THE CONSISTENCY-ACCEPTABILITY DIVERGENCE OF LLMS IN JUDICIAL DECISION-MAKING: TASK AND STAKEHOLDER DIMENSIONS

Mingda Zhang[1] and Qing Xu[2]

[1]School of Software, Yunnan University, Kunming 650500, China
[2]School of Law, Yunnan University, Kunming 650500, China

ABSTRACT

The integration of large language model (LLM) technology into judicial systems is fundamentally transforming legal practice worldwide. However, this global transformation has revealed an urgent paradox requiring immediate attention. This study introduces the concept of "consistency-acceptability divergence" for the first time, referring to the gap between technical consistency and social acceptance. While LLMs achieve high consistency at the technical level, this consistency demonstrates both positive and negative effects. Through comprehensive analysis of recent data on LLM judicial applications from 2023-2025, this study finds that addressing this challenge requires understanding both task and stakeholder dimensions. This study proposes the Dual-Track Deliberative Multi-Role LLM Judicial Governance Framework (DTDMR-LJGF), which enables intelligent task classification and meaningful interaction among diverse stakeholders. This framework offers both theoretical insights and practical guidance for building an LLM judicial ecosystem that balances technical efficiency with social legitimacy.

Keywords Large language models · Judicial decision-making · Consistency-acceptability divergence · Instrumental and value rationality · Technology acceptance · AI governance

1   Introduction

The integration of large language model (LLM) technology into judicial systems represents a fundamental shift in how legal practice operates globally. Since ChatGPT became publicly available, LLM applications in law have grown rapidly, expanding from basic document processing and legal research to core judicial decision-making activities [1]. For instance, the Shenzhen Intermediate Court's integration of LLMs into its decision-making processes marks a significant milestone in practical implementation. This trend has spread to multiple jurisdictions, including the United States, United Kingdom, Colombia, and India, making it a truly global phenomenon [2, 3].

However, LLMs demonstrate clear dual-edged characteristics when pursuing decision consistency. Recent findings from Apple's research team reveal an important insight: artificial intelligence (AI) "reasoning" models, including Claude, DeepSeek-R1, and o3-mini, do not actually perform genuine reasoning. Instead, they rely on pattern memorization [4]. While the Shenzhen Court's implementation achieved notable efficiency gains, it also exposed a critical limitation—the mechanical consistency based on pattern memorization rather than true reasoning. Research shows that GPT-4o functioning as "Judge AI" behaves like a formalistic judge, heavily influenced by precedents but lacking empathy, unlike human judges. This results in typical "machine coldness" characteristics [5]. Although this consistency built on pattern memorization appears impressive in technical metrics, it becomes a major barrier to social recognition and legitimacy.

This excessive focus on consistency has created an acceptability crisis in judicial applications. While LLMs achieve consistency through standardization and patterning, which enhances formal efficiency, this approach triggers widespread social acceptance barriers. The reason is that it neglects the contextual nature and value pluralism inherent in judicial

decision-making. This consistency-acceptability divergence shows systematic characteristics across both task and stakeholder dimensions. In the task dimension, technical tasks gain higher acceptance because their clear rules and verifiable results align well with consistency logic. In contrast, value-judgment-intensive judicial tasks face sharp drops in acceptability due to mechanical consistency. This mechanical approach not only fails to handle case specificity but also produces automation bias—people need to know the correct answer beforehand to identify LLM errors. In the stakeholder dimension, different groups show varied attitudes based on their positions: judges view consistency as threatening judicial independence and show low acceptance; lawyers passively accept it under efficiency pressure while harboring internal doubts; the public experiences the contradiction of recognizing technological potential while lacking substantive trust; and marginalized groups oscillate between hoping for fairness and fearing bias perpetuation. These differentiated patterns reflect how consistency, as instrumental rationality, conflicts with the value rationality held by different groups [6, 7].

Weber's rationalization theory provides a useful framework for understanding this phenomenon. He distinguished between instrumental rationality (pursuing maximum efficiency, standardization, and calculability) and value rationality (pursuing substantive justice based on beliefs). This distinction reveals the "iron cage" dilemma caused by expanding instrumental rationality in modernization [8]. In LLM judicial applications, this tension appears particularly acute. On one hand, LLMs achieve judgment standardization through algorithmic consistency, representing the ultimate instrumental rationality—reducing complex legal reasoning to calculable pattern matching. On the other hand, the essence of justice requires value rationality—understanding case particularity, empathizing with parties' circumstances, and pursuing social justice. This structural conflict between instrumental and value rationality forms the deep root of the consistency-acceptability divergence.

Habermas's communicative rationality theory offers a potential solution for reconciling this tension. Through rational dialogue among different parties, it is possible to acknowledge instrumental efficiency while maintaining value pluralism. This allows technical logic and justice demands to achieve dynamic balance through equal deliberation [9, 10]. However, current practice lacks this communicative bridge. Technical systems optimize consistency metrics following instrumental rationality logic, while social concerns based on value rationality cannot be effectively integrated. Judges fear professional judgment replacement, lawyers sacrifice value considerations under efficiency pressure, and the public distrusts mechanized justice—all manifestations of this conflict. Without communicative rationality mechanisms, consistency's instrumental advantages cannot translate into social acceptance due to lacking value recognition. This ultimately becomes the core obstacle to establishing legitimacy for LLM judicial applications [11].

The absence of communicative rationality mechanisms not only intensifies the conflict between instrumental and value rationality but also fundamentally undermines the social foundation for LLM judicial applications. The core challenge is constructing a social legitimacy foundation—achieving widespread recognition of both technical legitimacy and legality at the social level. The European Union's classification of judicial AI as high-risk systems through the Artificial Intelligence Act, requiring strict data quality control and human supervision, reflects this concern [12]. The lack of legitimacy recognition stems from multiple factors. Technically, LLMs' statistical learning architecture struggles to understand the deep logic and value connotations of legal reasoning. Institutionally, effective cross-group deliberation mechanisms are lacking. Culturally, different judicial traditions lead to varied responses. Only by recognizing how consistency's dual nature produces different effects across tasks and groups, and by constructing governance frameworks that leverage consistency's advantages while avoiding negative impacts, can LLM judicial applications obtain sustainable social foundations [13].

To address these theoretical gaps and practical challenges, this study aims to achieve three core objectives:

First, this study conceptualizes the core paradox in LLM judicial applications as the consistency-acceptability divergence. By integrating the latest empirical research, this study systematically analyzes how this divergence emerges, revealing how consistency undermines substantive justice while enhancing efficiency. This study examines the structural conflict between instrumental and value rationality and establishes a theoretical framework for evaluating this phenomenon.

Second, this study constructs a dual-dimensional analytical framework to reveal the divergence's essence. From the task dimension, this study analyzes differentiated acceptability effects and deep fractures across different judicial tasks. From the stakeholder dimension, this study examines differentiated response patterns and structural constraints among different groups. This approach helps deeply understand the social roots of the divergence phenomenon.

Third, this study constructs LLM judicial governance theory based on social legitimacy foundations. By integrating Weber's rationalization theory and Habermas's communicative rationality framework, this study proposes the Dual-Track Deliberative Multi-Role LLM Judicial Governance Framework (DTDMR-LJGF). This framework represents a paradigm shift from technological determinism to rational pluralistic integration.

2 Results

2.1 Judicial Task Dimension: Cross-Task Value Rationality Differences and Technical Boundaries

AI application acceptance in the judicial field shows significant variation across different tasks, clearly demonstrating consistency's dual-edged effects. To systematically analyze this phenomenon, this study compiled the latest empirical data from major legal industry research institutions and academic institutions between 2024-2025. Table 1 synthesizes acceptance data covering nine judicial task types. Data sources include large-scale surveys from authoritative institutions such as the Clio Legal Trends Report, Thomson Reuters Future of Professionals Report, and Stanford RegLab, as well as practical data from U.S. judicial systems, corporate legal departments, and international law firms.

These data reflect a complete spectrum from highly standardized auxiliary tasks (such as legal research and contract review) to decision-making tasks requiring intensive value judgments (such as judicial decision support and sentencing recommendations). Through systematic compilation of acceptance indicators, rational characteristics, and data sources for each task type, Table 1 demonstrates not only LLMs' differentiated adaptability across judicial tasks but also reveals how consistency produces dramatically different social effects in different contexts. This variation ultimately affects the establishment of legitimacy foundations.

Table 1: Differentiated Analysis of LLM Acceptance in Legal Tasks (2024-2025)

Task Type	Research Content	Key Indicators	Rational Characteristics	Data Sources
Legal Research & Case Analysis	Survey of 1,028 legal professionals analyzing AI adoption growth trends	79% of lawyers use LLM tools, 60% year-over-year increase	Mature technology, controlled risks, significant efficiency gains	[14, 15]
Contract Review	Survey of 800 lawyers and 712 international legal professionals	74% use LLM tools, 92% believe it improves work	High standardization, verifiable accuracy	[16, 17]
Document Review	Using fine-tuned Mistral 7B model to detect racial restrictive covenants	99% recall rate, 100% precision, saving 86,500 person-hours	Cost less than 2% compared to proprietary models	[18]
E-Discovery	Survey of 186 organizations (65% Fortune 500) evaluating Technology-Assisted Review	Only 12% corporate use, but courts widely accept TAR	Mature technology, accuracy superior to manual review	[19]
Transcription Services	Survey of legal professionals from various firm sizes	22.1% of law firms use	Mature technology, continuously improving accuracy	[20]
Legal Consultation & Advice	Survey of 1,200+ international professionals	26% actively use generative AI, but 83% consider AI providing legal advice inappropriate	Involves professional core values, higher ethical risks	[7, 21]
Court Representation	Research on AI application possibilities in court representation	96% believe AI representing clients in court "goes too far"	Interpersonal trust irreplaceable, legal reasoning complexity	[7]
Judicial Decision Support	Survey of 1,000 adults; Fine et al. experimental study with 1,800 participants	Judicial system trust drops to 35%, pure expertise judges significantly preferred	Systematic risk of algorithmic bias, threat to procedural justice	[22, 6]
Sentencing Recommendations	Evaluation of three major legal AI research tools reliability	Professional legal LLM tools show 17%+ hallucination rate (Westlaw AI exceeds 34%)	Prominent bias and errors, fairness controversies	[23]

Based on the analysis of acceptance differentiation across judicial tasks, this study identifies three essential mechanisms. These mechanisms collectively reveal how consistency's dual nature produces differentiated effects across task types, thereby affecting the social legitimacy foundation of LLM judicial applications.

First, this study observes an epistemological fracture: the fundamental incompatibility between knowledge representation and meaning generation. The surface-level task acceptance gradient—from high acceptance of technical tasks to low acceptance of creative tasks—essentially reflects a divide between two fundamentally different ways of understanding knowledge. Technical tasks achieve high acceptance rates, such as legal research (79% lawyer usage), contract review (74% usage rate, 92% believe it improves work), and document processing (99% recall rate, 100% precision). These tasks succeed because they essentially involve knowledge representation—encoding, retrieving, and

recombining established information according to formal rules [24]. In these contexts, consistency becomes a positive force for enhancing efficiency and accuracy.

However, tasks requiring judgment and creativity, such as court representation and judicial decision-making (trust at only 35%), involve meaning generation—creatively constructing new understanding frameworks within specific contexts [22, 6]. Here, consistency becomes a negative factor that limits innovation and individualized judgment. LLM systems can only reproduce existing bias patterns from training data; they cannot generate transcendent justice understanding in new judicial contexts [25].

Second, an ontological gap exists: the categorical leap from computational rationality to practical wisdom. The root of acceptance differentiation lies in a fundamental gap between two forms of rationality required in judicial practice. Technical tasks can be effectively processed by LLMs, such as e-discovery (despite only 12% corporate usage, courts widely accept it) and transcription services (22.1% law firm usage) [19, 20]. These tasks succeed because they can be reduced to computational rationality—optimization processes based on clear rules and quantifiable goals.

In contrast, judicial decision-making requires Aristotelian practical wisdom—the ability to balance multiple values and make prudent judgments in specific contexts [26]. When legal consultation is considered as involving professional core values and court representation is viewed as requiring irreplaceable interpersonal trust [7], consistency's limitations in the realm of practical wisdom become evident. This type of wisdom requires emotional resonance, moral intuition, and cultural understanding—embodied and embedded cognitive abilities that mechanical consistency cannot provide [27]. Concerns about systematic risk of algorithmic bias in judicial decision-making essentially represent an intuitive rejection of replacing practical wisdom with computational rationality.

Third, an axiological paradox emerges: the challenge of legitimacy reconstruction from procedural propriety to substantive justice. Task acceptance differentiation reflects inherent tensions in judicial legitimacy foundations. High acceptance of technical tasks builds upon procedural legitimacy logic—as long as LLM systems follow established procedures and produce verifiable results, their application possesses formal legitimacy.

However, ultimate judicial legitimacy derives from achieving substantive justice. This requires not only correct procedures but also just outcomes, meaningful processes, and socially recognized judgments [28]. The justice threat paradox facing judicial decision support reveals a troubling pattern: the more strictly LLM systems follow formal consistency, the more likely they deviate from substantive justice requirements, thereby undermining their social legitimacy foundation. High hallucination rates in sentencing recommendations (professional legal LLM tools showing 17%+, Westlaw AI exceeding 34%) [23] represent not merely technical defects but reflect the fundamental contradiction between formalized judicial procedures and the pursuit of substantive justice [29].

The three mechanisms above reveal fundamental flaws in current LLM judicial applications: applying a single instrumental rationality logic to all judicial tasks while ignoring different task types' differentiated consistency requirements. Current LLMs adopt a unified probabilistic calculation model for all task types. This rationality monism strategy inevitably leads to systematic failures in value-intensive tasks. The real solution lies in recognizing consistency's dual-edged characteristics and constructing differentiated task processing frameworks accordingly. This means allowing instrumental rationality to fully leverage consistency's efficiency advantages within its effective domain, while protecting value rationality's dominant position in judicial core functions. Ultimately, through rationality pluralism institutional design, it is possible to resolve consistency's dual-edged effects and create conditions for building social legitimacy foundations.

2.2 Stakeholder Dimension: Cross-Group Value Rationality Differentiation and Communication Possibilities

Stakeholder attitudes toward LLM judicial applications present systematic differentiation patterns. This profoundly reflects how consistency's dual nature triggers varied responses among different groups, affecting the establishment of social legitimacy foundations. To comprehensively understand this phenomenon, this study integrated empirical data spanning all levels of the judicial ecosystem from 2023-2025. Table 2 compiles attitude surveys and usage data covering 20 stakeholder groups. Data sources include Thomson Reuters Institute annual judicial reports, American Bar Association technology surveys, Pew Research Center national opinion polls, and behavioral science research from academic institutions including Stanford and Harvard.

Table 2: Value Recognition Differences in LLM Judicial Applications from Cross-Group Perspectives

Group Type	Main Content	Key Findings	Main Concerns	Research Sources
Decision Makers				
Judges (USA)	Survey of 1700+ professionals on generative AI attitudes	9% believe should use generative LLMs; 68% express uncertainty	Dehumanization of justice; judicial independence	[7, 30]
Judges (Shenzhen)	First judicial vertical LLM deployment covering 85 process nodes	Full system deployment for almost all civil and commercial cases	Over-reliance concerns; responsibility attribution unclear	[31, 2]
Judges (UK Welsh)	Judicial Attitude Survey with 94% salaried judges participating	Open but cautious attitude toward LLM assistance	Manipulation risks; need for clear usage guidelines	[32]
Lawyers (International)	Survey of 440+ lawyers from large and medium firms on ChatGPT	82% believe applicable, but only 3% actually using generative LLMs	Accuracy concerns (62%); data security and client privacy	[15]
Law Firms	Research on firm-level AI policy development and management	15% of firms warn against unauthorized use; only 6-41% have LLM policies	Efficiency pressure vs. replacement fears; need clear policies	[15, 33]
Lawyers (USA)	Analysis based on tens of thousands of user data and 1000+ responses	79% use LLMs (increased from 19% in 2023); 74% hourly work automatable	Concern about impact on revenue model (hourly billing)	[14]
Lawyers (Florida)	Florida Bar member opinion survey on generative AI	80% don't use generative LLMs; 79% believe "very strict regulation" needed	Cannot replace critical thinking and client explanation ability	[34]
Legal Professionals	Multiple industry surveys analyzing legal professionals' AI acceptance	54% list security issues as major implementation barrier	Job replacement fears; balancing benefits with ethical concerns	[35]
Public and Court Users				
Public (USA)	National survey of 10,701 adults on ChatGPT trust and usage	Only 2% highly trust ChatGPT for election info; 52% more worried than excited	Lack of empathy and emotional understanding	[36]
Court User Public	NCSC & Thomson Reuters research on AI opportunities in courts	Over 70% support auxiliary functions; only 27% believe courts provide adequate help	Functional acceptance vs. decisional resistance contradiction	[37]
General Public (Trust)	Study on public trust in AI-assisted judges by race/ethnicity	Pure expertise judges most trusted; AI-assisted judges less trusted	AI seen as amplifying existing biases; procedural justice concerns	[6]
Vulnerable Groups				
Minority Groups	AI impact on minority communities' civil rights	Enforcement agencies' AI decisions disproportionately impact immigrants/minorities	AI perpetuating historical biases; lack of transparency	[38]
Elderly (Judicial)	First case using AI-generated deceased victim video in court	Low acceptance for AI use in court proceedings	Deepfake technology abuse concerns; loss of human element	
Low-Income Groups	DoNotPay helped overturn $4M+ fines but fined for false advertising	AI tools help save millions but quality concerns remain	New inequality if high-quality AI legal services require payment	[39]
Crime Victims	AI-generated deceased victim video used in sentencing hearing	First use may set legal precedent; victim rights groups cautious	Consent and ethical issues; authenticity questions	[40]
Experts and Special Groups				
Legal Scholars	Research on generative AI legal ethics and practical applications	AI will become core legal service technology; existing ethical rules sufficient	"Hallucination" risks; over-reliance weakening independent judgment	[41]
Compliance Officers	Survey on corporate legal departments' AI compliance confidence	Only 40% highly confident in LLM-related judicial compliance	IP disputes; deepfake risks in litigation; unclear norms	[42]
Disabled Persons	AI systems need inclusive design to serve all populations	Court LLM tools often lack accessibility design features	Cannot understand and accommodate special needs in proceedings	

Based on the systematic analysis of stakeholder groups across multiple judicial domains, this study finds that the consistency-acceptability divergence in LLM judicial applications stems from three structural constraints. These constraints collectively reveal how consistency's dual nature hinders the establishment of communicative rationality mechanisms, thereby affecting the construction of social legitimacy foundations.

First, hierarchical power distribution produces a power-acceptance inverse effect. Decision-making power levels show strong conservative tendencies, with only 9% of U.S. judges believing LLMs should be used and 68% expressing uncertainty [7, 30]. Although the Shenzhen Court fully deployed an intelligent trial system covering 85 process nodes, management still expresses deep concerns [31, 2]. Among UK Welsh judges, 94% participated in attitude surveys but maintain cautious positions [32].

Professional power levels exhibit huge gaps between cognition and behavior. While 82% of international lawyers believe LLMs are applicable, only 3% actually use them [15]. U.S. lawyer usage grew from 19% in 2023 to 79%, yet 80% of Florida lawyers still don't use them [14, 34]. Executive power levels show significantly lagging institutionalized responses, with 15% of law firms issuing warnings against unauthorized use and only 10%-41% establishing LLM usage policies [15, 14].

Audience power levels demonstrate separation between expectations and trust. While 70% of the public supports auxiliary functions, only 2% highly trust LLM key decisions [36, 37]. This gradient difference embodies the power protection mechanism explained by Bourdieu's field theory—groups with higher power positions increasingly view consistency as erosion of their professional autonomy. They tend to maintain existing power structure stability through institutional resistance strategies [43].

Second, uneven distribution of cultural capital forms a professional knowledge-technical concern positive correlation pattern. Legal professionals universally show high concern levels: 54% list security issues as major barriers [35], 79% believe in very strict regulation needs [34], legal scholars insist on the centrality of human judgment [41], and only 40% of corporate compliance officers feel highly confident in LLM judicial compliance [42].

In contrast, the general public presents contradictory attitudes, with 52% more worried than excited. Special groups like people with disabilities, low-income groups, and the elderly face more specific difficulties [36, 39]. This difference embodies the structural tension between professional rationality and technical rationality within Weber's rationalization theory framework. Professional groups, drawing on their deep legal knowledge and accumulated experience, can more keenly identify consistency's inherent limitations in complex legal reasoning, value judgments, and ethical considerations. This leads them to generate technical caution based on professional responsibility [44].

Third, cultural solidification of value habitus leads to cross-regional culture-technology adaptability differentiation. China demonstrates ambition for large-scale technical practice, with the Shenzhen Court becoming the first to integrate LLMs into judicial decision-making on a large scale, covering almost all civil and commercial cases [31, 2]. The United States presents internal differentiation patterns, with 58% of judges uncertain about usage [7], while 74% of lawyers' hourly billing work faces automation threats [14]. The UK maintains a prudent attitude, with judges holding open but cautious positions toward LLMs in trial assistance [32].

These regional differences reflect varying understandings of consistency value under different judicial traditions and institutional environments. China's collective interest priority and efficiency-oriented culture views consistency as positive value. America's individual rights protection tradition views consistency as a potential threat. Britain's incremental reform tradition seeks balance between the two [2].

The interaction of these three constraints leads to the absence of communicative rationality mechanisms, preventing consistency's technical advantages from translating into social recognition. Instead, they become obstacles to establishing legitimacy foundations. This manifests in three aspects: institutional supply-demand imbalance between regulatory needs and policy provision, with 60% completely unaware of LLM detection tools [34]; cognitive-behavioral deviation between acceptance and practice, with 51% believing in application but only 3% actually using [15]; and attitude differentiation based on field positions rather than common value pursuits. Crime victims worry about LLM evidence authenticity [40], while disabled persons face technological exclusion.

Due to the lack of situations emphasizing equal participation, mutual understanding, and consensus orientation as outlined in communicative rationality theory, stakeholders respond strategically based on partial understandings of consistency rather than seeking communicative understanding based on common interests and value consensus. This prevents effective mediation of the fundamental tension between formal rational consistency pursued by technical systems and value rational pluralism upheld by social groups through democratic deliberation. Ultimately, this leads to the solidification and institutional reproduction of the consistency-acceptability divergence phenomenon.

3 Discussion

Through systematic analysis, this study reveals the core paradox of consistency-acceptability divergence in LLM judicial decision-making. This study finds that this divergence stems from structural tensions between consistency's dual characteristics and value rationality requirements. As illustrated in Figure 1, this divergence manifests through two key dimensions: in the task dimension, it appears as three mechanisms—epistemological fracture, ontological gap, and axiological paradox; in the stakeholder dimension, it emerges as three constraints—power-acceptance inverse effect, professional knowledge-technical concern positive correlation, and culture-technology adaptability differentiation [6, 45]. The theoretical framework presented in Figure 1 demonstrates how these dual dimensions converge to create the fundamental structural conflict between instrumental rationality (efficiency pursuit) and value rationality (justice quest). Theoretically, proposing the consistency-acceptability divergence conceptual framework holds innovative significance. The analytical perspective of this study integrating Weber's rationalization theory and Habermas's communicative rationality theory reveals the possibility of achieving dynamic balance between instrumental and value rationality through communicative approaches. Practically, the findings of this study provide differentiated AI application strategy guidance for judicial institutions. For technical tasks, institutions should fully leverage consistency's efficiency advantages. For value judgment tasks, they must establish strict human supervision mechanisms. Through the DTDMR-LJGF framework, institutions can construct multi-stakeholder deliberation mechanisms. This research also warns of risks from the marginalized group paradox and forced adoption phenomena, providing direct guidance for formulating balanced, prudent, human-centered judicial AI policies [46].

Figure 1 | Consistency-acceptability divergence in judicial LLMs

Theoretical framework showing the structural tension between technical consistency and social acceptability in judicial large language models, manifesting through task-specific (epistemological, ontological, axiological) and stakeholder-specific (power, cultural, habitus) mechanisms.
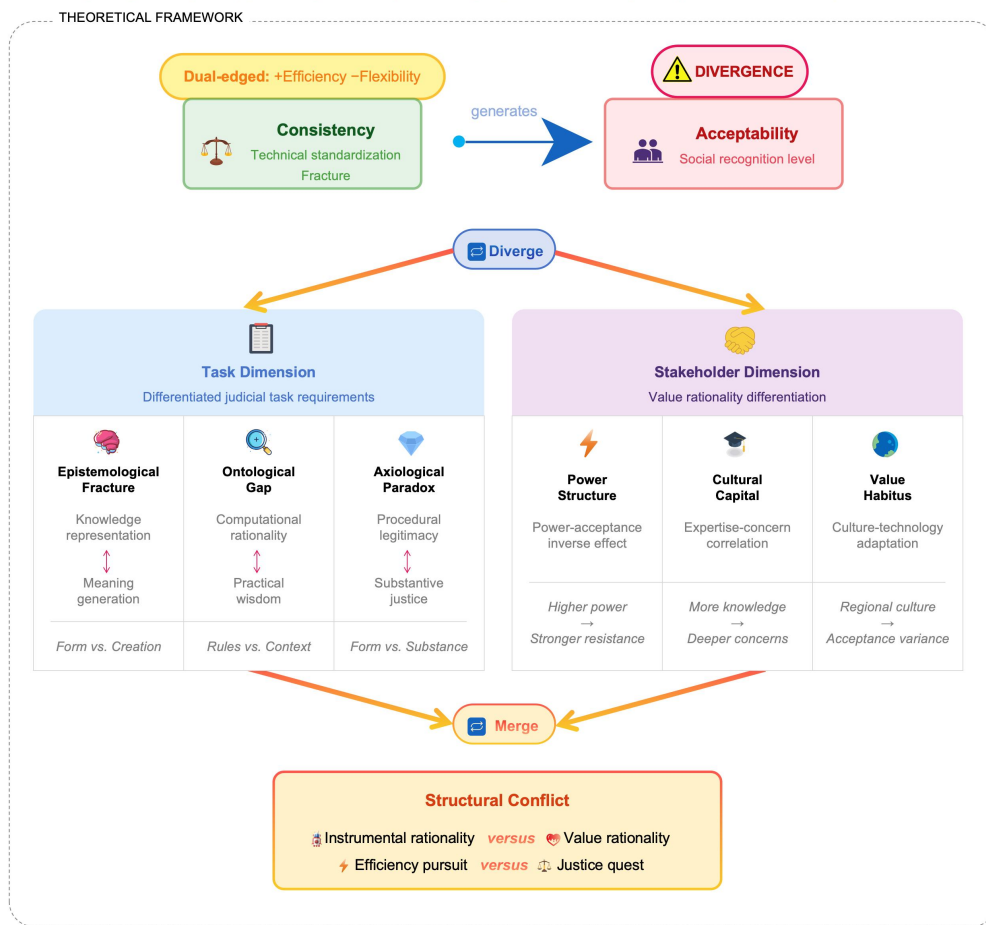
Figure 1: Consistency-acceptability divergence in judicial LLMs. The theoretical framework showing the structural tension between technical consistency and social acceptability in judicial large language models, manifesting through task-specific (epistemological, ontological, axiological) and stakeholder-specific (power, cultural, habitus) mechanisms.

To address the consistency-acceptability divergence, this study proposes the Dual-Track Deliberative Multi-Role LLM Judicial Governance Framework (DTDMR-LJGF). This framework fundamentally reshapes LLM judicial governance logic. As shown in Figure 2, it achieves precise task routing through an intelligent routing layer. Procedural tasks are directed to the formal rationality track for efficient processing, while value judgment tasks activate the substantive rationality track's multi-role deliberation mechanism (including judge agents, lawyer agents, and jury agents). The framework innovatively introduces a dynamic context interaction interface as a bidirectional interactive value calibration space for deep human-machine integration. Relying on a three-layer dynamic architecture—input processing layer, dual-track processing system, and integrated validation system—it maintains consistency advantages while preserving value pluralism. Through differentiated processing strategies using shadow jury mechanisms and rapid correction mechanisms, it achieves the practical implementation of communicative rationality within technical systems. This provides an operational solution that leverages technical advantages while maintaining judicial values for establishing social legitimacy foundations [47].

In terms of research perspective and solutions, this paper achieves important breakthroughs. By taking consistency's dual characteristics as the core analytical concept and constructing a task-stakeholder dual-dimensional analytical framework, this study achieves for the first time a three-dimensional holistic understanding of LLM judicial applications. This breaks through the fragmented landscape of technical performance evaluation, social acceptance research, and ethical analysis [25].

Through systematic integration of multi-source data from 2023-2025, this study constructs the most comprehensive empirical dataset of LLM judicial applications to date. Cross-validation reveals differentiated effect patterns of consistency across different contexts [2]. Theoretically, by tracing surface phenomena to deep tensions between instrumental and value rationality, the DTDMR-LJGF framework represents a fundamental reconstruction of AI system architecture based on communicative rationality theory. By embedding multi-stakeholder deliberation processes that incorporate diverse values into AI operational mechanisms, it indicates a technically feasible and ethically sound direction for judicial AI development.

However, this study has several limitations. First, there is geographical and cultural imbalance in data collection, with samples mainly concentrated in developed regions of North America, Europe, and East Asia. Insufficient attention to developing regions may limit the full capture of consistency's differentiated manifestations across cultural contexts. Second, the temporal window concentrated in 2023-2025 makes it difficult to observe long-term evolution patterns. Third, methodological constraints from original research limitations mean successful cases are more easily cited while failure experiences often remain hidden. Fourth, theoretically, translating communicative rationality from philosophical theory into specific technical design and institutional arrangements requires further exploration. These limitations clarify the research's scope while indicating directions for future research: expanding geographical coverage, conducting long-term tracking, establishing standardized evaluation systems, and advancing theory-practice translation.

4   Materials and Methods

This study comprehensively employed interdisciplinary perspectives and methods crossing legal philosophy and artificial intelligence computational science. Under this framework, this study proposed the theoretical innovation of consistency-acceptability. Building on this foundation, this study adopted systematic literature analysis to examine empirical research and practical data on LLM judicial application acceptance published between 2023 and 2025 [48].

Literature searches covered major academic databases including Web of Science, Scopus, Google Scholar, and SSRN. This study used combinations of keywords such as "large language models," "LLMs," "ChatGPT," "AI" with "judicial," "legal," "court," "justice," "acceptance," "adoption," and "perception" [49].

After systematic screening [50], the final dataset integrated four main source types: annual survey reports from industry research institutions (Thomson Reuters Institute, Clio, Bloomberg Law, American Bar Association, etc.); empirical research from academic institutions (Stanford RegLab, Harvard Law School, University of Pennsylvania, etc.); judicial practice data (Shenzhen Intermediate Court, U.S. Federal Courts, UK Welsh Courts, etc.); and public opinion surveys (Gallup, Pew Research Center, National Center for State Courts, etc.). Data collection methods included questionnaire surveys, judicial case analysis, controlled experiments, and in-depth interviews. These multi-source data collectively constitute the empirical foundation for analyzing LLM judicial application acceptance.

Data availability: All data analyzed are from publicly available sources cited in the references.

Figure 2. Architecture of Dual-Track Deliberative Multi-Role LLM Judicial Governance Framework (DTDMR-LJGF)

A novel framework integrating formal and substantive rationality for enhanced judicial AI governance

📤 INPUT PROCESSING LAYER — Entry
Integrated Case Information Processing

🔀 INTELLIGENT ROUTING LAYER — Decision

| **Complexity** Task Classification | **Value Density** Ethical Assessment | **Resources** Optimization |

Technical Tasks → Formal Track | Value Jugement Tasks → Substantial Track

Dual-Track Processing System

⚡ FORMAL RATIONALITY
Maximize Consistency

Task Types:
- Document Analysis
- Evidence Classification
- Procedural Compliance
- Pattern Matching

Metrics:
√ Speed √ Consistency √ Accuracy

👥 SUBSTANTIAL RATIONALITY
Avoid Negative Effects

⚖️ Judge Agent
Principles: Independence, Justice
Chain: Law→Facts→Application→Values

👨‍💼 Lawyer Agents
Setup: Prosecution & Defense
Function: Adversarial Reasoning

🏛️ Jury Agent
Base: Common Sense & Customs
Role: Social Values Balance

🔁 DYNAMIC CONTEXT INTERACTION — H-AI

| **📝 Structured** Context Input | **📖 Narrative** Details | **⚡ Values** Calibration | **👀 Feedback** Real-time |

Adaptive Strategy:
- High Context → **Shadow Jury** validation
- Routine → **Quick Correction** protocol

✅ INTEGRATED VALIDATION — QA

| **Cross-track** Consistency Check | **Stakeholder** Value Alignment | **Balance** Efficiency-Pluralism |

Technical Tasks → Formal Track | Value-laden Tasks → Substantial Track

📤 FINAL OUTPUT — Result

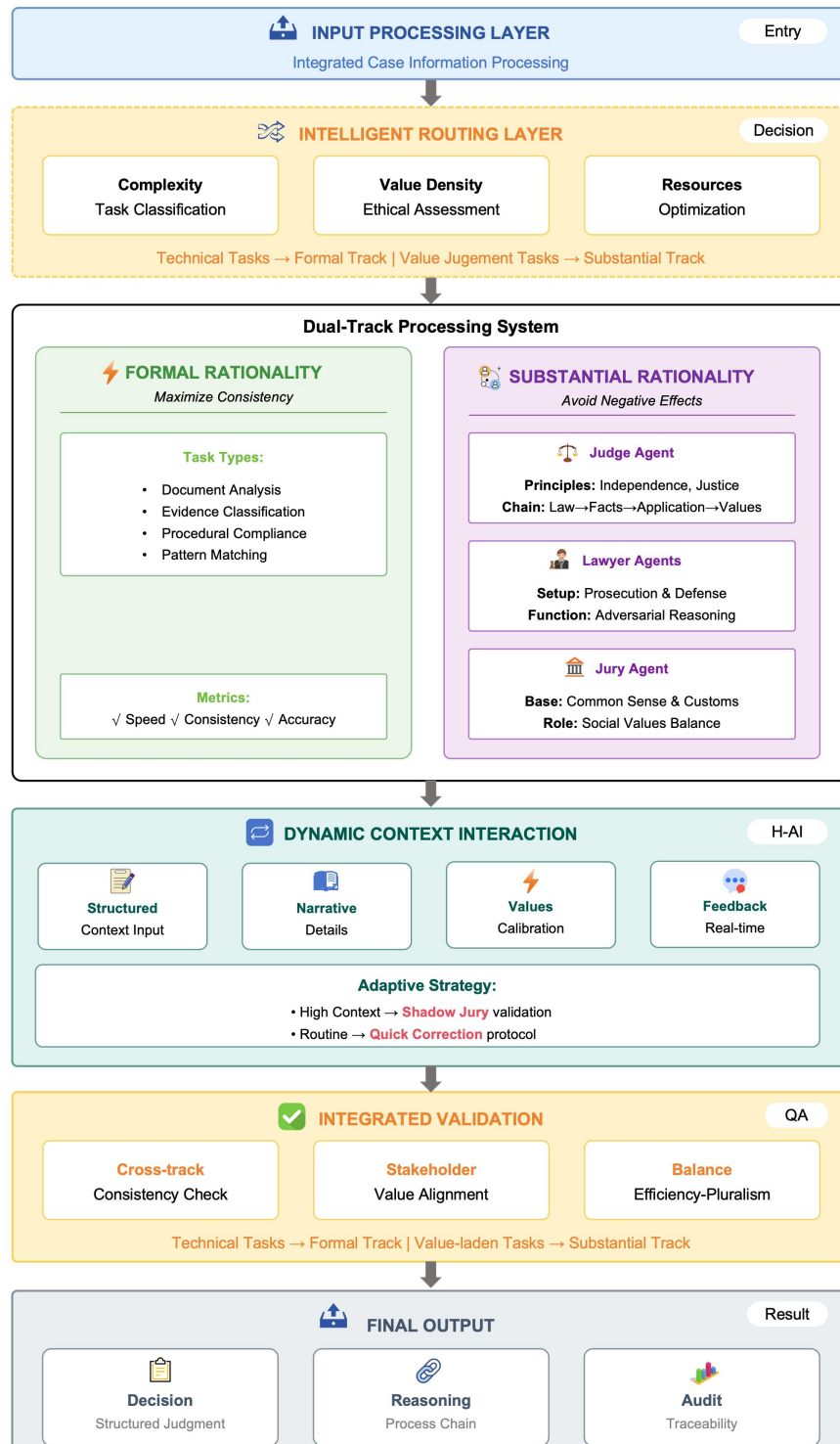| **📋 Decision** Structured Judgment | **🔗 Reasoning** Process Chain | **📊 Audit** Traceability |

Figure 2: Architecture of Dual-Track Deliberative Multi-Role LLM Judicial Governance Framework (DTDMR-LJGF). A novel framework integrating formal and substantive rationality for enhanced judicial AI governance.

References

[1] Thomson Reuters Institute, Georgetown Law Center on Ethics and the Legal Profession, 2025 report on the state of the US legal market. Technical report (2025). https://www.thomsonreuters.com/en-us/posts/legal/state-of-the-us-legal-market-2025 (accessed 17 June 2025).

[2] J. Z. Liu, X. Li, How do judges use large language models? Evidence from Shenzhen. *J. Legal Anal.* **16**, 235–262 (2025).

[3] Council of Europe, European ethical charter on the use of artificial intelligence in judicial systems and their environment (Updated guidance). Technical report (2024). https://www.coe.int/en/web/cepej/ (accessed 17 June 2025).

[4] Apple Machine Learning Research, The illusion of thinking: Understanding the strengths and limitations of reasoning models via the lens of problem complexity. arXiv [Preprint] (2025). https://doi.org/10.48550/arXiv.2506.06941 (accessed 17 June 2025).

[5] E. A. Posner, S. Saran, Judge AI: Assessing large language models in judicial decision-making. SSRN. https://doi.org/10.2139/ssrn.5068825. Deposited 15 January 2025.

[6] A. Fine, E. R. Berthelot, S. Marsh, Public perceptions of judges' use of AI tools in courtroom decision-making: An examination of legitimacy, fairness, trust, and procedural justice. *Behav. Sci.* **15**, 476 (2025).

[7] Thomson Reuters Institute, 2024 report on the state of the U.S. legal market. Technical report (2024).

[8] M. Weber, *Economy and Society: An Outline of Interpretive Sociology*, G. Roth, C. Wittich, Eds. (Bedminster Press, 1968).

[9] J. Habermas, *The Theory of Communicative Action: Vol. 1. Reason and the Rationalization of Society*, T. McCarthy, Trans. (Beacon Press, 1984).

[10] J. Habermas, *The Theory of Communicative Action: Vol. 2. Lifeworld and System: A Critique of Functionalist Reason*, T. McCarthy, Trans. (Beacon Press, 1987).

[11] J. Morison, A. McInerney, Artificial intelligence at the bench: Legal and ethical challenges of informing—or misinforming—judicial decision-making through generative AI. *Data Policy* **6**, e8 (2024).

[12] EU Artificial Intelligence Act, Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence. *Off. J. Eur. Union* **L 1689**, 1–144 (2024).

[13] Creative Rights Institute, 2024 blueprint of global AI legislative policy efforts: A comprehensive analysis of worldwide AI governance frameworks. Technical report (2024). https://www.creativerightsinstitute.com/reports/2024aiblueprint (accessed 17 June 2025).

[14] Clio, 2024 legal trends report. Technical report (2024). https://www.clio.com/resources/legal-trends/2024-report/ (accessed 17 June 2025).

[15] Thomson Reuters Institute, Law firm financial index Q4 2024 report. Technical report (2024).

[16] Ironclad, State of AI in Legal Report 2024. Technical report (2024). https://www.ironcladapp.com/ (accessed 17 June 2025).

[17] Wolters Kluwer Legal & Regulatory, 2024 future ready lawyer survey, 6th ed. Technical report (2024). https://www.wolterskluwer.com/en/solutions/legal-regulatory/resources/future-ready-lawyer-survey-2024 (accessed 17 June 2025).

[18] F. Surani *et al.*, AI for scaling legal reform: Mapping and redacting racial covenants in Santa Clara County. Stanford RegLab. https://reglab.github.io/racialcovenants/ (accessed 17 June 2025).

[19] Corporate Legal Operations Consortium, CLOC Global Institute 2024: Recharge - Embracing experimentation in legal ops. Technical report (2024). https://cloc.org/cgi (accessed 17 June 2025).

[20] Rev, AI in law: Insights from the 2025 legal technology survey. Technical report (2025).

[21] Thomson Reuters Institute, Future of professionals report. Technical report (2025).

[22] Gallup, 2024 Bentley-Gallup Business in Society Report. Technical report (2024). https://www.gallup.com/ (accessed 17 June 2025).

[23] V. Magesh *et al.*, Hallucination-free? Assessing the reliability of leading AI legal research tools. arXiv [Preprint] (2024). https://doi.org/10.48550/arXiv.2405.20362 (accessed 17 June 2025).

[24] J. H. Choi, D. Schwarcz, AI tools for lawyers: A practical guide. *Minn. Law Rev. Headnotes* **108**, 1–65 (2023).

[25] X. Jiang *et al.*, A survey on LLM-as-a-judge. arXiv [Preprint] (2024). https://arxiv.org/abs/2411.15594 (accessed 17 June 2025).

[26] H. Y. Jabotinsky, M. Lavi, AI in the courtroom: The boundaries of RoboLawyers and RoboJudges. SSRN. https://ssrn.com/abstract=4883326. Deposited 2025.

[27] A. Blair-Stanek, N. Holzenberger, B. Van Durme, Towards robust legal reasoning: Harnessing logical LLMs in law. arXiv [Preprint] (2025). https://arxiv.org/html/2502.17638v1 (accessed 17 June 2025).

[28] Y. A. Arbel, Judicial economy in the age of AI. SSRN. https://ssrn.com/abstract=4873649. Deposited 2025.

[29] State v. Loomis, 881 N.W.2d 749 (Wis. 2016), cert. denied, 137 S. Ct. 2290 (2017).

[30] American Bar Association, Year I report on the impact of AI on the practice of law. Technical report (2024).

[31] Shenzhen Intermediate People's Court, Shenzhen Intermediate Court's "AI-assisted trial system" launched today. *Shenzhen News Network* (2024). https://www.sznews.com/news/content/2024-06/28/content_31044219.htm (accessed 17 June 2025).

[32] C. Thomas, 2024 UK judicial attitude survey. Technical report (2024).

[33] International Legal Technology Association, 2024 Technology Survey. Technical report (2024). https://www.iltanet.org/resources/publications/surveys/ts24 (accessed 17 June 2025).

[34] The Florida Bar, 2024 membership opinion survey on artificial intelligence. *The Florida Bar News* (2024). https://www.floridabar.org/news/tfb-news/ (accessed 17 June 2025).

[35] Bloomberg Law, 2024 legal ops and tech survey. Technical report (2024). https://www.bloomberglaw.com/bloomberg-law-research-reports (accessed 17 June 2025).

[36] Pew Research Center, How the U.S. public and AI experts view artificial intelligence. Technical report (2024). https://www.pewresearch.org/internet/2024/08/28/how-the-u-s-public-and-ai-experts-view-artificial-intelligence/ (accessed 17 June 2025).

[37] National Center for State Courts, AI Policy Consortium for Law and Courts: Opportunities and challenges of artificial intelligence in judicial systems. Technical report (2024). https://www.ncsc.org/ai-policy-consortium (accessed 17 June 2025).

[38] Brennan Center for Justice, An agenda to strengthen U.S. democracy in the age of AI. Technical report (2024). https://www.brennancenter.org/our-work/research-reports/agenda-strengthen-us-democracy-age-ai (accessed 17 June 2025).

[39] Norton Rose Fulbright, 19th annual litigation trends survey. Technical report (2024). https://www.nortonrosefulbright.com/en/knowledge/publications/ai-legal-implications (accessed 17 June 2025).

[40] G. E. Marchant, Navigating AI in court systems: Ethics, legal frameworks, and practical tools. Technical report (2024).

[41] A. M. Perlman, The legal ethics of generative AI. SSRN. https://ssrn.com/abstract=4735389. Deposited 22 February 2024.

[42] Legal Dive, Legal teams putting AI use cases to test as trust concerns persist (2024). https://www.legaldive.com/ (accessed 17 June 2025).

[43] P. Bourdieu, The force of law: Toward a sociology of the juridical field. *Hastings Law J.* **38**, 805–853 (1987).

[44] T. Kim, W. Peng, Do we want AI judges? The acceptance of AI judges' judicial decision-making on moral foundations. *AI Soc.*, in press.

[45] ITM Web of Conferences, Reducing judicial inconsistency through AI: A review of legal judgement prediction models in *ITM Web Conf.* **73**, 02009 (2025).

[46] P. W. Grimm, C. Coglianese, M. R. Grossman, AI in the courts: How worried should we be? *Judicature* **107** (2024). https://judicature.duke.edu/articles/ai-in-the-courts-how-worried-should-we-be/ (accessed 17 June 2025).

[47] X. Liu, Y. Hou, Analyzing the justification for using generative AI technology to generate judgments based on the virtue jurisprudence theory. *Inf. Commun. Technol. Law* **33** (2024).

[48] N. Fabiano *et al.*, How to optimize the systematic review process using AI tools. *JCPP Adv.* **4**, e12234 (2024).

[49] A. Alshami *et al.*, Harnessing the power of ChatGPT for automating systematic review process: Methodology, case study, limitations, and future directions. *Systems* **12**, 164 (2024).

[50] Q. Khraisha *et al.*, Can large language models replace humans in systematic reviews? Evaluating GPT-4's efficacy in screening and extracting data from peer-reviewed and grey literature in multiple languages. *Res. Synth. Methods* **15**, 616–626 (2024).