes.

Table B2. *Heterogeneity Based on Courts' Upstream Pressure: Pre-Reform Caseload (Overall)*

	# of cases (logged)		Night Light Intensity (logged)		Rate of Work Overtime	
	(1) High Caseload	(2) Low Caseload	(3) High Caseload	(4) Low Caseload	(5) High Caseload	(6) Low Caseload
IVT Reform (=1)	0.778***	0.276*	-0.108*	-0.024	-0.167***	0.025
	(0.262)	(0.157)	(0.063)	(0.043)	(0.060)	(0.051)
Year-Month fixed effects	YES	YES	YES	YES	YES	YES
Court fixed effects	YES	YES	YES	YES	YES	YES
Observations	2,047	1,782	1,757	1,628	1,757	1,628
R-squared	0.647	0.686	0.873	0.907	0.176	0.139

Notes: The estimations are based on court-month level data. Robust standard errors adjusted for clustering at the individual level are given in parentheses. ***, **, and * indicate statistical significance at 1%, 5% and 10%, respectively. *Pre-Reform Caseload* refers to the number of cases handled by the court in 2016. Courts with a caseload above the sample median are classified as *High Caseload*, while those with a caseload at or below the median are classified as *Low Caseload*.

Appendix C Robustness Checks

This appendix reports the detailed results of the robustness checks summarized in the main text.

Excluding the Influence of COVID-19. The COVID-19 pandemic and associated containment measures substantially disrupted court operations in Shanghai during the first half of 2022, potentially affecting judicial efficiency. This concern is particularly relevant because Shanghai served as the pilot region for the *IVT Reform without Clerk*, which is the primary policy of interest in this study. If pandemic-related disruptions coincided with the implementation of the reform, the estimated treatment effects could reflect external operational shocks rather than the causal impact of AI-assisted judicial reforms.

To address this concern, we first exclude cases adjudicated in 2022 and re-estimate the models. In addition, because the COVID-19 outbreak began in late 2019, we further restrict the sample by excluding all observations after 2019 to eliminate potential confounding effects from the pandemic. As shown in Table B3, the results remain consistent with the baseline estimates, supporting the robustness of our findings.

Table B3. *Exclude the Influence of Covid-19*

	Adjudication Duration (logged)		
	Exclude cases adjudicated in 2022		Exclude post-2019 observations
	(1)	(2)	(3)
IVT Reform (=1)	-0.173***		-0.207***
	(0.025)		(0.028)
IVT Reform with Clerk (=1)		-0.173***	
		(0.025)	
IVT Reform without Clerk (=1)		-0.165***	
		(0.036)	
Year-Month fixed effects	YES	YES	YES
Court fixed effects	YES	YES	YES
Case Characteristic Controls	YES	YES	YES
Observations	524,473	524,473	305,032
R-squared	0.254	0.254	0.217

Notes: All estimations are conducted using the subsample of court documents identified as judgments. Because the *IVT Reform without Clerk* was implemented after 2020—beyond our post-2019 sample window in Column (3)—we do not conduct the two-stage regression. Robust standard errors adjusted for clustering at the court-by-year-month level are given in parentheses. ***, **, and * indicate statistical significance at 1%, 5% and 10%, respectively.

Control for personnel change. It is possible that the estimated effects only reflect personnel changes within courts rather than the behavioral mechanism we propose. If the results are driven by compositional changes in judicial personnel, namely the entry or exit of judges or clerks with systematically different adjudication styles or workloads, rather than by changes in judicial behavior induced by the

reform, our interpretation of the reform's impact would be confounded.

To address this concern, we identify the presiding judge and clerk for each case and include individual fixed effects in all regressions. This specification absorbs time-invariant heterogeneity in judicial personnel, such as differences in professional experience, adjudication styles, and work efficiency, that could otherwise bias the estimates. As reported in Table B4, the estimated effects remain robust after controlling for individual fixed effects, indicating that our findings are unlikely to be driven by personnel reallocation.

Table B4. *Control for Personnel Changes*

	Adjudication Duration (logged)	
	(1)	(2)
IVT Reform (=1)	-0.077***	
	(0.018)	
IVT Reform with Clerk (=1)		-0.077***
		(0.018)
IVT Reform without Clerk (=1)		-0.088***
		(0.025)
Year-Month fixed effects	YES	YES
Court fixed effects	YES	YES
Judge fixed effects	YES	YES
Clerk fixed effects	YES	YES
Case Characteristic Controls	YES	YES
Observations	573,832	573,832
R-squared	0.542	0.542

Notes: The estimations are conducted using the subsample of court documents identified as judgments. Robust standard errors adjusted for clustering at the court-by-year-month level are given in parentheses. ***, **, and * indicate statistical significance at 1%, 5% and 10%, respectively.

Appendix D Impact on Documentation Quality

This appendix reports detailed results on the impact of the IVT reform on the quality of judicial documentation, complementing the judgment quality results in the main text.

We assess documentation quality using generative-AI-based scores along four dimensions: (1) *Language and Expression*, capturing clarity, formality, and coherence; (2) *Fact Presentation*, measuring accuracy, consistency, and objectivity; (3) *Structure and Organization*, assessing completeness and logical flow; and (4) *Format and Consistency*, reflecting adherence to judicial writing conventions. Each dimension is rated on a continuous 1–5 scale, with higher values indicating better quality. The full prompt is provided in Appendix E.

Table B5 reports the results. The reform significantly improves *Language and Expression* (Columns (1)–(2)) and *Fact Presentation* (Columns (3)–(4)): post-reform judgments exhibit clearer language and more accurate, objective factual narratives. By contrast, the effects on *Structure and Organization* and *Format and Consistency* (Columns (5)–(8)) are statistically insignificant. Because these latter dimensions are largely dictated by standardized templates that leave little room for variation, they serve as a natural placebo, indicating that the observed gains reflect substantive improvements in content rather than mechanical differences in formatting.

Table B5. *Impact of IVT Reform on Documentation Quality: Generative AI-Based Scores*

	Text Score of							
	Expression		Fact		Structure		Format	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IVT Reform (=1)	0.007***		0.007**		-0.001		-0.003	
	(0.002)		(0.003)		(0.001)		(0.003)	
IVT Reform with Clerk (=1)		0.007***		0.007**		-0.001		-0.003
		(0.002)		(0.003)		(0.001)		(0.003)
IVT Reform without Clerk (=1)		0.005		0.010**		0.001		-0.002
		(0.004)		(0.004)		(0.002)		(0.004)
Year-Month fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
Court fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
Judge fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
Clerk fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
Case Characteristic Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	631,610	631,610	631,610	631,610	631,610	631,610	631,610	631,610
R-squared	0.268	0.268	0.252	0.252	0.163	0.163	0.210	0.210

Notes: All estimations are conducted using the subsample of court documents identified as judgments. Robust standard errors adjusted for clustering at the court-by-year-month level are given in parentheses. ***, **, and * indicate statistical significance at 1%, 5% and 10%, respectively. Documentation quality scores are generated using DeepSeek-based generative AI.

Taken together, the reform improves substantive aspects of documentation quality—particularly

linguistic clarity and factual accuracy—while leaving template-driven features unchanged. This pattern reinforces the broader conclusion that the IVT reform raises trial efficiency without sacrificing judicial quality.

Appendix E Prompt for Scoring Using Generative AI

You are an assistant for evaluating the writing quality of judicial decisions. Please score each document based on the following standards.Scores must be strictly within the 1-5 range, and decimals (e.g., 2.5, 3.7) are allowed, rounded to one decimal place. 1 = Very poor, 5 = Excellent. Scoring Dimensions and Standards are as follows. **Note:** Do not assign the same score to all documents. You must distinguish quality levels based on actual differences.

1. **Language & Expression**
    - Clarity: Is the language concise and easy to understand?
    - Formality: Does it conform to the formal and professional style of judicial documents?
    - Coherence: Are sentences and paragraphs connected smoothly?
    - Reference for scoring:
        - 1 point = obscure, confusing, or frequent errors
        - 3 points = generally clear but with redundancy or lack of formality
        - 5 points = fluent, concise, professional
2. **Structure & Organization**
    - Completeness: Does it include essential parts such as case facts, reasoning, and judgment result?
    - Orderliness: Is the reasoning logically clear and well-structured?
    - Reference for scoring:
        - 1 point = missing sections or disorganized logic
        - 3 points = basically complete but average in logic
        - 5 points = complete and logically coherent
3. **Fact Presentation**
    - Accuracy: Are the facts consistent throughout?
    - Objectivity: Is the description neutral and free from emotional bias?
    - Reference for scoring:
        - 1 point = frequent errors, serious contradictions
        - 3 points = generally accurate but with vagueness or minor bias
        - 5 points = accurate, consistent, and objective

4. **Format & Consistency**

   - Formatting: Does it follow standard judicial document conventions (paragraphs, numbering, punctuation)?
   - Consistency: Are key details such as parties, dates, and amounts consistent throughout?
   - Reference for scoring:
     - 1 point = chaotic formatting, contradictory information
     - 3 points = basically standard but with minor errors
     - 5 points = fully standardized and consistent

Appendix F Heterogeneity Based on Individual Experience

We next explore heterogeneity by individuals' experience. The IVT reform is not merely a matter of technological substitution; rather, it represents a process of human-machine integration. The effectiveness of AI adoption is likely to depend on judicial personnel's ability to learn, adapt to, and incorporate such AI-assisted tool into their daily workflow. In this sense, experience may play a critical role in shaping the dynamics of human-AI interaction. An important question, therefore, is whether the benefits of AI adoption accrue uniformly across judicial personnel, or whether its impact systematically varies with individuals' professional expertise and familiarity with judicial procedures. We construct proxies for individuals' experience and incorporate them into our empirical specification. As reported in Table B6, individuals with greater *AI Exposure* exhibit substantially larger efficiency gains from the IVT reform, suggesting that accumulated experience with AI tools enhances judicial personnel's ability to effectively integrate the technology into their workflow. A plausible explanation is that repeated interaction with AI-assisted processes improves users' familiarity with the system, reduces coordination frictions, and facilitates smoother human-machine collaboration. These findings highlight the complementary relationship between human expertise and AI tools, underscoring that successful digitalization reforms depend not only on technological capability but also on users' capacity to adapt and learn.

Table B6. *Heterogeneity Based on Individual Experience*

	Adjudication Duration (logged, average)	Hearing Duration (logged)	Appealed (=1)
	(1)	(2)	(3)
IVT Reform with Clerk (=1)	-0.071***	-0.068***	-0.005***
	(0.017)	(0.022)	(0.001)
IVT Reform without Clerk (=1)	-0.053	-0.144***	-0.015***
	(0.046)	(0.039)	(0.003)
AI Exposure	0.054**	0.161***	0.013***
	(0.026)	(0.044)	(0.003)
IVT Reform with Clerk (=1) × AI Exposure	-0.050*	-0.162***	-0.014***
	(0.026)	(0.044)	(0.003)
IVT Reform without Clerk (=1) × AI Exposure	-0.051*	-0.160***	-0.013***
	(0.026)	(0.044)	(0.003)
Year-Month fixed effects	YES	YES	YES
Court fixed effects	YES	YES	YES
Observations	254,357	144,281	502,309
R-squared	0.101	0.097	0.128

Notes: The estimations are derived from civil and criminal cases aggregated at the individual-month level. Robust standard errors adjusted for clustering at the individual level are given in parentheses. ***, **, and * indicate statistical significance at 1%, 5% and 10%, respectively. The variable *AI Exposure* captures an individual's accumulated experience with AI-assisted court processes. It is defined as the number of months elapsed from the individual's first appearance in the sample after the IVT reform was implemented in their court up to month t. This measure enables us to assess how direct and continuous exposure to AI systems shapes judicial efficiency.