

Review

Open Access



From single-agent to multi-agent: a comprehensive review of LLM-based legal agents

Se Yang, Zhe Yang, Yutong Liu, Hongtao Wang 

Department of Computer, North China Electric Power University, Baoding 071003, Hebei, China.

Correspondence to: Prof. Hongtao Wang, Department of Computer, North China Electric Power University, Baoding 071003, Hebei, China. E-mail: wanght@ncepu.edu.cn

How to cite this article: Yang, S.; Yang, Z.; Liu, Y.; Wang, H. From single-agent to multi-agent: a comprehensive review of LLM-based legal agents. *AI Agent* 2025, 1, 4. <https://dx.doi.org/10.20517/aiagent.2025.06>

Received: Sep 27 2025 **First Decision:** 9 Oct 2025 **Revised:** 17 Oct 2025 **Accepted:** 22 Oct 2025 **Published:** 18 Nov 2025

Academic Editor: Tongliang Liu **Copy Editor:** Shu-Yuan Duan **Production Editor:** Shu-Yuan Duan

Abstract

With the growing application of artificial intelligence (AI) in the legal domain, large language model (LLM)-based legal agents have achieved remarkable progress. This survey provides a comprehensive review of the applications and developments of LLM-driven agents in law. Firstly, we outline the core legal tasks, including legal information retrieval, question answering, judgment prediction, and legal text generation, along with the corresponding evaluation benchmarks. Then, we analyze the technical challenges faced by both single-agent and multi-agent systems in legal scenarios and summarize the prevailing research methods. Finally, we discuss future directions for legal agents, including enhancing single-agent trustworthiness through explainability, boosting multi-agent efficiency with collaborative AI techniques, enabling cross-jurisdictional interoperability via legal knowledge graphs, and establishing ethical governance with quantifiable metrics. By synthesizing existing research, this survey aims to offer theoretical insights and practical guidance for the sustainable advancement of legal agents.

Keywords: Legal agent, legal task, multi-agent systems, legal artificial intelligence, agent-based simulation

INTRODUCTION

In recent years, large language models (LLMs) have achieved rapid advancements in the fields of natural language processing (NLP) and artificial intelligence (AI). LLMs have demonstrated remarkable capabilities in language understanding and reasoning, driving the development of agent technologies across various application scenarios such as code generation^[1], medical decision-making and services^[2], and news veracity



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



detection^[3]. These advances provide a crucial foundation for intelligence in interdisciplinary domains, and also bring new opportunities to traditional fields that heavily rely on specialized expertise.

In the legal domain, texts exhibit high levels of specialization and require complex, multi-layered reasoning, posing significant challenges for intelligent agent systems^[4]. Existing research primarily falls into two directions: single-agent and multi-agent approaches. The single-agent approach utilizes a single LLM to handle specific legal tasks^[5-7]; its performance is often constrained when faced with problems requiring diverse expertise or integrated reasoning. To overcome these limitations, researchers have turned to multi-agent systems, which simulate complex legal decision-making environments through the collaboration of multiple specialized agents^[8,9]. Nonetheless, this architectural shift also introduces new challenges, including significant communication overhead, difficulties in maintaining output consistency, and a substantial increase in system complexity. To better leverage these agents for tackling legal tasks, it is necessary to conduct a comprehensive review of current legal agent research, systematically analyze the challenges and methodologies of single- and multi-agent systems in legal tasks, and outline future research directions.

While numerous surveys have explored the application of LLMs in the legal domain, they still exhibit certain limitations. For instance, Feng *et al.* conducted a survey exclusively focused on legal judgment prediction tasks^[10], Yang *et al.* reviewed only legal question answering tasks^[11]; however, neither of them provided a systematic summary of tasks across the broader legal domain. Similarly, Siino *et al.* emphasized the potential of LLMs but considered only single-agent scenarios, overlooking critical aspects such as multi-agent collaboration and cross-task generalization^[12]. Furthermore, existing research on legal multi-agent systems primarily concentrates on specific task scenarios, lacking a systematic analysis of the challenges and methodologies pertinent to multi-agent settings^[13,14]. Therefore, although current studies have laid a foundation for legal AI research, they are overall limited by insufficient task coverage, an overemphasis on single-agent approaches, and a lack of systematic synthesis regarding the challenges in multi-agent systems.

This paper systematically examines core tasks in the legal domain and the challenges they face, and analyzes how LLM agents have been applied to address these issues in recent years. A systematic outline of this survey is presented in [Figure 1](#). The main contributions of this paper are summarized as follows:

- (1) A comprehensive review of core tasks in the legal domain and their evaluation benchmarks in recent years (Section LEGAL TASKS AND BENCHMARKS).
- (2) An analysis of the key technical challenges and research methodologies for single-agent systems in legal scenarios (Section CHALLENGES AND METHODS FOR SINGLE AGENTS).
- (3) A systematic summary of the challenges and approaches related to multi-agent systems in the legal domain (Section CHALLENGES AND METHODS FOR MULTI-AGENTS).
- (4) An outlook on future research directions, providing guidance for the development of legal intelligent agents (Section FUTURE PROSPECT).

LEGAL TASKS AND BENCHMARKS

This section systematically reviews the primary tasks of legal agents, provides a comparative analysis of the strengths and weaknesses of single-agent vs. multi-agent approaches in handling legal tasks, and examines the mainstream benchmarks from the past three years.

Legal tasks

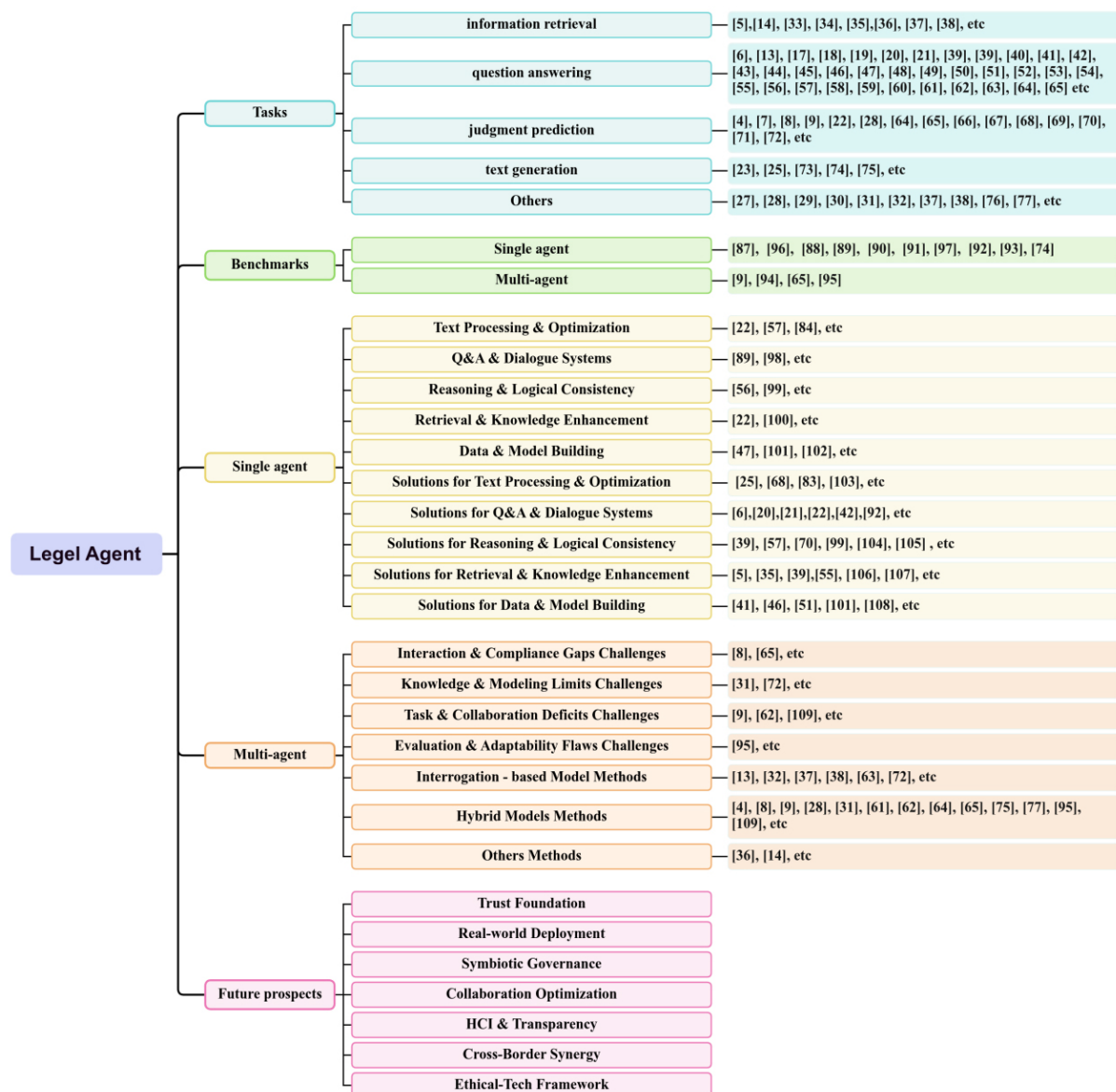


Figure 1. A systematic outline of this survey.

With the advancement of LLMs and intelligent agent technologies, legal agents have been widely applied to various legal tasks to enhance the efficiency of legal work. Core tasks in the legal domain include legal information retrieval, legal question answering, legal judgment prediction, and legal text generation. Additionally, tasks such as compliance analysis, due diligence, and legal ontology integration are also included. These tasks cover the major scenarios within the legal field and constitute the core task framework for legal agent applications. Next, we provide a detailed introduction to these primary legal tasks.

Legal information retrieval is a foundational task that supports nearly all upper-layer legal applications. Its goal is to accurately and efficiently identify the most relevant information from a large volume of legal documents in response to the specific needs of legal professionals. Depending on the retrieval objective, it can be primarily categorized into: (1) case law retrieval^[15,16], which involves retrieving historically decided cases with reference value that are similar to the current case based on its factual circumstances; (2) statutory provision retrieval^[17], which aims to rapidly locate relevant legal statutes and regulations based on

a legal question or case facts. An efficient legal information retrieval system can significantly enhance the productivity of legal research and practice, and serves as a core component in many legal intelligent agent systems.

Legal question answering primarily involves retrieving and generating accurate, interpretable legal answers based on natural language questions posed by users, using legal knowledge bases or case databases. Unlike general question answering, legal question answering requires not only understanding complex legal language but also strictly adhering to statutory provisions and precedent logic. Currently, this task has been widely applied in intelligent legal consultation systems^[6], public legal education platforms^[18], and online legal service platforms^[19-21].

Legal judgment prediction aims to forecast potential court rulings based on case facts, legal provisions, and relevant precedents, including the determination of charges, sentencing recommendations, and compensation amounts^[10]. Its core objective is to generate judgment predictions that conform to judicial reasoning by taking as input case information such as case summaries, relevant legal articles, and parties' actions, thereby assisting lawyers, judges, and litigants in better understanding the likely trajectory of a case. Unlike traditional prediction tasks, legal judgment prediction must strictly adhere to legal logic and judicial interpretations^[22], posing higher demands on current technologies.

Legal text generation aims to produce professional texts that conform to legal linguistic norms and logical structures using natural language generation techniques, while avoiding factual errors and hallucinated information^[23]. Depending on the type of content generated, the main categories include: (1) legal summarization^[24], which involves summarizing lengthy legal documents by extracting key points such as core facts, issues in dispute, and conclusions; (2) judgment rationale generation^[25], which generates logically rigorous explanations for judicial decisions based on given case facts and applicable legal provisions; and (3) contract drafting^[26], which involves generating or revising contractual texts according to clause templates and mutual agreements.

In addition to the aforementioned core tasks, legal agents are also being progressively applied to more complex legal practice scenarios. Examples include compliance analysis^[27-29] and due diligence^[30]. The task of compliance analysis aims to help enterprises automatically review whether their business operations comply with relevant regulations, thereby mitigating potential legal risks. Due diligence is commonly conducted in mergers and acquisitions, as well as in investment and financing processes, with the goal of automating the review of contracts, financial documents, and potential legal risks. Furthermore, multi-agent systems encompass the specialized task of legal ontology integration^[31,32], which involves constructing structured legal knowledge graphs across domains and jurisdictions to achieve unified representation and semantic interconnection of fragmented legal knowledge. This task provides a foundational knowledge base for coordinated decision-making in multi-agent systems.

Table 1 summarizes recent legal tasks and their associated literature from the past three years. **Figure 2** illustrates the distribution of various legal tasks in the field of legal intelligence. As shown in **Table 1** and **Figure 2**, we can see that: (1) The number of research literature under the single-agent paradigm is significantly more than that of multi-agent, which reflects the advantages of the former in technology maturity; (2) Research in both paradigms is heavily concentrated on Legal Questions and Answers, indicating that this task is a fundamental and pivotal challenge in the legal domain; (3) The task of "legal ontology integration" only appears in multi-agent research, highlighting their distinct capability in handling complex problems that necessitate deep collaboration and knowledge integration.

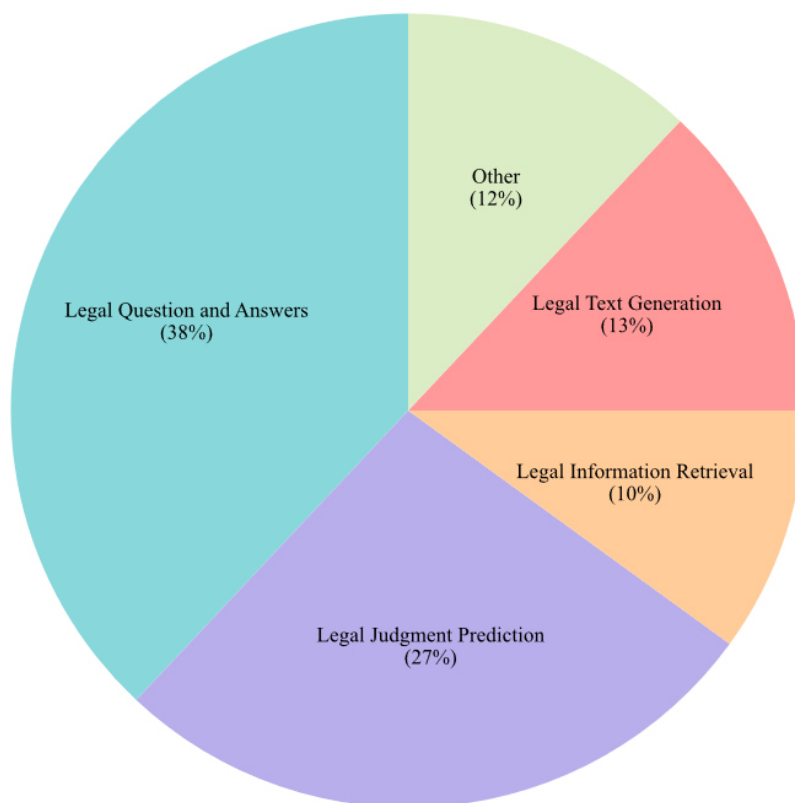


Figure 2. Distribution of legal tasks.

Table 1. Legal tasks and related literature

	Legal tasks	Works (Single-agent)	Works (Multi-agent)
Main tasks	Legal Information Retrieval	[5,33-35]	[14,36-38]
	Legal Question and Answers	[6,17-21,39-60]	[13,61-65]
	Legal Judgment Prediction	[7,22,66-71]	[4,8,9,28,64,65,72]
	Legal Text Generation	[23,25,73,74]	[75]
Other tasks	Legal Compliance analysis	[27-29,76]	[77]
	Legal Due diligence	[30]	[37,38]
	Legal Ontology Integration	-	[31,32]

Figure 3 displays the publication volume statistics for the Legal Agent domain during the period 2022-2025. From Figure 3, we can observe several important trends:

- (1) Overall Growth Trend: Both single-agent and multi-agent publications show significant growth from 2022 to 2025, indicating that the Legal Agent field is in a rapid development phase.
- (2) Technology Maturity Gap: Single-agent publications consistently outnumber multi-agent ones, reflecting the more mature application of single-agent technology in the legal domain. The gap is particularly evident in 2022, with 11 single-agent publications vs. 0 multi-agent publications, demonstrating different stages of technological development.
- (3) Growth Rate Analysis: Although starting from a lower base, multi-agent research shows rapid growth beginning in 2023, indicating rapidly increasing research interest in this direction.

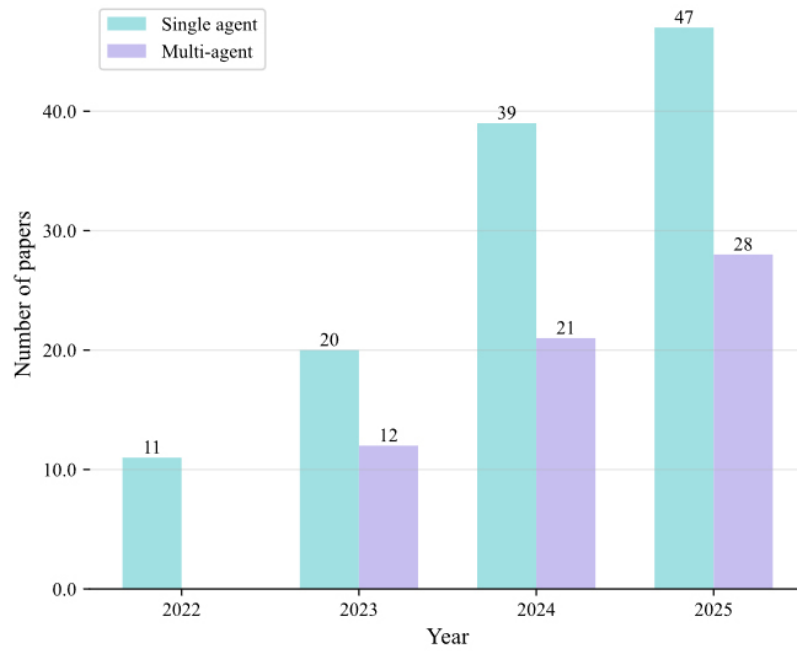


Figure 3. Legal agent publications (2022-2025).

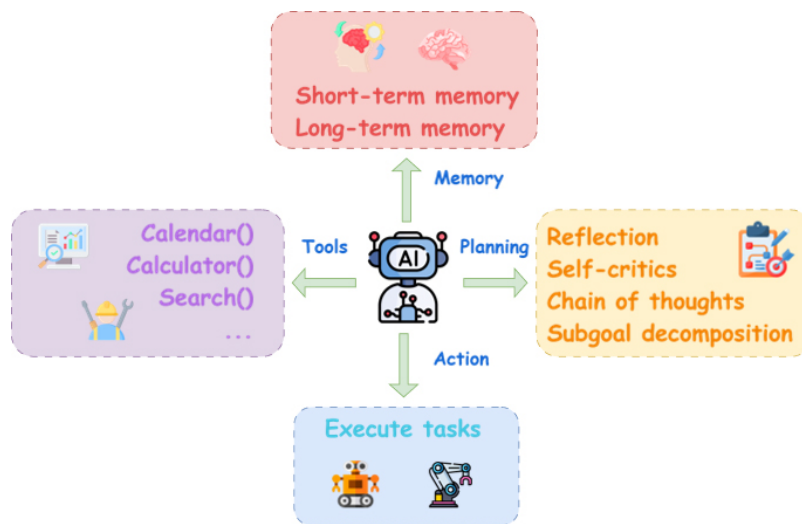


Figure 4. Agent architecture.

(4) Research Focus Evolution: The steady growth in single-agent research confirms its current dominance, while the rapid expansion of multi-agent research suggests it may become an important future trend.

Single-agent vs. multi-agent systems

A typical intelligent agent architecture comprises four core components: planning, memory, tools, and action, as shown in Figure 4.

The planning module acts as the agent's "brain", responsible for high-level reasoning and decision-making. It guides the agent toward its ultimate goal by decomposing complex tasks, formulating step-by-step strategies, and engaging in continuous reflection and adjustment.

The memory module provides the agent with persistent information storage and retrieval, enabling it to learn from past interactions and maintain state across dialogues and tasks.

The tools module significantly expands the agent's capabilities by accessing external resources to acquire real-time information or perform computations beyond the scope of its internal model.

The action module translates internal decisions and plans into concrete, executable commands, outputting them to the environment and thereby closing the critical loop from "thinking" to "acting".

This integrated architecture allows the agent to handle complex real-world tasks in a dynamic, self-adaptive, and incrementally learning manner. Following this overview, we will discuss the advantages and disadvantages of single-agent and multi-agent systems.

Single-agent systems

(1) Strengths.

The single agent features a simple structure and straightforward training and deployment processes, reducing system complexity and development costs^[18,22,78,79]. Only one model needs to be trained and maintained, resulting in low resource consumption, which makes the single agent suitable for resource-limited environments^[18,79,80]. This efficiency enables strong performance in legal tasks with high real-time demands, such as legal information retrieval, question answering, judgment prediction, and text generation^[25,81].

Additionally, single agents can quickly handle standardized legal tasks. For example, in legal compliance analysis, they can efficiently identify compliance issues in legal texts and generate reports^[81]. In legal due diligence, single agents can process tasks within a single legal domain; for instance, Law Large Language Model (LawLLM) focuses on the U.S. legal system and can efficiently complete due diligence tasks related to U.S. law^[82]. Similarly, in legal ontology integration, single agents can rapidly integrate ontological knowledge from a single legal field, providing efficient solutions^[81].

(2) Weaknesses.

Single-agent systems rely on the knowledge and capabilities of a single model, which can lead to knowledge limitations and insufficient generalization ability^[76,83]. For example, when dealing with complex legal issues or tasks involving intersections of multiple legal fields, they may fail to make full use of multi-source knowledge and information from different perspectives. This limitation restricts the comprehensiveness and accuracy of their results^[12,47].

Moreover, single-agent systems tend to lack flexibility when handling cross-domain or cross-border legal tasks. In compliance analyses involving multiple legal systems, they may exhibit knowledge blind spots^[81]. Similarly, in legal due diligence or ontology integration, they often face difficulties combining information or ontological knowledge from different legal systems, leading to incomplete or suboptimal solutions.

Multi-agent systems

(1) Strengths.

Through collaboration among multiple agents, multi-agent systems can integrate diverse knowledge and skills, improving the accuracy and comprehensiveness of task execution^[48,70,73,84]. For example, by combining LLMs with knowledge sources such as knowledge graphs, multi-agent systems can better handle complex legal issues and provide more comprehensive solutions^[48,70]. Interactions and learning between agents also

Table 2. Strengths and weaknesses of single-agent vs. multi-agent systems in the legal domain

	Single agent	Multi-agent
Strengths	Simple structure, Low resource consumption, Fast response	Knowledge integration, High accuracy, Strong generalization ability
Weaknesses	Knowledge limitations, Poor generalization ability	High resource requirements, High complexity, Difficult deployment, High maintenance costs

enhance their generalization ability, enabling adaptation to a wide range of legal tasks, including legal information retrieval, question answering, judgment prediction, and text generation^[12,73,75,85].

By leveraging multiple agents, these systems can analyze legal issues from different perspectives, yielding more thorough solutions. In legal compliance analysis, multi-agent systems can integrate knowledge from different legal fields to deliver more accurate assessments. In legal due diligence, they can combine legal knowledge from multiple countries, allowing them to manage complex cross-border tasks. In legal ontology integration, multi-agent systems can merge ontological knowledge from various legal systems, achieving cross-system integration and more comprehensive solutions^[73].

(2) Weaknesses.

Multi-agent systems require the design and implementation of interaction mechanisms among multiple agents, which increases system complexity as well as development and maintenance costs, and raises the risk of errors^[48,69]. They also demand greater computing and storage resources to support agent operation and collaboration, making deployment in resource-limited environments challenging^[75,86].

Multi-agent systems face high coordination costs and increased system complexity. In tasks such as legal compliance analysis, legal due diligence, and legal ontology integration, synchronizing information and coordinating decisions among multiple agents requires additional mechanisms to ensure efficiency and consistency, which may slow system operation^[73].

A more concise comparison is shown in [Table 2](#).

Benchmarks

To objectively evaluate mainstream LLMs on the legal tasks defined in Section **Legal tasks**, researchers have established specialized benchmarks. The evolution of these benchmarks reveals a clear trajectory: a shift from single-agent static tasks to multi-agent dynamic interactions, an expansion from English to multilingual coverage, and a progression from basic cognitive skills to sophisticated practical applications.

Regarding single-agent benchmarks, early examples such as LegalBench^[80] comprehensively assessed six categories of cognitive skills, including legal issue spotting, rule recall, and application, through 162 tasks. Subsequently, language-specific benchmarks emerged to address linguistic gaps, exemplified by ArabLegalEval^[87] for Arabic and the Korean Legal Benchmark^[88] for Korean, reflecting a push for greater diversity. Concurrently, Chinese single-agent benchmarks developed rapidly. LawBench^[89] pioneered this effort by systematically constructing 20 tasks focused on memory, understanding, and application within China's mainland legal system. LexEval^[90] expanded this foundation with 23 tasks, becoming the most comprehensive Chinese benchmark at the time and emphasizing logical reasoning and ethical judgment. By 2025, the focus shifted towards practical utility with benchmarks such as LAiW (a Chinese legal AI

benchmark for practical applications, comprising 14 tasks across 3 domains)^[91], which emphasized case and document processing, and UCL-Bench (a Chinese User-Centric Legal Benchmark, comprising 22 tasks across 5 distinct legal scenarios)^[92], which adopted a user-centric design to mirror real-world legal services. Furthermore, the emergence of specialized benchmarks, such as JuDGE (Judgment Document Generation for Chinese Legal System)^[74] for Chinese judgment document generation, indicates a maturation in single-agent evaluation from broad coverage to deeper, more specialized assessments.

The emergence of multi-agent legal benchmarks signifies a paradigm shift, moving from evaluating static, individual capabilities to assessing dynamic, collaborative interactions. SimuCourt (a judicial benchmark that encompasses 420 Chinese judgment documents)^[9] established a foundational simulated judicial environment for three major case types and two trial levels. Subsequent benchmarks then specialized further: LegalAgentBench (a comprehensive benchmark specifically designed to evaluate LLM Agents in the Chinese legal domain)^[93] introduced complex tasks such as multi-hop reasoning to address profound legal complexities; Multi-stage Interactive Legal Evaluation (MILE)^[65] focused on evaluating intensive dynamic interactions beyond static limitations; and J1-Eval (a fine-grained evaluation framework for assessing task performance and procedural compliance in dynamic legal environments)^[94] constructed a procedurally constrained, multi-role environment for fine-grained, full-process assessment from consultation to court.

This evolution, as systematically presented in [Tables 3 and 4](#), illustrates a clear pathway from the broad coverage of single-agent evaluation toward the dynamic interaction and full-process simulation of multi-agent systems, marking the strategic advancement of legal AI from foundational cognition to collaborative intelligence in complex practice.

CHALLENGES AND METHODS FOR SINGLE AGENTS

Single Agent refers to an intelligent system mode in which only one independent intelligent entity, equipped with autonomous perception, decision-making, and execution capabilities, independently completes environmental interaction and task goal achievement in a specific task scenario. In this chapter, the challenges and research methods faced in the development process of legal single agents will be described.

Challenges

For the legal tasks introduced in Section **Legal task**, LLMs face numerous challenges that significantly affect their reliability, practicality, and adoption in real-world legal contexts. These challenges can be categorized into five main areas: legal text processing and optimization, legal question and answering systems, legal reasoning and logical consistency, legal retrieval and knowledge enhancement, and legal data and model building.

Legal text processing and optimization

Legal texts are typically lengthy, complex, and filled with dense legal terminology, making it difficult for LLMs to efficiently process and extract relevant information. In terms of long text processing, legal documents such as court judgments and contracts often exceed the context window limitations of most LLMs, resulting in incomplete information extraction^[57]. Regarding unstructured documents, legal files usually lack a standardized structure, making the extraction of key information challenging^[83]. Additionally, the complexity and length of legal texts pose challenges to traditional retrieval methods, leading to inaccurate retrieval results^[22].

Legal question and answering systems

Table 3. Benchmarking mainstream single agent (2023–2025)

Benchmark	Year	Language	Task Num	Taxonomy	Data Size
LegalBench ^[80]	2023	English	162	Issue-spotting, Rule-recall, Rule-application, Rule-conclusion, Interpretation, Rhetorical-understanding	91,290
LegalEval-Q ^[96]	2025	English	1	Legal Text Quality Evaluation	10,000
ArabLegalEval ^[88]	2024	Arabic	3	Multiple Choice Questions, Question & Answer pairs, Arabic translations of tasks from LegalBench	37,583
Korean Legal Benchmark ^[89]	2024	Korean	15	Legal knowledge, legal reasoning, The Korean bar exam	3,308
LawBench ^[90]	2024	Chinese	20	Memorization, Understanding, Applying	-
LexEval ^[91]	2024	Chinese	23	Memotization, Understanding, Logic Inference, Discrimination, Generation, Ethic	14,150
CogniBench ^[97]	2025	Chinese	1	Assess Cognitive Faithfulness of LLM	24,084
LAiW ^[92]	2025	Chinese	14	Fundamental Information Retrieval, Legal Principles Inference, Advanced Legal Application	11,605
UCL-Bench ^[93]	2025	Chinese	22	Memory, Understand, Apply, Analyze, Evaluate, Create	1,100
JuDGE ^[74]	2025	Chinese	1	Judgment Document Generation	2,505

Benchmark: The name of the evaluated legal benchmark; Year: The publication year of the benchmark; Language: The primary language of the datasets used in the benchmark; Task Num: The number of distinct legal tasks defined within the benchmark; Taxonomy: The classification scheme for legal tasks as proposed by the original benchmark authors; Data Size: The total volume of the benchmark's dataset, reported as the number of instances. LegalBench: a collaboratively constructed large-scale English legal reasoning benchmark; LegalEval-Q: an English benchmark specifically designed to evaluate the quality of legal texts; ArabLegalEval: Arabic Legal Reasoning and Question Answering Benchmark; Korean Legal Benchmark: Korean Legal Benchmark Dataset, covering legal knowledge, reasoning, and tasks related to the Korean bar exam; LawBench: Chinese Legal Language Understanding and Application Evaluation Benchmark, which includes memory, understanding, and application tasks; LexEval: Chinese Legal Reasoning Comprehensive Evaluation Benchmark, which assesses memory, comprehension, logical reasoning, ethical judgment, etc.; CogniBench: a Chinese benchmark for evaluating the cognitive faithfulness of large language models; LAiW: a Chinese legal AI benchmark for practical applications, comprising 14 tasks across 3 domains UCL-Bench: a Chinese User-Centric Legal Benchmark, comprising 22 tasks across 5 distinct legal scenarios; LLM: large language model; JuDGE: Judgment Document Generation for Chinese Legal System.

Legal question and answering (Q&A) systems need to balance the accuracy, interpretability, and smooth interaction of answers. However, LLMs often fail to provide clear reasoning chains when processing complex legal language, resulting in answers that lack interpretability^[97]. Moreover, user-friendly interaction is an essential component of legal question and answering systems, but LLMs perform poorly when handling incomplete or ambiguous user queries, affecting the interaction experience^[88].

Legal reasoning and logical consistency

Legal reasoning requires strict adherence to legal logic and judicial interpretation, but LLMs frequently generate answers with logical inconsistencies or hallucinations. This not only affects the accuracy of reasoning but also leads to a lack of credibility in the generated legal opinions^[97]. Furthermore, LLMs often struggle to provide clear and logical reasoning chains when dealing with complex legal cases, further limiting their application in the legal field^[56].

Table 4. Benchmarking mainstream multi-agent (2023-2025)

Benchmark	Year	Language	Task num	Taxonomy	Data size
SimuCourt ^[9]	2024	Chinese	1	Judicial Decision-Making	420
LegalAgentBench ^[93]	2025	Chinese	6	1-hop to 5-hop, Writing	160k
MILE ^[65]	2025	Chinese	1	Dynamic Legal Interaction	693
J1-Eval ^[94]	2025	Chinese	6	Knowledge Questioning, Legal Consultation, Complaint Drafting, Defence Drafting, Civil Court, Criminal Court	508

SimuCourt: A judicial benchmark that encompasses 420 Chinese judgment documents; LegalAgentBench: a comprehensive benchmark specifically designed to evaluate LLM Agents in the Chinese legal domain; MILE: Multi-stage Interactive Legal Evaluation; J1-Eval: a fine-grained evaluation framework for assessing task performance and procedural compliance in dynamic legal environments.

Legal retrieval and knowledge enhancement

Legal retrieval involves accurately identifying relevant legal information from a large volume of documents, but the complexity and ambiguity of legal terminology pose challenges to retrieval. Traditional retrieval methods perform poorly when dealing with long texts and complex legal terms, leading to inaccurate retrieval results^[22]. Additionally, LLMs face difficulties in integrating legal knowledge, as they struggle to effectively combine legal statutes, case law, and professional knowledge, affecting the accuracy and comprehensiveness of retrieval^[98].

Legal data and model building

Developing effective legal LLMs requires high-quality data and robust model training, but the scarcity of data in the legal field is particularly prominent. High-quality labeled data is often scarce in the legal domain, especially in resource-poor regions, limiting model training and optimization^[99]. Moreover, general-purpose LLMs typically lack in-depth knowledge of specific legal systems, resulting in poor performance in specific legal tasks^[47]. Finally, the complexity and diversity of the legal field also place higher demands on the generalization ability of models, increasing the difficulty of model building^[100].

Methods

In response to the challenges in performance, text processing, and evaluation, researchers have proposed systematic solutions from multiple technical perspectives. This section reviews current technical approaches and innovations aimed at addressing the multifaceted challenges of deploying LLMs in the legal domain. It encompasses five areas: legal text processing and optimization, legal Q&A and dialogue systems, legal reasoning and logical consistency, legal retrieval and knowledge enhancement, and data and model building. These areas provide theoretical and practical support for the development of legal AI systems.

Solutions for legal text processing and optimization

The primary challenge in legal text processing involves handling extensive, intricate, and term-heavy documents. Researchers have proposed various innovative solutions to address these challenges. For long document splitting, Shen *et al.* proposed a lightweight solution using improved rotary position encoding (NTK-RoPE), which extends the context length beyond 8K tokens via Low-Rank Adaptation (LoRA) fine-tuning, thereby significantly enhancing processing capabilities^[83]. For addressing special needs in text classification, Johnson *et al.* innovatively reframed multi-label classification as an instruction generation task, solving label imbalance^[101]. Prasad *et al.* proposed the Multi-stage Encoder-based Supervised with-clustering (MESc) hierarchical framework for unstructured long documents using sharding, clustering, and

aggregation^[68]. For document generation, Lin *et al.* used a “legal constituent elements decomposition” method to ensure generated text strictly follows legal element order^[25]. Overall, a complete technical chain covering encoding extension, structure modeling and lightweight fine-tuning has been established, effectively addressing the unique challenges posed by long legal texts.

Solutions for legal Q&A and dialogue systems

Legal Q&A systems need to balance answer accuracy, explainability, and smooth interaction. Researchers have proposed multi-level technical solutions. To enhance reasoning capabilities, Jiang and Yang embedded legal syllogism into prompt templates^[22]. Dai *et al.* built the LAiW benchmark to test model logic consistency^[91]. To improve answer reliability, Louis *et al.* proposed a “retrieve-then-read” framework to provide traceable legal bases for answers^[42]. For better interaction, Yao *et al.* designed an interactive clarification system to address missing information in user queries^[20]. Considering user emotion, Mishra *et al.* included emotional factors in a reinforcement learning (RL) framework^[6]. Moreover, John *et al.* explored deep learning methods, verifying that deep generative models are feasible for legal dialogue^[21]. These studies form a progressive technical chain, including syllogistic prompting, retrieval augmentation, interactive clarification, and RL fine-tuning, laying a solid foundation for legal AI services.

Solutions for legal reasoning and logical consistency

The core challenge lies in enabling LLMs to reason logically akin to judges while mitigating hallucinations. Researchers developed various innovative methods. For building clear logic frameworks, Deng *et al.* embedded judicial syllogism into prompt engineering^[97]. Kaczmarczyk *et al.* achieved strict logical consistency^[57]. For discriminative reasoning needs, Deng *et al.* proposed a three-stage Ask-Discriminate-Predict (ADAPT) process^[97]. Wang *et al.* further formalized it into differentiable rules^[70]. To improve causal reasoning, Chen *et al.* developed the causality-aware self-attention mechanism (CASAM) to help models focus on key judgment words^[102]. To enhance system robustness, Hu *et al.* used two-stage training to reduce hallucinated law citations^[103]. Nguyen *et al.* added aspect constraints to ensure output consistency^[40]. Current research forms a progressive path, encompassing explicit structure, discriminative matching, causal intervention, and adversarial robustness, progressively addressing issues of logical consistency and hallucinations.

Solutions for legal retrieval and knowledge enhancement

Legal retrieval confronts challenges stemming from term ambiguity and long text noise. To enhance text representation, Zhou *et al.* proposed a “compress then retrieve” method^[5]. Zeng *et al.* introduced Step-Back Prompting^[104]. Akbar *et al.* developed a “State-Wise Index” strategy^[55]. To strengthen knowledge fusion, Li *et al.* proposed the Unified Legal Retriever (UniLR) framework, unifying key element supervision and graph knowledge enhancement^[35]. Wiratunga *et al.* embedded case-based reasoning (CBR) cycles into the retrieval augmented generation (RAG) process^[39]. To optimize retrieval strategy, Akbar *et al.* proposed a strategy adaptive to query complexity^[55]. Tang *et al.* built a global case graph and used degree regularization loss^[105]. These studies evolved along three lines, including fine-grained representation, structured fusion, and adaptive strategy, significantly improving retrieval accuracy and system robustness, thereby providing substantial support for high-trust legal AI applications.

Solutions for legal data and model building

The development of legal LLMs confronts challenges such as scarce data and inadequate model adaptability. To address data scarcity, Li *et al.* proposed Continuous Semantic Augmentation Fine-Tuning (CSAFT) to generate diverse training samples through semantic expansion^[41]. Sheik *et al.* used pre-trained models to generate pseudo-labeled data^[99]. Zhou *et al.* developed a knowledge-guided generation

framework (KGDG) to ensure data quality^[106]. For model optimization, Fei *et al.* used a two-stage fine-tuning strategy to improve model generalization and specialization^[51]. Al-Qaesm *et al.* used quantization and synthetic data for efficient resource use^[46]. Research has formed a technical system consisting of data enhancement, model fine-tuning, and knowledge fusion, which effectively supports the practical application and deployment of legal AI systems.

Overall, the evolution of legal LLMs is progressing along three major trends: knowledge enhancement, structured reasoning, and humanized service. A comprehensive technical chain is gradually taking shape, consisting of data enhancement, knowledge fusion, structured reasoning, and interaction optimization. These methods work together to improve the reliability, usability, and accessibility of legal LLMs, thereby laying a solid foundation for their eventual mature and reliable integration into judicial practice.

CHALLENGES AND METHODS FOR MULTI-AGENTS

Multi-agent is an intelligent system model composed of multiple intelligent entities with autonomous decision-making capabilities. These entities work together to accomplish complex task objectives that are difficult for a single agent to achieve, through information exchange, collaborative division of labor, or competition mechanisms. In this chapter, the challenges and research methods faced in the development of legal multi-agent systems will be described.

Challenges

In the field of legal multi-agent systems, in addition to the challenges that single-agent systems may face, there remain multiple challenges when simulating real judicial scenarios. These challenges primarily manifest in four dimensions: dynamic interaction capabilities, depth of knowledge integration, task coverage completeness, and evaluation system adaptability. This section will discuss these dimensions in detail.

Insufficient dynamic interaction and procedural compliance

Existing legal systems based on LLMs face difficulties when simulating the dynamic processes of real courts. Static legal models have significant flaws in multi-round debates: on the one hand, the models cannot flexibly apply legal provisions for in-depth adversarial debates and are prone to overfitting in standardized tasks^[8]; on the other hand, real legal scenarios require high levels of professionalism and strong interactivity, which often leads to issues such as misunderstanding of demands or omission of facts in LLMs^[65]. The scarcity and privacy of dynamic interaction data further complicate the model's adaptability in complex scenarios.

Limitations in legal knowledge integration and cognitive modeling

There are structural barriers to integrating multi-agent systems with legal ontologies. Existing multi-agent system platforms cannot effectively interface with structured models that store legal knowledge, making it difficult for agents to access and apply legal knowledge in real time^[31]. A deeper challenge lies in the inadequacy of cognitive modeling: although traditional dynamic epistemic logic (DEL) can formalize changes in knowledge states, it cannot address the critical issue of "attention limitations" in legal scenarios, the degree to which agents focus on specific information. This makes predicting defendants' behaviors and accurately modeling their subjective intentions very challenging^[72].

Lack of complex task handling and collaboration mechanisms

Current legal AI technologies mostly focus on isolated judicial stages, such as legal Q&A or case retrieval, making it difficult to cover the entire process from courtroom debate to judgment^[9]. In terms of

collaboration, single-agent systems cannot simulate the collaborative workflow of professionals in a real law firm, leading to overly generalized answers in scenarios such as Chinese legal consultations, lacking personalization^[62]. Although LLMs show potential in judicial decision-making, traditional legal judgment prediction methods still overly rely on historical case datasets, failing to fully leverage the collaborative potential of multi-agent systems and the dynamic reasoning abilities of LLMs^[107].

Inadequate evaluation systems and adaptability to dynamic environments

Current legal intelligence benchmarks largely adopt static, non-interactive paradigms, such as multiple-choice questions and single tasks, which do not reflect the dynamic characteristics of real legal practice, including multi-role interactions, procedural compliance requirements, and incremental updates of information. This limitation in evaluation methods hinders the comprehensive assessment of agents' performance in complex legal environments, thereby obstructing the iterative optimization of systems^[94].

Methods

We classify the system into debate-based, interrogation-based, mediation-based, and other models based on the interaction methods among agents in a multi-agent system. This section will discuss the corresponding legal agent systems from different dimensions based on their interaction modes.

Interrogation-based model

In the research of legal multi-agent systems, the interrogation-based model is typically used to simulate the interrogation process in legal scenarios, where interactions between agents primarily occur through questions and answers, and the exchange of information, rather than intense adversarial debates. This subsection will focus on discussing the core approach to problem-solving in this model, which mainly includes three aspects: structured questioning and multi-round interaction guidance, decomposition of legal issues into sub-problems, and the use of external tools.

The model guides agents to reason step by step through structured questioning and multi-round interactions to solve complex legal tasks. For example, in the RAG combined system developed by Mamalis *et al.*, agents search the General Data Protection Regulation (GDPR) legal database and generate targeted queries to guide LLMs in extracting key legal clauses^[13]. In the PAKTON (a fully open-source, end-to-end, multi-agent framework with plug-and-play capabilities) framework, the “questioner” agent iterates questioning the “researcher” agent to refine the legal report, integrating the ReAct (an approach to integrating reasoning and acting in LLMs) paradigm to combine reasoning and actions, filling knowledge gaps, and generating structured legal conclusions^[38]. Similarly, in the DEL model, the “questioner” agent performs multiple rounds of questioning to update the cognitive states of agents, verifying the logical consistency between legal clauses^[72]. Such models emphasize multi-round interactions and evidence verification, greatly enhancing the transparency and interpretability of legal analysis.

Meanwhile, researchers have invented a core mechanism that uses a method of breaking down legal issues, dividing them into logically related sub-problems, and then guiding the model through iterative questioning for deep reasoning. For instance, LegalGPT (Legal Chain of Thought for the Legal Large Language Model Multi-agent Framework) developed by Shi *et al.* further designed three types of thinking chains (COT) for legal exams, consultations, and judgment prediction, where the consultation COT requires users to structure their questions into “problem summary - scenario description - specific question - expected solution - special constraints”^[63]. Based on this structure, agents generate step-by-step reasoning paths, combined with external knowledge bases to correct hallucination issues.

Moreover, to further enhance the system's efficiency and flexibility, external tools and ontology designs can help agents more accurately retrieve legal provisions and contract contents, thereby reducing costs and improving analytical capabilities. For instance, the LAW (Legal Agentic Workflows for Custody and Fund Services Contracts) framework introduced by Watson *et al.* employs code-generation agents that dynamically call dedicated tools through a “question-tool invocation-result validation” chain process to handle complex tasks such as contract lifecycle calculations^[37]. Ontology-based legal systems simulate the interrogation process through middleware, with agent behavior validated through ontology rules, forming a “behavior-questioning-compliance feedback” loop to ensure decisions comply with legal norms^[32].

Hybrid models

The interactions in multi-agent systems usually present diverse forms. Based on three basic interaction methods: debate, interrogation, and mediation, this paper categorizes the hybrid interaction models in existing literature into three types: debate and interrogation coexistence models, interrogation and mediation coexistence models, and models that combine interrogation, debate, and mediation, and conducts a systematic discussion.

Debate and Interrogation Coexistence. In this hybrid model, agent interactions combine both debate and interrogation mechanisms. On the one hand, the parties involved (e.g., plaintiff and defendant) engage in confrontation through evidence presentation and cross-examination, simulating an adversarial litigation process, while the judge remains neutral and plays the role of a referee. On the other hand, the judge may intervene proactively (e.g., when the authenticity of evidence is in doubt) to exercise the right of examination, such as directly questioning witnesses or retrieving evidence.

The AgentsCourt framework, proposed by He *et al.*, combines interrogation and debate models by simulating courtroom procedures^[9]. The judge guides fact-finding and legal basis retrieval through structured questioning, while in the free debate phase, the plaintiff and defendant agents engage adversarially, ultimately completing the entire process from case analysis to judgment generation. The Debate-Feedback mechanism further optimizes this coexistence model by dynamically merging the interrogation and debate components through smooth operations, enhancing the accuracy of the judgment^[107]. Zhang and Ashley's reflective multi-agent framework combines interrogation tasks (the fact analyst rigorously examines the basis of arguments) with debate tasks (plaintiff and defendant agents construct argument structures) through iterative reflection cycles, ensuring that the questioner rigorously checks the factual foundation of the debate while debaters construct a more rigorous three-tier argument structure^[64]. This approach helps reduce erroneous reasoning and enhances the persuasiveness of the argument. Furthermore, the AgentCourt framework presented by Chen *et al.* deepens the integration of the 2 models, replicating the binary structure of “judge-led interrogation - lawyer adversarial debate” in court, enabling knowledge co-evolution between 2 types of agents in civil case simulations through the Adversarial Evolution (AdvEvol) evolutionary mechanism^[8]. Overall, current legal multi-agent research combines reasoning mechanisms of interrogation and debate to develop a decision-making framework that is both rigorous in judicial processes and ethically adaptable.

Interrogation and Mediation Coexistence. In this mechanism, the interrogation agent (e.g., lawyer) is responsible for leading fact investigations and legal applications. If an agent's conclusion is contradictory or inconsistent, a mediation agent or some mediation mechanism needs to be initiated to intervene and ensure the consistency of reasoning and uniformity of conclusions.

In the AgentDevLaw middleware architecture proposed by Sperotto and Aguiar, the interrogation-based

model actively checks whether the fisherman's behavior complies with legal rules through government agents, while the mediation-based model involves system agents receiving violation signals from the legal ontology and dynamically generating sanction suggestions to resolve rule conflicts^[31]. In the recruitment audit framework reported by Fernández *et al.*, the interrogation model is led by expert agents to inspect the recruitment system (e.g., checking for age or racial discrimination), while the mediation model uses a rule engine to convert ethical and legal norms into executable conflict mediation strategies, ensuring the audit process is compliant and inclusive^[77]. In the MALR (Multi-Agent framework for improving complex Legal Reasoning capability) developed by Yuan *et al.*, the interrogation model strictly reviews legal elements through sub-task agents, while the mediation model corrects errors through an adaptive rule learning mechanism, optimizing the reasoning path to resolve confusion over charges^[4]. In the ChatLaw framework developed by Cui *et al.*, researchers perform structured questioning based on a legal knowledge base, while legal assistants guide users through conversational prompts to supplement vague information, ensuring the rigor and user-friendliness of consultations^[61]. The LawLuo framework further deepens this coexistence mechanism, where lawyer agents actively question case details in multi-round dialogues, while secretary agents organize the dialogue records to generate personalized reports. The Boss agent supervises the balance between the two modes, ensuring dynamic adaptation of legal agenda execution and user demand responses^[62]. The MASER (a Multi-agent Legal Simulation Driver) framework developed by Yue *et al.* strengthens this coexistence mode through a supervisory mechanism, with lawyer agents asking questions based on the preset legal agenda, while client agents simulate real user behavior with information gaps or emotional resistance, and supervisory agents correct conversation behaviors in real-time, seamlessly integrating the interrogation and mediation processes^[65]. This coexistence model significantly enhances both the professional rigor and interactive adaptability of legal services.

Interrogation, Debate, and Mediation Coexistence. This hybrid model combines all three interaction methods, forming a multi-stage dynamic collaborative process. Initially, the debate model dominates, with party agents presenting evidence for confrontation; then it shifts to an interrogation phase, where the judge agent leads the review of key evidence; finally, the mediation model intervenes, with the mediator agent promoting reconciliation and eventually forming a unified judgment.

This fusion paradigm simulates real legal scenarios and enhances decision robustness and fairness. For example, the J1-EVAL framework constructs a dynamic legal environment by simulating the prosecutor-defendant confrontation in criminal courts, lawyer debates in civil courts, and the court mediation stage, building a multi-role collaborative evaluation system^[94]. Courtroom-LLM further maps court procedures into multi-agent interactions, where prosecution and defense lawyers debate the case evidence, and the judge uses multiple rounds of questioning and reasoning to form a judgment while introducing mediation mechanisms to balance multiple positions and create a dynamic decision-making loop^[75]. AUTOLAW (a novel violation detection framework) innovatively adopts a jury system, where jurors evaluate based on legal knowledge through an interrogation-based process, and different legal role agents vote and negotiate based on case law and professional knowledge rankings, using a majority vote mechanism to reduce bias risks in single models^[28]. This coexistence model integrates legal logic rigor, critical adversarial debates, and inclusive mediation decision-making through multi-role division of labor and dynamic interaction.

Other

In addition to the above types of agents, multi-agent design can also be applied to the development of legal recommendation systems. The research on legal multi-agent recommendation systems focuses on the collaboration among agents and the integration of semantic knowledge, addressing the complexity of information in the legal field and the demand for personalized services. The early Infonorma system used a

layered architecture based on the Multi-Agent Application Engineering Methodology (MAAEM), utilizing the collaboration between user modeling agents and legal knowledge processing agents to solve the problem of unstructured legal information and achieve precise matching of user needs^[36]. Subsequent research further enhanced the role of ontology technology in legal knowledge representation. Drumond and Girardi, by constructing a semantic knowledge graph of legal ontologies, achieved automatic classification and similarity calculation of regulations and case laws, thus improving the adaptability and interpretability of recommendation results^[14]. These methods combine distributed decision-making by multi-agent systems and semantic Web technologies, offering efficient and accurate legal recommendation system solutions, making it an important direction in current legal AI research.

FUTURE PROSPECT

Over the next three years, legal agents will evolve in four linked steps: single-agent trust, multi-agent efficiency, cross-domain teamwork, and ethics governance. We will first build a safe base, then improve how agents cooperate, raise human-machine trust, and finally create a system that works across jurisdictions while staying both controllable and ethical.

Optimization of Cross Agent Collaboration Mechanism: Existing legal multi-agent architectures have achieved full process coverage from case analysis to judgment. However, task allocation and information redundancy among intelligent agents still affect the efficiency of the system. In the future, A fusion research method based on multi-agent RL and graph neural networks can be designed to create a more flexible task scheduling model, enhance collaboration between modules such as contract review, case retrieval, and evidence analysis, and avoid functional overlap.

Human computer interaction and increased transparency in decision-making: The trust of lawyers in AI largely depends on the interpretability and controllability of the system. For example, although the “dispute resolution” function possessed by legal agents has produced certain effects, current multi-agent systems still face the problem of “black box”. In the future, visualization tools can be developed to showcase the reasoning process of intelligent agents, and every data access, information transmission, and decision made by roles such as judges and lawyers will be recorded as immutable transactions, forming a clear chain of responsibility. Simultaneously, design a human-machine collaborative interface to enable lawyers to make real-time adjustments to the output of the intelligent agent.

Cross-jurisdictional system collaboration: Legal intelligent agent systems need to adapt to different legal modules and systems. In the future, it is hoped that a multilingual ontology legal knowledge graph can be constructed to unify the modeling of legal concepts under different legal systems. At the same time, by combining neural machine translation with specific word embedding techniques in the legal field, cross-linguistic alignment of legal concepts can be achieved.

Ethical and technological governance framework: With the increasing application of legal agents, ethical issues are becoming increasingly prominent. We suggest introducing a set of AI legal ethics quantitative evaluation indicators, including core measurement indicators such as source traceability, decision fairness, and program consistency, to measure the reliability and stability of intelligent agents' reasoning paths and conclusions when dealing with similar cases.

CONCLUSION

In this paper, we conducted a comprehensive survey of single-agent systems and multi-agent systems in the legal field. We outlined the core tasks and evaluation benchmarks in the legal field, examined LLM-based

agents in the legal domain, including their architecture, research methods, and challenges. Although these intelligent agent systems can improve efficiency in tasks such as legal retrieval, contract review, and judgment prediction, there are still many unresolved challenges, especially in terms of specialized governance, cross-jurisdictional system collaboration, and the collaboration of intelligent agents. Future work should focus more on the refinement and automation of specific laws, constructing systems that adapt to different legal systems, and developing robust communication protocols between intelligent agents to build a more comprehensive and resilient future legal technology system.

DECLARATIONS

Authors' contributions

Made substantial contributions to the conception and design of the survey, and led the writing of the original draft: Yang, S.

Performed literature investigation, data analysis, and contributed to the writing of the original draft: Yang, Z.

Participated in literature investigation, provided critical resources, and assisted in writing review and editing: Liu, Y.

Supervised the entire project, administered the process, reviewed and edited the manuscript: Wang, H.

Availability of data and materials

Not applicable.

Financial support and sponsorship

None.

Conflicts of interest

Hongtao Wang is an Editorial Board Member of the journal *AI Agent*. Hongtao Wang was not involved in any steps of the editorial process, including reviewers' selection, manuscript handling, and decision-making. The other authors declare that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2025.

REFERENCES

1. Zheng, Q.; Xia, X.; Zou, X.; et al. CodeGeeX: a pre-trained model for code generation with multilingual benchmarking on HumanEval-X. In *KDD '23: The 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Long Beach, USA, August 6-10, 2023; Association for Computing Machinery: New York, USA, 2023; pp 5673-84. [DOI](#)
2. Wang, W.; Ma, Z.; Wang, Z.; et al. A survey of LLM-based agents in medicine: how far are we from baymax? In *63rd Annual Meeting of the Association for Computational Linguistics, Findings of the Association for Computational Linguistics: ACL 2025*, Vienna, Austria, July 27-30, 2025; Association for Computational Linguistics: Stroudsburg, USA, 2025; pp 10345-59. [DOI](#)
3. Wan, H.; Feng, S.; Tan, Z.; Wang, H.; Tsvetkov, Y.; Luo, M. DELL: Generating reactions and explanations for LLM-based misinformation detection. In *62nd Annual Meeting of the Association for Computational Linguistics, Findings of the Association for Computational Linguistics: ACL 2024*, Bangkok, Thailand and virtual meeting, July 11-16, 2024; Association for Computational Linguistics: Stroudsburg, USA, 2024; pp 2637-67. [DOI](#) [PubMed](#)

4. Yuan, W.; Cao, J.; Jiang, Z.; et al. Can large language models grasp legal theories? Enhance legal reasoning with insights from multi-agent collaboration. In *The 2024 Conference on Empirical Methods in Natural Language Processing, Findings of the Association for Computational Linguistics: EMNLP 2024*, Miami, USA, November 11-16, 2024; Association for Computational Linguistics: Stroudsburg, USA, 2024; pp 7577-97. DOI
5. Zhou, Y.; Huang, H.; Wu, Z. Boosting legal case retrieval by query content selection with large language models. In *the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region, Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region*, Beijing, China, November 26-28, 2023; Association for Computing Machinery: New York, USA, 2023 pp 176-84. DOI
6. Mishra, K.; Priya, P.; Ekbal, A. Help me heal: a reinforced polite and empathetic mental health and legal counseling dialogue system for crime victims. *AAAI*. **2023**, 37, 14408-16. DOI
7. Zhang, Y.; Tian, Z.; Zhou, S.; et al. RLJP: legal judgment prediction via first-order logic rule-enhanced with large language models. *arXiv* 2025. arXiv:2505.21281. Available online: <https://doi.org/10.48550/arXiv.2505.21281> (accessed 11 November 2025).
8. Chen, G.; Fan, L.; Gong, Z.; et al. AgentCourt: simulating court with adversarial evolvable lawyer agents. In *63rd Annual Meeting of the Association for Computational Linguistics, Findings of the Association for Computational Linguistics: ACL 2025*, Vienna, Austria, July 27-30, 2025. Association for Computational Linguistics: Stroudsburg, USA, 2025; pp 5850-65. DOI
9. He, Z.; Cao, P.; Wang, C.; et al. AgentsCourt: building judicial decision-making agents with court debate simulation and legal knowledge augmentation. In *The 2024 Conference on Empirical Methods in Natural Language Processing, Findings of the Association for Computational Linguistics: EMNLP 2024*, Miami, USA, November 11-16 2024; Association for Computational Linguistics: Stroudsburg, USA, 2024; pp 9399-416. DOI
10. Feng, Y.; Li, C.; Ng, V. Legal judgment prediction: a survey of the state of the art. In *ThirtyFirst International Joint Conference on Artificial Intelligence (IJCAI-22), Proceedings of the Thirty-first International Joint Conference on Artificial Intelligence*, Vienna, Austria, July 23-29, 2022; International Joint Conferences on Artificial Intelligence Organization: Marina del Rey, USA, 2022; pp 5461-9. DOI
11. Yang, X.; Wang, Z.; Wang, Q.; Wei, K.; Zhang, K.; Shi, J. Large language models for automated Q&A involving legal documents: a survey on algorithms, frameworks and applications. *IJWIS*. **2024**, 20, 413-435. DOI
12. Siino, M.; Falco, M.; Croce, D.; Rosso, P. Exploring LLMs applications in law: a literature review on current legal NLP approaches. *IEEE Access*. **2025**, 13, 18253-76. DOI
13. Mamalis, M. E.; Kalampokis, E.; Fitsilis, F. A large language model agent based legal assistant for governance applications. In *23rd Annual International Conference on Digital Government Research, Proceedings of the 23rd Annual International Conference on Digital Government Research*, Seoul, Korea, June 15-17, 2022; Association for Computing Machinery: New York, USA, 2022; pp 286-301. DOI
14. Drumond, L.; Girardi, R. A multi-agent legal recommender system. *Artif. Intell. Law*. **2008**, 16, 175-207. DOI
15. Liu, B.; Hu, Y.; Wu, Y.; et al. Investigating conversational agent action in legal case retrieval. In *45th European Conference on Information Retrieval, Proceedings of the 45th European Conference on Information Retrieval, Part I*, Dublin, Ireland, April 2-6, 2023; Springer: Cham, Switzerland, 2023; pp 622-35. DOI
16. Liu, B.; Wu, Y.; Zhang, F.; et al. Query generation and buffer mechanism: towards a better conversational agent for legal case retrieval. *Inform. Process. Manag.* **2022**, 59, 103051. DOI
17. Xie, N.; Bai, Y.; Gao, H.; et al. DeliLaw: A Chinese legal counselling system based on a large language model. In *CIKM '24: The 33rd ACM International Conference on Information and Knowledge Management, Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, Boise, USA; October 21-25 2024; Association for Computing Machinery: New York, USA, 2024; pp 5299-303. DOI
18. Yuan, M.; Kao, B.; Wu, T. H.; et al. Bringing legal knowledge to the public by constructing a legal question bank using large-scale pre-trained language model. *Artif. Intell. Law*. **2023**, 32, 769-805. DOI
19. Amato, F.; Fonisto, M.; Giacalone, M.; Sansone, C. An intelligent conversational agent for the legal domain. *Information*. **2023**, 14, 307. DOI
20. Yao, R.; Wu, Y.; Zhang, T.; et al. Intelligent legal assistant: an interactive clarification system for legal question answering. In *WWW '25: The ACM Web Conference 2025, Companion Proceedings of the ACM on Web Conference 2025*, Sydney, Australia, April 28-May 2, 2025; Association for Computing Machinery: New York, USA, 2025; pp 2935-8. DOI
21. John, A. K.; Di Caro, L. Robaldo, L.; Boella, G. Legalbot: a deep learning-based conversational agent in the legal domain. In *22nd International Conference on Applications of Natural Language to Information Systems Lecture Notes in Computer Science (LNCS)*, Liège, Belgium, June 21-23, 2017; Springer: Cham, Switzerland, 2017; Vol. 10260, pp 171-5. DOI
22. Jiang, C.; Yang, X. Legal syllogism prompting: teaching large language models for legal judgment prediction. In *Nineteenth International Conference on Artificial Intelligence and Law (ICAIL 2023), Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law*, Braga, Portugal, June 19-23, 2023; Association for Computing Machinery: New York, USA, 2023; pp 417-21. DOI
23. Zhang, D.; Petrova, A.; Trautmann, D.; Schilder, F. Unleashing the power of large language models for legal applications. In *CIKM '23: The 32nd ACM International Conference on Information and Knowledge Management, Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, Birmingham, United Kingdom, October 21-25, 2023; Association for Computing Machinery: New York, USA, 2023; pp 5257-8. DOI
24. Shukla, A.; Bhattacharya, P.; Poddar, S.; et al. Legal case document summarization: extractive and abstractive methods and their evaluation. In *2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing, Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Online, November 20-23, 2022; Association for Computational Linguistics: Stroudsburg, USA, 2022; pp 1048-64. DOI
25. Lin, C. H.; Cheng, P. J. Assisting drafting of Chinese legal documents using fine-tuned pre-trained large language models. *Rev. Socionetwork. Strat.* **2025**, 19, 83-110. DOI
26. Aggarwal, V.; Garimella, A.; Srinivasan, B. V.; N, A.; Jain, R. ClauseRec: a clause recommendation framework for AI-aided contract authoring. In *2021 Conference on Empirical Methods in Natural Language Processing, Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Online and Punta Cana, Dominican Republic, November 7-11, 2021; Association for Computational Linguistics: Stroudsburg, USA, 2021; pp 8770-6. DOI

27. Hassani, S.; Sabetzadeh, M.; Amyot, D.; Liao, J. Rethinking legal compliance automation: opportunities with large language models. In *2024 IEEE 32nd International Requirements Engineering Conference (RE), Proceedings of the 2024 IEEE 32nd International Requirements Engineering Conference (RE)*, Reykjavik, Iceland, June 24-28, 2024; IEEE: Piscataway, NJ, USA, 2024; pp 432-40. DOI
28. Nguyen, T. D.; Pham, L. H.; Sun, J. AUTOLAW: enhancing legal compliance in large language models via case law generation and jury-inspired deliberation. *arXiv* 2025, arXiv:2505.14015. Available online: <https://doi.org/10.48550/arXiv.2505.21281> (accessed 11 November 2025).
29. Feretzakis, G.; Vagena, E.; Kalodanis, K.; Peristera, P.; Kalles, D.; Anastasiou, A. GDPR and large language models: technical and legal obstacles. *Future. Internet*. **2025**, *17*, 151. DOI
30. Jang, M.; Stikkel, G. Leveraging natural language processing and large language models for assisting due diligence in the legal domain. In *2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Mexico City, Mexico, June 16-23, 2024; Association for Computational Linguistics: Stroudsburg, USA, 2024; pp 155-64. DOI
31. Sperotto, F. A.; de Aguiar, M. S. AgentDevLaw: A middleware architecture for integrating legal ontologies and multi-agent systems. In: Cerri R, Prati RC, Eds. *BRACIS 2020, Proceedings of the Brazilian Conference on Intelligent Systems*, Online, October 19-22, 2020; Springer Nature: Cham, Switzerland; 2020; pp 33-46. DOI
32. Sperotto, F. A.; Belchior, M.; de Aguiar, M. S. Ontology-based legal system in multi-agents systems. In *Mexican International Conference on Artificial Intelligence 2019, Proceedings of the 18th Mexican International Conference on Artificial Intelligence*, Xalapa, Mexico, October 27-November 2, 2019; Springer Nature: Cham, Switzerland; 2019; pp 507-21. DOI
33. Bui, M. Q.; Do, D. T.; Le, N. K.; et al. Data augmentation and large language model for legal case retrieval and entailment. *Rev. Socionetwork. Strat.* **2024**, *18*, 49-74. DOI
34. Kalra, R.; Wu, Z.; Gulley, A.; et al. HyPA-RAG: A hybrid parameter adaptive retrieval-augmented generation system for AI legal and policy applications. In *1st Workshop on Customizable NLP: Progress and Challenges in Customizing NLP for a Domain, Application, Group, or Individual, Proceedings of the 1st Workshop on Customizable NLP: Progress and Challenges in Customizing NLP for a Domain, Application, Group, or Individual (customnlp4u)*, Miami, USA, November 16, 2024. Association for Computational Linguistics; Stroudsburg, USA, 2024; pp. 237-56. DOI
35. Li, A.; Wu, Y.; Liu, Y. et al UniLR: Unleashing the power of LLMs on multiple legal tasks with a unified legal retriever. In *63rd Annual Meeting of the Association for Computational Linguistic, Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vienna, Austria. July 27-August 1, 2025; Association for Computational Linguistics: Stroudsburg, USA, 2025; pp 11953-67. DOI
36. Drumond, L.; Girardi, R.; Leite, A. Architectural design of a multi-agent recommender system for the legal domain. In *ICAIL07: 11th International Conference on Artificial Intelligence and Law, Proceedings of the 11th International Conference on Artificial Intelligence and Law*, Stanford California, June 4-8 2007; Association for Computing Machinery: New York, USA, 2007; pp 183-7. DOI
37. Watson, W.; Cho, N. Srishankar, N., et al. LAW: legal agentic workflows for custody and fund services contracts. *arXiv* 2024, arXiv:2412.11063. Available online: <https://arxiv.org/abs/2412.11063> (accessed 11 November 2025).
38. Raptopoulos, P.; Filandrianos, G.; Lymperaio, M. PAKTON: a multi-agent framework for question answering in long legal agreements. *arXiv* 2025, arXiv:2506.00608. Available online: <https://arxiv.org/abs/2506.00608> (accessed 11 November 2025).
39. Wiratunga, N.; Abeyratne, R.; Jayawardena, L.; et al. CBR-RAG: case-based reasoning for retrieval augmented generation in LLMs for legal question answering. In *32nd International Conference on CaseBased Reasoning Research and Development (ICCBR 2024), Proceedings of the 32nd International Conference on CaseBased Reasoning Research and Development (ICCBR 2024)*, Mérida, Mexico, July 1-4 2024; Springer Nature: Cham, Switzerland; 2023; pp 445-60. DOI
40. Nguyen, H. T.; Satoh, K. ConsRAG: minimize LLM hallucinations in the legal domain. In *Frontiers in Artificial Intelligence and Applications*; Savelka, J., Harasta, J., Novotna, T., et al., Eds.; IOS Press: Amsterdam, The Netherlands, 2024. pp 327-32. DOI
41. Li, B.; Fan, S.; Huang, J. CSAFT: continuous semantic augmentation fine-tuning for legal large language models. In *33rd Artificial Neural Networks and Machine Learning - ICANN 2024, Lecture Notes in Computer Science*, LuganoViganello, Switzerland, September 17-20, 2024; Springer Nature: Cham, Switzerland, 2024; Vol. 15020, pp 293-307. DOI
42. Louis, A.; Van Dijk, G.; Spanakis, G. Interpretable long-form legal question answering with retrieval-augmented large language models. *AAAI*. **2024**, *38*, 22266-75. DOI
43. Do Espírito Santo F. O.; Marques Peres S.; De Sousa Gramacho G.; Brandão, A. A. F.; Cozman, F. G. Legal document-based, domain-driven Q&A system: LLMs in perspective. In *2024 International Joint Conference on Neural Networks (IJCNN), International Joint Conference on Neural Networks, IJCNN 2024: Proceedings*, Yokohama, Japan, June 30-July 4, 2024. IEEE; 2024. pp 1-9. DOI
44. Wan, Z.; Zhang, Y.; Wang, Y.; Cheng, F.; Kurohashi, S. Reformulating domain adaptation of large language models as adapt-retrieve-revise: a case study on Chinese legal domain. In *63rd Annual Meeting of the Association for Computational Linguistics (ACL 2024), Findings of the Association for Computational Linguistics: ACL 2024*, Bangkok, Thailand and virtual meeting, July 11-16, 2024; Association for Computational Linguistics: Stroudsburg, USA, 2024; pp 5030-41. DOI
45. El Hamdani, R.; Bonald, T.; Malliaros, F. D.; Holzenberger, N.; Suchanek, F. The factuality of large language models in the legal domain. In *CIKM '24: The 33rd ACM International Conference on Information and Knowledge Management, Short paper at 33rd ACM International Conference on Information and Knowledge Management (CIKM 2024)*, Boise, USA, October 21-25, 2024; Association for Computing Machinery: New York, USA, 2024; pp 3741-6. DOI
46. Al-qaesm, R.; Hendi, M.; Tantour, B. Alkafi-llama3: fine-tuning LLMs for precise legal understanding in Palestine. *Discov. Artif. Intell.* **2025**, *5*, 107. DOI
47. Zhu, S.; Pan, L.; Li, B.; Xiong, D. LANDeRMT: detecting and routing language-aware neurons for selectively finetuning LLMs to machine translation. In *62nd Annual Meeting of the Association for Computational Linguistics, Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Bangkok, Thailand, August 11-16, 2024; Association for Computational Linguistics: Stroudsburg, USA, 2024; pp 12135-48. DOI
48. Li, B.; Fan, S.; Zhu, S.; Wen, L. CoLE: A collaborative legal expert prompting framework for large language models in law. *Knowl. Based. Syst.* **2025**, *311*, 113052. DOI
49. Yao, R.; Wu, Y.; Wang, C.; Xiong, J.; Wang, F.; Liu, X. Elevating legal LLM responses: harnessing trainable logical structures and semantic knowledge with legal reasoning. In *2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, Albuquerque, New Mexico, April 29-May 4, 2025; Association for Computational Linguistics: Stroudsburg, USA, 2025; pp 5630-42. DOI

50. Doyle, C.; Tucker, A. D. If you give an LLM a legal practice guide. In *CSLAW '25: Symposium on Computer Science and Law, Proceedings of the 2025 Symposium on Computer Science and Law*, Munich, Germany, March 25-27, 2025; Association for Computing Machinery: New York, USA, 2025; pp 194-205. DOI
51. Fei, Z. Zhang S.; Shen X.; et al. InternLM-law: an open-sourced Chinese legal large language model. In *31st International Conference on Computational Linguistics, Proceedings of the 31st International Conference on Computational Linguistics*, Dublin, Ireland, January 19-24, 2025; Association for Computational Linguistics: Stroudsburg, USA, 2025; pp 9376-92. <https://aclanthology.org/2025.coling-main.629/> (accessed 2025-11-12).
52. Rivas-Echeverría, F.; Ramos, L. T.; Ibarra, J. L.; Zerpa-Bonillo, S.; Arciniegas, S.; Asprino-Salas, M. LegalBot-EC: an LLM-based chatbot for legal assistance in Ecuadorian law. *arXiv* 2025, arXiv:2506.12346. Available online: <https://arxiv.org/abs/2506.12346> (accessed 11 November 2025).
53. Blair-Stanek, A.; Van Durme, B. LLMs provide unstable answers to legal questions. *arXiv* 2025, arXiv:2502.05196. Available online: <https://arxiv.org/abs/2502.05196> (accessed 11 November 2025).
54. Hannah, G.; Sousa, R. T.; Dasoulas, I.; D'amato, C. On the legal implications of large language model answers: a prompt engineering approach and a view beyond by exploiting Knowledge Graphs. *J. Web. Semant.* **2025**, *84*, 100843. DOI
55. Akbar, K. A.; Uddin, M. N.; Khan, L.; et al. Retrieval augmented generation-based large language models for bridging transportation cybersecurity legal knowledge gaps. *arXiv* 2025, arXiv:2502.18426. Available online: <https://doi.org/10.48550/arXiv.2505.18426>. (accessed 11 November 2025).
56. Pham, H. Q.; Nguyen, Q. V.; Tran, D. Q.; Nguyen, T. B.; Nguyen, K. V. Top2at ALQAC2024: large language models (LLMs) for legal question answering. *Int. J. As. Lang. Proc.* **2025**, *35*, 2450010. DOI
57. Kaczmarczyk, A.; Libal, T.; Smywiński-Pohl, A. A legal assistant for accountable decision-making. In Savelka, J.; Harasta, J.; Novotna, T.; Misek, J. Eds.; *Volume 395: Legal Knowledge and Information Systems*. IOS Press: Amsterdam, The Netherlands, 2024. pp 378-80. DOI
58. Büttner, M.; Habernal, I. Answering legal questions from laymen in German civil law system. In *18th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, St. Julian's, Malta, March 17-22, 2024; Association for Computational Linguistics: Stroudsburg, USA, 2024; pp 2015-27. DOI
59. Fratrič, P.; Holzenberger, N.; Amariles, D. R. Can AI expose tax loopholes? Towards a new generation of legal policy assistants. *arXiv* 2025, arxiv.2503.17339. Available online: <https://arxiv.org/abs/2503.17339> (accessed 11 November 2025).
60. Rebolledo-mendez, J. D.; Tonatiuh Gomez Briones, F. A.; Gonzalez Cardona, L. G. Legal artificial assistance agent to assist refugees. In *2022 IEEE International Conference on Big Data (Big Data), Proceedings of the 2022 IEEE International Conference on Big Data (Big Data)*, Osaka, Japan, December 17-20 2022. IEEE; 2022; pp 5126-8. DOI
61. Cui, J.; Ning, M.; Li, Z.; et al. ChatLaw: a multi-agent collaborative legal assistant with knowledge graph enhanced mixture-of-experts large language model. *arXiv* 2024, arXiv:2306.16092. Available online: <https://arxiv.org/abs/2306.16092> (accessed 11 November 2025).
62. Sun, J.; Dai, C.; Luo, Z.; Chang, Y.; Li, Y. LawLuo: a multi-agent collaborative framework for multi-round Chinese legal consultation. *arXiv* 2024, arXiv:2407.16252. Available online: <https://arxiv.org/abs/2407.16252> (accessed 11 November 2025).
63. Shi, J.; Guo, Q.; Liao, Y.; Liang, S. LegalGPT: legal chain of thought for the legal large language model multi-agent framework. In *International Conference on Intelligent Computing 2024, Lecture Notes in Computer Science*, Singapore, August 1-4, 2024; Springer Nature: Singapore; 2024; Vol. 14880, pp 25-37. DOI
64. Zhang, L.; Ashley, K. D. Mitigating manipulation and enhancing persuasion: a reflective multi-agent approach for legal argument generation. *arXiv* 2025, arXiv:2506.02992. Available online: <https://arxiv.org/abs/2407.16252> (accessed 11 November 2025).
65. Yue, S.; Huang, T.; Jia, Z.; et al. Multi-agent simulator drives language models for legal intensive interaction. In *2025 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2025), Findings of the Association for Computational Linguistics: NAACL 2025*, Albuquerque, New Mexico, April 29-May 4, 2025; Association for Computational Linguistics: Stroudsburg, USA, 2025; pp 6537-70. DOI
66. Shui, R.; Cao, Y.; Wang, X.; Chua, T. S. A comprehensive evaluation of large language models on legal judgment prediction. In *EMNLP 2023, Findings of the Association for Computational Linguistics: EMNLP 2023*, Singapore, December 6-10, 2023; Association for Computational Linguistics: Stroudsburg, USA, 2023; pp 7337-48. DOI
67. Wu, Y.; Zhou, S.; Liu, Y.; et al. Precedent-enhanced legal judgment prediction with LLM and domain-model collaboration. In *2023 Conference on Empirical Methods in Natural Language Processing, Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, Singapore, December 6-10, 2023; Association for Computational Linguistics: Stroudsburg, USA, 2023; pp 12060-75. DOI
68. Prasad, N.; Boughanem, M.; Dkaki, T. Exploring large language models and hierarchical frameworks for classification of large unstructured legal documents. In *European Conference on Information Retrieval (ECIR) 2024, Proceedings of European Conference on Information Retrieval (ECIR) 2024 in part II of the Lecture Notes in Computer Science series*, Glasgow, UK, March 24-28, Springer Nature: Cham, Switzerland, 2024; pp 424-32. DOI
69. Xia, Y.; Luo, X. Legal judgment prediction with LLM and graph contrastive learning networks. In *Proceedings of the 2024 8th International Conference on Computer Science and Artificial Intelligence (CSAI), CSAI 2024: 2024 8th International Conference on Computer Science and Artificial Intelligence (CSAI)*, Beijing, China, December 6-8 2024; Association for Computing Machinery: New York, USA, 2024; pp 424-32. DOI
70. Wang, X.; Zhang, X.; Hoo, V.; Shao, Z.; Zhang, X. LegalReasoner: a multi-stage framework for legal judgment prediction via large language models and knowledge integration. *IEEE. Access.* **2024**, *12*, 166843-54. DOI
71. Nigam, S. K.; Patnaik, B. D.; Mishra, S.; Shallum, N.; Ghosh, K.; Bhattacharya, A. NyayaAnumana & INLegalLlama: the largest Indian legal judgment prediction dataset and specialized language model for enhanced decision analysis. In *31st International Conference on Computational Linguistics, Proceedings of the 31st International Conference on Computational Linguistics*, Abu Dhabi, United Arab Emirates, January 19-24, 2025; Association for Computational Linguistics: Stroudsburg, USA, 2025; pp 11135-60. <https://aclanthology.org/2025.coling-main.738/> (accessed 2025-11-12).
72. Goto, T.; Sano, K.; Tojo, S. Modeling predictability of agent in legal cases. In *2016 International Conference on Agents, Proceedings of the 2016 International Conference on Agents*, Matsue, Japan, September 28-30, 2016; IEEE: Now York, USA; pp 13-14. DOI

73. Ghosh, S.; Verma, D.; Ganesan, B.; Bindal, P.; Kumar, V.; Bhatnagar, V. InLegalLLaMA: Indian legal knowledge enhanced large language model. In *1st International OpenKG Workshop: Large Knowledge-Enhanced Models, Proceedings of the 1st International OpenKG Workshop: Large Knowledge-Enhanced Models*, Jeju Island, South Korea, August 3, 2024; CEUR Workshop Proceedings: Online, 2024; pp 37-46. <https://api.semanticscholar.org/CorpusID:274280428> (accessed 2025-11-12).
74. Su, W.; Yue, B.; Ai, Q.; et al. JuDGE: benchmarking judgment document generation for Chinese legal system. In *48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25), Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25)*, July 13-18 2025; Padua, Italy; Association for Computing Machinery: New York, USA; 2025. pp 3573-83. DOI
75. Jung, S.; Jung, J. Courtroom-LLM: a legal-inspired multi-LLM framework for resolving ambiguous text classifications. In *31st International Conference on Computational Linguistics, Proceedings of the 31st International Conference on Computational Linguistics*, Abu Dhabi, United Arab Emirates, January 19-24, 2025; Association for Computational Linguistics: Stroudsburg, USA, 2025; pp 7367-85. <https://aclanthology.org/2025.coling-main.493/> (accessed 2025-11-12).
76. Hassani, S. Enhancing legal compliance and regulation analysis with large language models. In *2024 IEEE 32nd International Requirements Engineering Conference (RE)*, Reykjavik, Iceland, June 24-28, 2024; IEEE: New York, USA, 2024; pp 507-11. DOI
77. Fernández Martínez, M. d. C.; Fernández, A. AI in recruiting. multi-agent systems architecture for ethical and legal auditing. In *Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19), Proceeding of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*, Macao, China August 10-16, 2019; International Joint Conferences on Artificial Intelligence Organization: Macao, China, 2019; pp 6428-9. DOI
78. Deroy, A.; Ghosh, K.; Ghosh, S. How ready are pretrained abstractive models and LLMs for legal case judgement summarization. *arXiv* 2023, arXiv:2306.01248. Available online: <https://arXiv.org/abs/2306.01248> (accessed 11 November 2025).
79. Vats, S.; Zope, A.; De, S.; et al. LLMs - the good, the bad or the indispensable?: A use case on legal statute prediction and legal judgment prediction on Indian court cases. In *EMNLP 2023, Findings of the Association for Computational Linguistics: EMNLP 2023*, Singapore, December 6-10, 2023; Association for Computational Linguistics: Stroudsburg, USA, 2023; pp 12451-74. DOI
80. Guha, N.; Nyarko, J.; Ho, D. E.; et al. Legalbench: A collaboratively built benchmark for measuring legal reasoning in large language models. In *37th Conference on Neural Information Processing Systems (NeurIPS 2023), Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*, New Orleans, USA, December 10-16, 2023; Neural Information Processing Systems Foundation, Inc.: California, USA; 2023; pp 12451-74. DOI
81. Zhao, X.; Qiao, X.; Ou, K.; et al. HW-TSC at SemEval-2024 Task 5: Self-Eval? A confident LLM system for auto prediction and evaluation for the legal argument reasoning task: In *18th International Workshop on Semantic Evaluation (SemEval-2024), Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, Mexico City, Mexico, June 20-21, 2024; Association for Computational Linguistics: Stroudsburg, USA, 2024; pp 1806-10. DOI
82. Shu, D.; Zhao, H.; Liu, X.; Demeter, D.; Du, M.; Zhang, Y. LawLLM: law large language model for the US legal system. In *CIKM '24: The 33rd ACM International Conference on Information and Knowledge Management, Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, CIKM 2024*, Boise, USA, October 21-25, 2024; Association for Computing Machinery: New York, USA, 2024; pp 4882-9. DOI
83. Shen, C.; Ji, C.; Yue, S.; et al. Empowering LLMs for long-text information extraction in Chinese legal documents. In *CCF International Conference on Natural Language Processing and Chinese Computing, Lecture Notes in Computer Science*, Hangzhou China, November 1-3, 2024; Wong, D.F., Wei, Z., Yang, M., Eds.; Springer Nature: Singapore, 2024; Vol. 15359, pp 457-69. DOI
84. Gray, M.; Zhang, L.; Ashley, K. D. Generating case-based legal arguments with LLMs. In *CS&LAW '25: Symposium on Computer Science and Law, Proceeding of CS&LAW '25: Proceedings of the 2025 Symposium on Computer Science and Law*, Munich, Germany, March 25-27, 2025; Association for Computing Machinery: New York, USA, 2025; pp 160-8. DOI
85. Bali, Y.; Çiloğlu, B. Leveraging large language models for natural language processing based tasks in the legal domain: a short survey. In *International Conference on Computational Science and Its Applications (ICCSA 2025), Proceedings of the International Conference on Computational Science and Its Applications (ICCSA 2025)*, Istanbul, Türkiye, June 30-July 3, 2025; Springer Nature Switzerland AG: Cham, Switzerland, 2025; pp 166-78. DOI
86. Liang, H.; Zhang, X.; Mahari, R.; et al. Leveraging large language models for learning complex legal concepts through storytelling. *arXiv* 2024, arXiv:2402.17019. Available online: <https://arXiv.org/abs/2402.17019> (accessed 4 August 2025).
87. Hijazi, F.; Alharbi, S.; Alhussein, A. et al ArabLegalEval: A multitask benchmark for assessing Arabic legal knowledge in large language models. In *The Second Arabic Natural Language Processing Conference, Proceedings of The Second Arabic Natural Language Processing Conference*, Bangkok, Thailand, August 16, 2024; Association for Computational Linguistics: Stroudsburg, USA, 2024; pp 225-49. DOI
88. Kim, Y.; Choi, Y.; Choi, E.; Choi, J.; Park, H. J.; Hwang, W. Developing a pragmatic benchmark for assessing Korean legal language understanding in large language models. In *The 2024 Conference on Empirical Methods in Natural Language Processing, Findings of the Association for Computational Linguistics: EMNLP 2024*, Miami, USA, November 11-16, 2024; Association for Computational Linguistics: Stroudsburg, USA, 2024; pp 5573-95. DOI
89. Fei, Z.; Shen, X.; Zhu, D.; et al. LawBench: Benchmarking legal knowledge of large language models. In *2024 Conference on Empirical Methods in Natural Language Processing, Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Miami, USA, November 11-16, 2024; Association for Computational Linguistics: Stroudsburg, USA, 2024; pp 7933-62. DOI
90. Li, H.; Chen, Y.; Ai, Q.; Wu, Y.; Zhang, R.; Liu, Y. LexEval: a comprehensive chinese legal benchmark for evaluating large language models. *arXiv* 2023, arXiv:2310.05620. Available online: <https://arXiv.org/abs/2409.20288> (accessed 11 November 2025).
91. Dai, Y.; Feng, D.; Huang, J. et al. LAiW: A Chinese legal large language models benchmark. *arXiv* 2023, arXiv:2310.05620. Available online: <https://arXiv.org/abs/2310.05620> (accessed 11 November 2025).
92. Gan, R.; Feng, D.; Zhang, C. et al UCL-Bench: A Chinese user-centric legal benchmark for large language models. In *2025 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2025), Findings of the Association for Computational Linguistics: NAACL 2025*, Albuquerque, New Mexico, April 29-May 4, 2025; Association for Computational Linguistics: Stroudsburg, USA, 2025; pp 7945-88. DOI
93. Li, H.; Chen, J.; Yang, J. et al LegalAgentBench: evaluating LLM agents in legal domain. In *63rd Annual Meeting of the Association for Computational Linguistic, Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vienna, Austria, July 27-30, 2025; Association for Computational Linguistics: Stroudsburg, USA, 2025; pp 2322-44. DOI

94. Jia, Z.; Yue, S.; Chen, W. et al. Ready Jurist One: benchmarking language agents for legal intelligence in dynamic environments. *arXiv* 2025, arXiv:2507.04037. Available online: <https://arxiv.org/abs/2507.04037> (accessed 11 November 2025).
95. Li, Y. H.; Wu, G. S. LegalEval-Q: a new benchmark for the quality evaluation of LLM-generated legal text. *arXiv* 2025, arXiv:2505.24826. Available online: <https://arxiv.org/abs/2505.24826> (accessed 11 November 2025).
96. Tang, X.; Li, J.; Hu, K. et al. CogniBench: a legal-inspired framework and dataset for assessing cognitive faithfulness of large language models. In *63rd Annual Meeting of the Association for Computational Linguistic, Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vienna, Austria, July 27-30, 2025; Association for Computational Linguistics: Stroudsburg, USA, 2025; pp 21567-85. DOI
97. Deng, C.; Mao, K.; Zhang, Y.; Dou, Z. Enabling discriminative reasoning in LLMs for legal judgment prediction. In *The 2024 Conference on Empirical Methods in Natural Language Processing, Findings of the Association for Computational Linguistics: EMNLP 2024*, Miami, USA, November 11-16, 2024; Association for Computational Linguistics: Stroudsburg, USA, 2024; pp 784-96. DOI
99. Sheik, R.; Siva Sundara, K. P.; Nirmala, S. J. Neural data augmentation for legal overruling task: small deep learning models vs. large language models. *Neural. Process. Lett.* **2024**, *56*, 121. DOI
100. Krumov, K.; Boytcheva, S.; Koytchev, I. SU-FMI at SemEval-2024 Task 5: from BERT fine-tuning to LLM prompt engineering - approaches in legal argument reasoning. In *18th International Workshop on Semantic Evaluation (SemEval-2024), Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, Mexico City, Mexico, June 20-21, 2024. Association for Computational Linguistics: Stroudsburg, USA, 2024; pp 1652-8. DOI
101. Johnson, E.; Holt, X.; Wilson, N. Improving the accuracy and efficiency of legal document tagging with large language models and instruction prompts. *arXiv* 2025, arXiv: 2504.21202. Available online: <https://arxiv.org/abs/2504.09309> (accessed 11 November 2025).
102. Chen, H.; Zhang, L.; Liu, Y.; Yu, Y. Rethinking the development of large language models from the causal perspective: a legal text prediction case study. *AAAI*. **2024**, *38*, 20958-66. DOI
103. Hu, Y.; Gan, L.; Xiao, W.; Kuang, K.; Wu, F. Fine-tuning large language models for improving factuality in legal question answering. *arXiv* 2025, arXiv:2501.06521. Available online: <https://arXiv.org/abs/2501.06521> (accessed 11 November 2025).
104. Zeng, G.; Tian, G.; Zhang, G.; Lu, J. RoSiLC-RS: a robust similar legal case recommender system empowered by large language model and step-back prompting. *Neurocomputing*. **2025**, *648*, 130660. DOI
105. Tang, Y.; Qiu, R.; Huang, Z. UQLegalAI@COLIEE2025: advancing legal case retrieval with large language models and graph neural networks. *arXiv* 2025, arXiv:2505.20743. Available online: <https://arXiv.org/abs/2505.20743> (accessed 11 November 2025).
106. Zhou, Z.; Yu, K. Y.; Tian, S. Y.; et al. LawGPT: knowledge-guided data generation and its application to legal LLM. *arXiv* 2025, arXiv:2502.06572. Available online: <https://arxiv.org/abs/2502.06572> (accessed 11 November 2025).
107. Chen, X.; Mao, M.; Li, S.; Shangguan, H. Debate-feedback: a multi-agent framework for efficient legal judgment prediction: In *2025 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2025), Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)*, Albuquerque, New Mexico, April 29-May 4, 2025; Association for Computational Linguistics: Stroudsburg, USA, 2025; pp 462-70. DOI