AI Open 5 (2024) 181–196

Contents lists available at ScienceDirect

AI Open

journal homepage: www.keaipublishing.com/en/journals/ai-open

Large language models in law: A survey

Jinqi Lai [a], Wensheng Gan [a,*], Jiayang Wu [a], Zhenlian Qi [b], Philip S. Yu [c]

[a] Jinan University, Guangzhou 510632, China
[b] Guangdong Eco-Engineering Polytechnic, Guangzhou 510520, China
[c] University of Illinois Chicago, Chicago, USA

ARTICLE INFO

Keywords:
Artificial intelligence
LLMs
Justice
Legal model

ABSTRACT

The advent of artificial intelligence (AI) has significantly impacted the traditional judicial industry. Moreover, recently, with the development of AI-generated content (AIGC), AI and law have found applications in various domains, including image recognition, automatic text generation, and interactive chat. With the rapid emergence and growing popularity of large models, it is evident that AI will drive transformation in the traditional judicial industry. However, the application of legal large language models (LLMs) is still in its nascent stage. Several challenges need to be addressed. In this paper, we aim to provide a comprehensive survey of legal LLMs. We not only conduct an extensive survey of LLMs but also expose their applications in the judicial system. We first provide an overview of AI technologies in the legal field and showcase the recent research in LLMs. Then, we discuss the practical implementations presented by legal LLMs, such as providing legal advice to users and assisting judges during trials. In addition, we explore the limitations of legal LLMs, including data, algorithms, and judicial practice. Finally, we summarize practical recommendations and propose future development directions to address these challenges.

1. Introduction

Artificial intelligence (AI) (Brynjolfsson and Mcafee, 2017; Fetzer and Fetzer, 1990; Zhang and Lu, 2021) was initially defined as “the science and engineering of making intelligent machines”. Its goal is to enable machines to understand data and make decisions akin to human intelligence. With the advancement of machine learning and the introduction of AI learning frameworks, AI technologies started gaining prominence and finding applications in various domains. Deep learning was formally proposed in 2006 (Goodfellow et al., 2016; LeCun et al., 2015). With the continuous development of technologies such as big data (Shorten and Khoshgoftaar, 2019), the Internet of Things (IoTs) (Laghari et al., 2021; Sun et al., 2023) and AI, deep learning has made significant progress. For example, the IBM question-and-answer robot is widely used in the field of speech recognition (Saon et al., 2015), as well as the emergence of driverless technology. AI-generated content (AIGC) (Cao et al., 2023; Wu et al., 2023) is a new AI technology that follows in the footsteps of professional-generated content (PGC) and user-generated content (UGC) (Tu et al., 2021). AIGC encompasses various techniques, such as generative adversarial networks (GANs) (Pan et al., 2019) and diffusion models (Croitoru et al., 2023).

Foundation models (Bommasani et al., 2021), also known as large models, represent an important direction in the development of AI technologies and were introduced in 2021. Large models are used in text transformation tasks such as machine translation (Sutskever et al., 2014), interactive chat responses (Zhang et al., 2020), and content generation (Liu et al., 2021b). In the legal field, speech recognition can be applied to court transcripts to improve efficiency. In the text-to-image domain, large models can generate images based on user-provided descriptions (Venugopalan et al., 2015). Attention mechanism (Niu et al., 2021) was proposed to address the challenges of large models in processing long sequences. They assign weights to important information by processing input queries, keys, and values. It can enhance the performance of large models (Niu et al., 2021). With the application of attention mechanisms in large models, pre-trained models with different parameter levels continue to emerge. The Transformer (Vaswani et al., 2017) laid the foundation for large language model architectures. In addition, Google introduced the BERT model (Devlin et al., 2019) for large-scale pre-training, and OpenAI developed GPT (Radford et al., 2018). In 2020, GPT-3 (Brown et al., 2020) consists of 5.3 trillion parameters. In 2021, the first trillion-parameter language model, Switch Transformer (Fedus et al., 2022), was established. In 2022, Stability AI released the Diffusion model for

* Corresponding author.
E-mail addresses: jqlai499@gmail.com (J. Lai), wsgan001@gmail.com (W. Gan), csjywu1@gmail.com (J. Wu), qzlhit@gmail.com (Z. Qi), psyu@uic.edu (P.S. Yu).

https://doi.org/10.1016/j.aiopen.2024.09.002
Received 16 November 2023; Received in revised form 24 July 2024; Accepted 26 September 2024
Available online 17 October 2024

text-to-image generation (Ruiz et al., 2023). These LLMs are widely applied in fields (Gan et al., 2023b), such as education (Gan et al., 2023a; Kasneci et al., 2023), robotics (Zeng et al., 2023), and medicine (Hamet and Tremblay, 2017).

In recent years, applying LLMs in law has attracted a lot of attention. This is because LLM can learn massive amounts of knowledge and understand language. It also can reason about cases, assist judges in decision-making, and automatically generate documents. In the long history of law and justice, courtroom rulings with judges taking the lead and the balance of power between prosecution and defense have always been crucial factors in resolving legal cases. The law possesses authority, rigor, objectivity, and normativity, while the judicial process is characterized by periodicity and fairness. However, with the increasing population, the number of judicial cases has also grown, and the imbalance between the number of cases and the available human resources has led to prolonged judicial processes. Relying solely on human judgment is no longer sufficient to meet societal demands. Therefore, the application of artificial intelligence in the field of law has become significant (Rissland et al., 2003). In recent years, with the continuous development of deep learning and other AI technologies, the legal field has witnessed the emergence of more intelligent applications (Chen et al., 2019). For example, some regions have introduced the concept of smart courts, or AI courts, to assist in the adjudication process using AI technology. Legal applications of AI include legal research, document analysis, contract review, predictive analytics for case outcomes, and legal chatbots for providing basic legal information and guidance.

Until now, many legal LLMs with vertical applications in the legal field have been released. For example, LawGPT (Nguyen, 2023) is based on general Chinese-based models (e.g., Chinese-LLaMA), which can answer legal knowledge questions and case logic reasoning. ChatLaw (Cui et al., 2023), released by Peking University in 2023, is trained on a large number of legal news, forums, and judicial interpretation datasets, and can quickly match relevant cases consulted by users. LawGPT_zh, released by Shanghai Jiao Tong University, is an open-source Chinese legal LLM fine-tuned based on ChatGLM-6B LoRA 16-bit instructions, which can automatically generate legal documents and provide user consultation functions. In addition, the evaluation indicators for fine-tuning the legal LLMs model have also caused a lot of discussion, because it is very important to ensure transparency, accountability, and ethical considerations when implementing artificial intelligence in the legal process.

Research gaps. The law and AI applications have widely adopted many related AI technologies (Atkinson et al., 2020; Rissland et al., 2003; Sourdin, 2018; Surden, 2019). The rise of LLMs has brought more possibilities for assisting law management and decision-making. This paper is the first comprehensive review of legal LLMs, introducing the definition, significance, applications, and future. In our paper, we mainly address the following three questions. Furthermore, we propose some suggestions for improving legal LLMs.

- What is the future of law and AI?
- What are the characteristics of legal LLMs and what are their shortcomings?
- Will it be possible to replace the judges' role in the legal system?

Contributions. To fill in the gaps in current research, we first conducted a systematic literature review on the use of LLMs in the judicial field. What's more, we also show our opinions on the future legal LLM system. The main contributions of this paper are as follows:

- To the best of our knowledge, this is the first review article on legal LLMs. We first explain the basic concepts related to AI and law. Following that, we provide a detailed description of the characteristics of legal LLMs, showcasing the difference between traditional judgment and current judgment with AI.
- We demonstrate how judges can make fairer decisions with the help of legal LLMs. We also explore in detail how to promote the optimization of legal LLMs based on the characteristics of big legal data and the intricacies of judicial practice.
- We provide the latest research on legal LLMs, including studies conducted by companies and universities, as well as an in-depth investigation of fine-tuning model techniques and evaluation strategies.
- Finally, we summarize the key challenges and future directions of legal LLMs and provide suggestions and directions for improvement.

Organization. The organization of this paper is shown in Fig. 1. In Section 2, we introduced the key technologies and development history related to LLMs. We provide a brief overview of the traditional judiciary, legal big data, and the main characteristics of AI judiciary in Section 3. In Section 4, we based on the recent popular fine-tuned legal large models and model evaluation methods, focusing on their applications. In Section 5, we analyzed the challenges faced by legal LLMs from the perspectives of legal data, algorithms, traditional judiciary, judicial practice, and human social ethics. Then, we provided specific feasible directions for future research in Section 6. Finally, we presented the conclusion of this survey in Section 7.

2. Key technologies of LLMs

2.1. Related concepts

In this subsection, we explore several related concepts that are fundamental to understanding the field of AI. These concepts encompass various aspects of the foundation model, deep learning, and natural language processing (NLP). Understanding these concepts is crucial for grasping the underlying principles and techniques behind LLMs and their role in advancing NLP tasks.

Foundation Model: The concept of a foundation model, proposed by OpenAI, refers to a large-scale pre-training model (Bommasani et al., 2021). These models are trained on massive amounts of data and can be fine-tuned for various tasks. They provide a solid "foundation" for various applications. Foundation models are powerful models derived from pre-training on large-scale data. They serve as the basis for constructing various specific AI applications, providing a reliable starting point for multiple tasks. These foundation models employ various AI techniques and have achieved remarkable performance in natural language understanding and generation.

Deep Learning (DL) (LeCun et al., 2015) is a type of machine learning algorithm. The term "deep" implies the use of multiple layers in neural networks, enabling computers to learn from a large amount of data and extract useful features and patterns. With multi-layered neural networks, deep learning can handle large-scale data and perform deeper feature learning, classification, and prediction. Typical deep learning frameworks include PyTorch (Paszke et al., 2019), TensorFlow (Abadi et al., 2016), MindSpore (Han et al., 2021), and Caffe (Jia et al., 2014).

Self-attention Mechanism (Shaw et al., 2018) is a crucial technique in deep learning and has been extensively used in processing sequential data in NLP. In the embedding layer, the input sequence is encoded into different vectors. The self-attention mechanism then dynamically assigns and calculates attention weights for each part, and finally sends out the weighted sum as a sequence representation. The self-attention mechanism often employs a multi-head mechanism to capture different relationships and features, including long-range dependencies. It finds wide application in NLP models. For example, the Transformer model incorporates a self-attention mechanism to achieve optimization and breakthroughs (Yang et al., 2022).

Natural Language Processing (NLP) (Chowdhary and Chowdhary, 2020; Nadkarni et al., 2011) enables machines to understand, generate, and manipulate human language, facilitating seamless communication

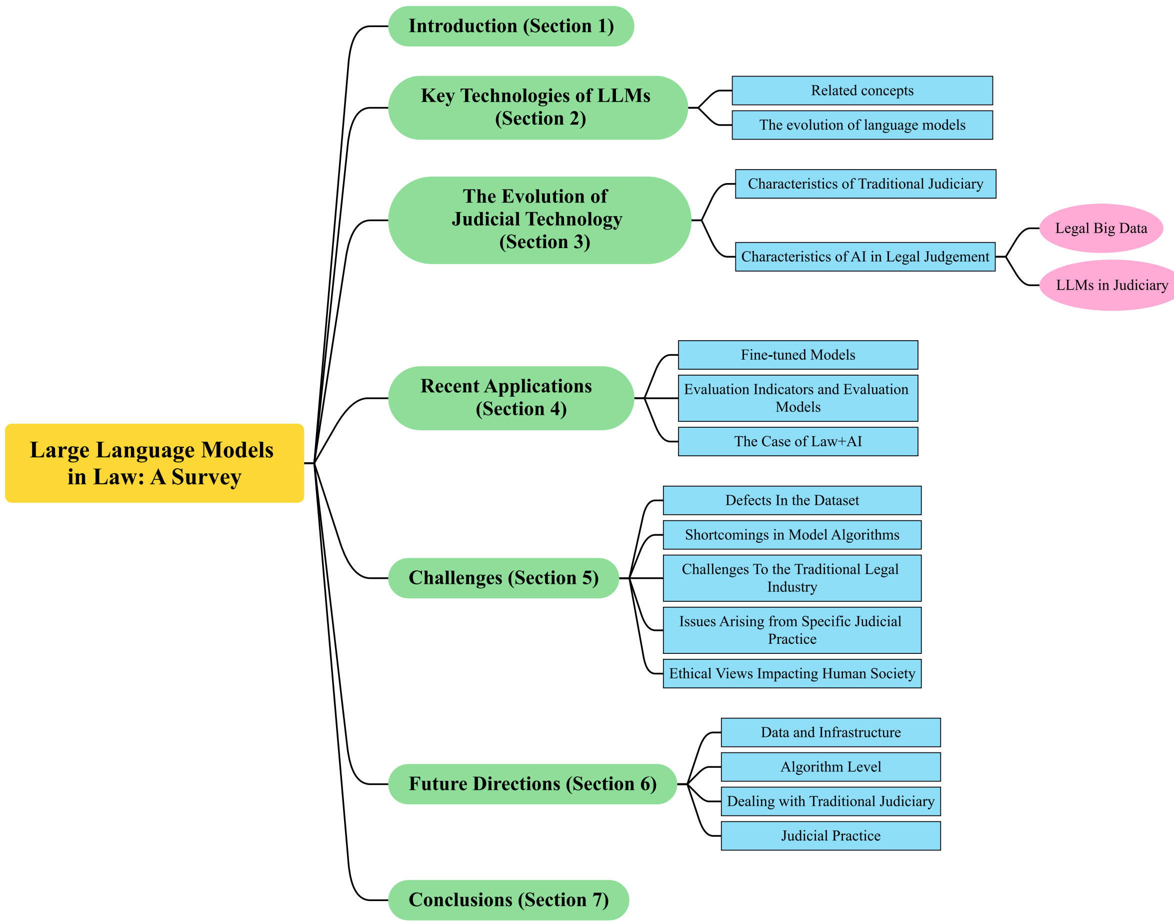

Fig. 1. The outline of our overview.

between humans and machines. NLP has a wide range of applications across various industries. In the legal domain, NLP is utilized as a tool for information extraction from legal documents. It could help to extract specific legal elements, such as case references or contract clauses (Sleimi et al., 2018), from large volumes of legal text.

LLMs (Wei et al., 2022; Zhao et al., 2023) are typically trained on vast amounts of unlabeled data through a pre-training stage (Carlini et al., 2021). The step of pre-training equips LLMs with the capability to perform various NLP tasks (Min et al., 2023), including automatic text generation and translation. What is more, the core technologies utilized in LLMs include fine-tuning and reward modeling. Nowadays, LLMs are widely used for different natural language tasks. They are employed in intelligent translation systems (Brants et al., 2007) enabling accurate translation, such as Google's neural machine translation (NMT) (Wu et al., 2016). Furthermore, LLMs play a crucial role in intelligent customer service and question-answering systems, such as virtual assistants Apple's Siri (Kepuska and Bohouta, 2018) and Amazon's Alexa (Hoy, 2018). These systems could help to understand and respond to user inquiries in a conversational manner. Moreover, LLMs have contributed to advancements in speech recognition such as Microsoft's Cortana (Hoy, 2018) and Google's speech recognition system (Kepuska and Bohouta, 2018). They could help convert spoken language into written text.

2.2. Evolution of LLMs in law

The evolution of language models can be described through six periods. As early as 1948, the emergence of N-gram models (Brown et al., 1992) divided the text into combinations of n-words and used statistical methods to predict the probability of the next word. These models were unable to capture complex language dependencies and semantic structural information. Then, in 1954, the bag-of-words model (Zhang et al., 2010) appeared, treating text as a collection of words but disregarding word order and semantic relationships. In 2003, neural probabilistic language models (Bengio et al., 2003) introduced neural networks in an attempt to improve model performance through the learning capacity of neural networks. However, due to scaling limitations in neural networks and training data, the processing capability of these models remained limited. Subsequently, in 2013, the emergence of the Word2Vec model (Church, 2017) provided a new way to represent the language model's capabilities by training and learning word vectors to capture semantic relationships between words.

The most significant breakthrough occurred in 2018, when pretraining language models such as BERT (Devlin et al., 2019) and GPT (Radford et al., 2018) were introduced. These models, based on the Transformer architecture, were pre-trained on large-scale text data and learned rich semantic and syntactic knowledge (Voita et al., 2019). In 2020, the T5 model was released, which transformed various NLP tasks into text-to-text transformation problems and achieved excellent performance (Ni et al., 2022). Besides, GPT-3 (Brown et al., 2020) became one of the largest pre-training language models, showcasing powerful generation capabilities. Moreover, between 2022 and 2023, models such as GPT-3.5 (Bhaskar et al., 2023) and GPT-4[1] further improved and enhanced the scale, performance, and generation capabilities. The consistent evolution of these language models provides more opportunities and possibilities for AI application in the legal domain.

[1] https://doi.org/10.48550/arXiv.2303.08774

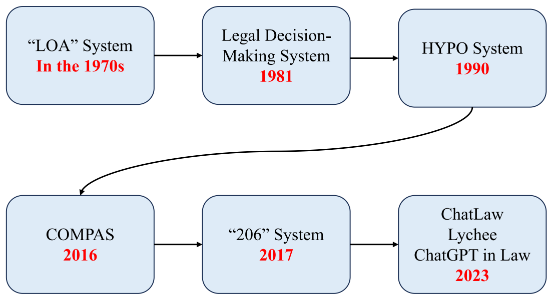

Fig. 2. The evolution of legal LLMs.

Examples of the application of legal LLMs in the field of justice can be described through a timeline, as shown in Fig. 2. In the 1970s, developed countries such as the United States developed legal reasoning systems, legal simulation analysis systems, and expert systems based on AI and applied them to judicial practice (Ashley, 1992). From the 1980s to the 1990s, China also began to develop legal expert systems based on AI technology, with relevant achievements such as the "practical criminal law expert" system and the "LOA Lawyer office automation system" (Aini et al., 2020). However, the functions of these systems are relatively simple, and more are used as data databases, which are limited by the level of hardware and software at that time, and their degree of intelligence is relatively limited. In 1981, Waterman and Peterson developed the LDS (Legal Decision-making System) (Waterman et al., 1986). Since the 1990s, the research focus of extraterritorial judicial artificial intelligence has shifted from the legal expert reasoning model based on legal rules to the legal analysis reasoning model based on judicial cases. In 1990, Professor Asheley of the University of Pittsburgh designed the earliest legal analytical reasoning model, the HYPO (Hypothetical Reasoning) system (Ashley, 1991).

Since 2013, many countries, including China, have begun to gradually carry out the construction of "smart courts" (Shi et al., 2021), using big data and AI technology to assist judges in trials and promote the transparency and visualization of trial processes. In 2016, several states in the United States adopted an AI judicial discretion tool called COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) (Brennan et al., 2009). The system assesses the risk of reoffending through the offender's interview and the information provided by the judicial department, and the evaluation results will serve as a valuable reference in judicial decision-making. In 2017, the "206 System" was born in Shanghai, China, and the large language model was used to assist the trial of criminal cases and help judges achieve fairer sentences (Cui, 2020). For example, by converting audio recordings of interrogations into text and then making simple adjustments, a complete interrogation transcript can be formed.

With the birth of various LLMs in recent years, the application of universal LLMs such as ChatGPT in the legal field has also attracted wide attention. In July 2023, the Peking University team released ChatLaw, a large model of Chinese law, that is vertically applied to the legal field to provide inclusive legal services for the public (Cui et al., 2023). The application of LLMs such as LaWGPT and Lychee in the legal field has also been discussed (Nguyen, 2023).

3. Evolution of judicial technology

In this section, we explore the development of judicial AI. By analyzing the shortcomings of traditional judiciary and the characteristics of legal big data, we further uncover the important features of the LLM judiciary and discuss them in depth.

3.1. Characteristics of traditional judiciary

We explore some characteristics of the traditional judiciary that are crucial for our subsequent understanding of the application of judicial artificial intelligence (Re and Solow-Niederman, 2019). These characteristics include reliance on human decision-making, a lack of flexibility, and resource consumption, among others. They are outlined as follows:

Relying on Human Decision-making: Traditional judicial systems primarily rely on the human decision-making of judges, prosecutors, and lawyers, including case hearings, judgments, and legal interpretations. During the process of reasoning and evidence collection in a case, they often need to refer to the specific circumstances of the case, legal provisions, and past precedents, combined with their professional knowledge, to make judgments and decisions. Finally, the judgment or defense is carried out in a trial.

Basing on Precedent: Traditional judiciary often relies on precedents in the decision-making process (Baude, 2020), such as previous judgments in similar cases and relevant provisions of laws. In many judicial systems, the highest court's judgments have authority and binding effect, guiding other courts. For example, the decisions of the highest court are considered authoritative for constitutional interpretation and legal application, and other courts often refer to those decisions in relevant cases.

Lacking Flexibility: In the field of the judiciary, when there is uncertainty within the legal norms or vague boundaries of legal concepts, judges need to make judgments and choices based on specific contexts rather than mechanically applying the law (Easterbrook, 1987). For example, when dealing with contracts, if a clause includes the concept of "reasonable time" but the contract does not provide a specific definition, the judge must determine what constitutes "reasonable time" according to the specific circumstances. In this case, the judge should consider the nature of the contract, the relationship between the parties, industry standards, and other relevant factors. As a result, judges have the ability to handle judicial cases with flexibility.

Time and Resource-consuming: Traditional judiciary requires a significant amount of human resources and time when dealing with a large number of cases. This leads to situations where there are many cases but few personnel, which can prolong the trial process. For example, the process of conducting case hearings, summoning witnesses, and collecting evidence may consume a considerable amount of time and resources. Similar hierarchical systems exist in other countries, too. For instance, the judicial system in the United Kingdom adopts a hierarchical system (Hanretty, 2020), including the Magistrates' Court, the Crown Court, and the Supreme Court.

3.2. Characteristics of AI in legal judgment

When we talk about the relationship between legal LLMs, judges, and users, the entire process is shown in Fig. 3. Legal LLMs can match similar cases to assist judges in decision-making, provide complete legal consultation for users, and so on. Legal professionals can use the logical reasoning capabilities of legal LLMs to understand the case process, assist judges in decision-making, quickly identify similar cases through language comprehension, analyze and summarize key case details, and use automated content generation capabilities to draft repetitive legal documents. By alleviating the issue of "too many cases, too few people", these AI systems can enhance judicial efficiency and quality (Re and Solow-Niederman, 2019). Some characteristics of legal big data and LLMs are shown below:

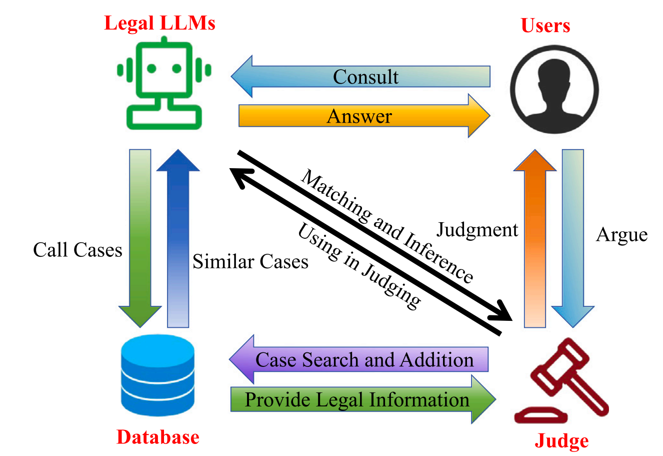

Fig. 3. The basic relationship between Users, Judge, Database, and Legal LLMs.

3.2.1. Characteristics of legal big data

Processing datasets is an important part of training a large model. Legal big data, compared to other datasets, exhibits non-structured, multi-sourced, timely, and privacy and security features, among others, which have garnered attention.

Unstructured: Legal data often exhibits unstructured characteristics, such as legal concepts, legislative texts, judgments, legal commentaries, etc. The textual data formats are inconsistent and difficult for computers to understand and process (Adnan and Akbar, 2019). Therefore, NLP and text analysis techniques are required to extract useful information and transform it into structured data that can be understood by AI.

Multilingual and Multicultural: Law covers multiple languages and cultures (Šarčević, 2016), so legal big data may involve texts in different languages. Cross-lingual analysis of legal data necessitates addressing issues such as translation, cultural differences, and legal terminology variations. For example, European Union regulations often have multiple official language versions, such as English, French, German, etc. Legal researchers need to compare different language versions of regulations to ensure an accurate understanding of their meanings.

Vast Scale and Complexity: Legal data usually contains a large amount of text, such as hundreds of pages of legislation or judgments. Moreover, legal data exhibits diverse characteristics, with each domain having its unique regulations and types of legal documents, requiring different analysis methods and specialized knowledge. For instance, the field of intellectual property involves various legal issues, including patents, trademarks, copyrights, etc. These datasets are extensive and require powerful computing capabilities for processing and analysis. Some legal data, such as legal statutes, can be complex, and due to their large volume and diversity, AI needs robust computational power for handling and analysis. For example, tax law is highly intricate, encompassing thousands of pages of tax regulations and judicial precedents. This requires large-scale legal databases and advanced search tools to assist tax professionals in research and analysis (Zhong et al., 2020b).

Timeliness: Law is an ever-changing field, with legal documents and regulations frequently revised and updated. Therefore, legal big data must be regularly updated to reflect the latest legal provisions. For example, in tax law, the government may introduce new tax regulations to adapt to economic changes. Thus, legal professionals and tax experts need to keep their research and advice up-to-date.

Data Multi-sourcing: Legal big data can come from multiple sources, including court records, government files, legal databases, and social media. Integrating data from these different sources and ensuring data consistency is a challenge. For example, legal researchers may need to access both federal and state court records to obtain comprehensive case information, which requires integrating data from different sources.

Privacy and Security Concerns: Legal data may contain labeled sensitive information, such as personal identifying information, details of privacy-related cases, etc. Therefore, operations like anonymization are needed to remove privacy labels during the collection, storage, and analysis of legal big data. Moreover, protecting user privacy is crucial (Liu et al., 2021a). For example, in criminal cases, court documents may contain personal identification information as well as the defendants' criminal histories. These pieces of information require strict privacy and security protection to prevent unauthorized access and disclosure.

3.2.2. Characteristics of LLM in judiciary

With the rapid development of AI technologies such as AIGC, the application of legal LLMs in the field of judiciary has become widespread. They exhibit the following characteristics, similar to Xu et al. (2024), as shown in Fig. 4:

Understanding Language: LLMs have the ability to interact with users and establish contextual relationships to perform various tasks (Du et al., 2022). Through deep learning on extensive legal datasets, LLMs can analyze and modify legal documents, checking grammar, symbols, sentence structures, etc. They can also extract key elements from legal documents, such as the disputed issues in a case, applicable laws, the identities of the parties involved, the relevance of evidence, and more. These extracted features are crucial inputs for legal decision-making algorithms. LLMs can quickly extract key points from legal documents, combine them with the judgment outcomes, and generate concise and accurate case summaries. Legal professionals can utilize LLMs to extract key points from legal documents, combine them with judgment outcomes, and generate concise and accurate case summaries, reducing time and effort while still producing high-quality work.

Generating Content: LLMs can automatically generate legal documents (Kanapala et al., 2019), case reports, legal contracts, etc. By inputting basic information about a legal case, such as party details, legal grounds, and evidence, AI algorithms can generate a draft legal document that complies with legal standards. For example, a lawyer can input the basic information of a legal case into an AI system, which will then generate an initial draft of a legal document containing relevant clauses, conditions, and vocabulary. The lawyer can review and modify the draft, saving time and effort while maintaining high-quality work. Furthermore, AI technologies can structure text, such as Markdown text, JSON text, Excel text, and so on. Users can input legal report files and specify the desired text format, generating structured texts that meet their requirements.

Applying Speech-to-Text: In traditional court proceedings, judges need to handwrite meeting records and submit them to typists to complete information recordings. With the continuous development of AI technology, speech-to-text conversion (Reddy and Mahender, 2013) has become widely used, reducing the incompleteness of manual records. AI-based speech-to-text platforms (Trivedi et al., 2018) also increase the transparency of AI involvement in decision-making. For example, in some foreign courts (Etulle et al., 2023), the process of speech-to-text conversion (Trivedi et al., 2018) is fully presented in the courtroom, ensuring the credibility of meeting records and increasing public trust in the judiciary. AI can also play an assisting role in judicial decision-making for judges (Xu, 2022).

Providing Legal Consultation: In the field of justice, LLMs can interact with users. Therefore, users can ask legal questions to the model and receive answers and suggestions based on the training data (Greenleaf et al., 2018). This approach can provide convenient and efficient legal consultation services for users, while also reducing the workload of professional lawyers. For example, in a civil lawsuit, the user can ask the model multiple questions, and the model can provide better judgment plans based on legal knowledge (Zhong et al., 2020a).

Matching Optimal Solutions for Cases: AI can extract the key features of a given case, deeply mine a large number of historical

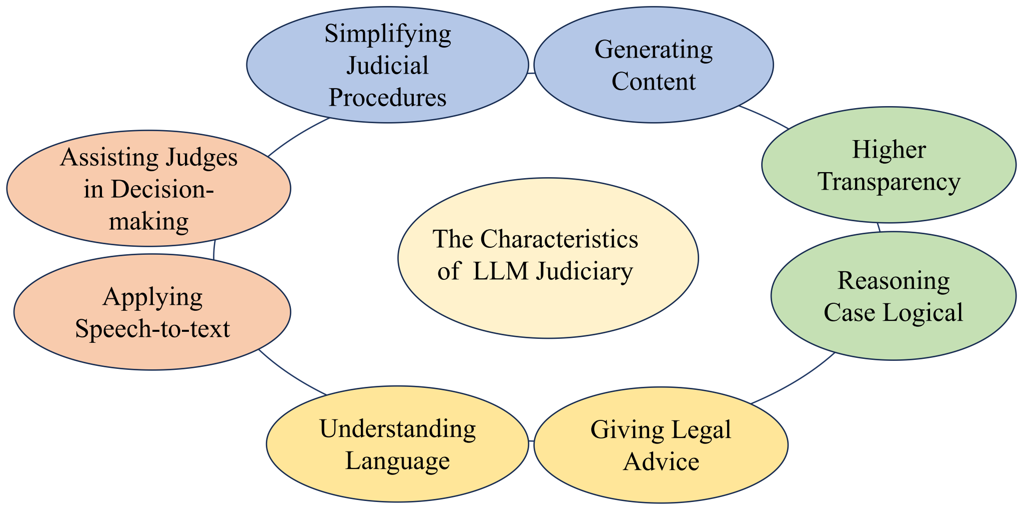

Fig. 4. Characteristics of LLM in Judiciary.

cases and judgment outcomes, and obtain the optimal solution for the case (Kallus, 2020). It can assist in comprehensive evidence-searching. Furthermore, judicial discretion models can capture a case's detailed points based on its features, maximizing judicial fairness.

Reasoning Case Logic: Interactive AI can perform case logic reasoning to a certain extent based on multi-round prompts from users and given case-related information (Atkinson et al., 2020). It analyzes a case's evidence chain by extracting relevant elements from the input legal documents and mining internal details. Legal LLMs conduct comprehensive analysis based on metrics such as the relevance, credibility, validity, and completeness of evidence, thus establishing a complete evidence chain for the case. Similarly, AI evaluates the authenticity and detail of case facts by analyzing the input legal text information and the feasibility of the facts. For example, in some areas, expert inspections have revealed inconsistencies in case facts within the dataset used to train ChatGPT. Legal LLMs, based on deep learning from extensive datasets, possess a certain degree of decision-making capability. It can propose judicial decision suggestions based on facts and legal regulations to assist judges in decision-making. For example, it includes the implementation of "smart courts" (Shi et al., 2021) and the popularity of legal LLMs (Zhong et al., 2020b).

Simplifying Judicial Procedures - Improving Judicial Efficiency: The "too many cases, too few people" issue has long been a problem in the judiciary. Much of the work of legal professionals relies on repetitive document-based tasks. AI technologies, focusing on how to incorporate more comprehensive legal big data into systems (Zhong et al., 2020b), have simplified judicial procedures and, to some extent, improved judicial efficiency (de Sousa et al., 2022). For example, AI applications in judicial scenarios have reduced judges' administrative work (Sourdin, 2018) and increased trial efficiency (de Sousa et al., 2022).

4. Recent applications

In this section, we combine the analysis of the latest popular ten fine-tuned legal LLMs to examine their distinct characteristics in handling legal matters. We then present feasible model evaluation metrics and methods, as well as discuss some AI legal case studies.

4.1. Fine-tuned models

We compared 10 recently popular legal LLMs, as shown in Table 1, summarizing the early basic models of these legal LLMs and the model's website. In addition, we compared these fine-tuned models from the perspective of the main functions of the models and found that most legal LLMs have the function of consulting and answering questions, and some legal LLMs perform better in explaining legal concepts and predicting judicial decisions. We also summarize the model characteristics of different legal LLMs from the perspective of training datasets.

LawGPT_zh: It is an open-source Chinese legal LLM based on ChatGLM-6B LoRA 16-bit instruction fine-tuning, released by Shanghai Jiao Tong University. The datasets include the existing legal question-and-answer data and high-quality legal question-and-answer text constructed by self-instruct based on legal provisions and real-life case guidance. The computing power is 4x NVIDIA 3090. This model enhances the performance, reliability, and professionalism of general language models in the legal domain.

LawGPT (Nguyen, 2023): It is based on general Chinese-based models (for example Chinese-LLaMA, Chinese-Alpaca-Plus-7B, etc.). This model contains official data such as judgments from the China judgment documents network, judicial examination data, crime prediction data, and legal question-and-answer data. It expands the scope of legal terminology and performs pre-training on large-scale Chinese legal text to enhance the basic semantic understanding abilities of large models in the legal domain. The computing power is 8× NVIDIA Tesla V100 32 G. LawGPT is also fine-tuned on legal dialogue question-and-answer datasets and judicial examination datasets to strengthen understanding and execution abilities in legal contexts.

LexiLaw: It is fine-tuned based on the ChatGLM-6B architecture. It enhances its performance and professionalism in providing legal consultation and support through fine-tuning legal domain datasets. The data sources include BELLE-1.5M general domain data, fine-tuned legal question and answer data, 50K legal documents, laws and regulations, and legal reference books. This model has 7x NVIDIA A100 GPUs and 40 GB of computing power. LexiLaw aims to provide accurate and reliable legal consultation services for legal professionals, students, and general users, addressing specific legal issues, legal articles, case analyses, and legal interpretations, as well as offering helpful recommendations and guidance.

Lawyer LLaMA (Huang et al., 2023; Touvron et al., 2023): As a Chinese legal LLM trained on a large-scale legal dataset, it first conducted pre-training on the legal corpus to learn the Chinese legal knowledge system systematically. On this basis, ChatGPT is used to collect a batch analysis of objective questions on China's National Unified Legal Professional Qualification Examination and responses to legal consultations. The collected data is used to fine-tune the model's instructions, allowing the model to learn to apply legal knowledge to capabilities in specific scenarios. Lawyer LLaMA can provide legal advice, analyze legal cases, and generate legal articles.

Table 1
The Fine-tuned models.

Legal LLMs	Traditional model	Main function	Website
LawGPT_zh	ChatGLM-6B LoRA 16 bit	Consultation Q&A	https://github.com/LiuHC0428/LAW-GPT
LaWGPT	Chinese-Alpaca-Plus-7B	Consultation Q&A	https://github.com/pengxiao-song/LaWGPT
LexiLaw	ChatGLM-6B	Consultation Q&A	https://github.com/CSHaitao/LexiLaw
Lawyer LLaMA	LLaMA	Explain legal concepts	https://github.com/AndrewZhe/lawyer-llama
HanFei	BLOOMZ-7B1	Explain legal concepts	https://github.com/siat-nlp/HanFei
ChatLaw	Jiangzi-13B, Anima-33B	Writing legal documents	https://github.com/PKU-YuanGroup/ChatLaw
Lychee	GLM-10B	Consultation Q&A	https://github.com/davidpig/lychee_law
WisdomInterrogatory	Baichuan-7B	Consultation Q&A	https://github.com/zhihaiLLM/wisdomInterrogatory
JurisLMs	AI judge, AI Lawyer	Judge prediction	https://github.com/seudl/JurisLMs
Fuzi.mingcha	ChatGLM	Case analysis	https://github.com/irlab-sdu/fuzi.mingcha

HanFei: This legal LLM is a fully parameterized Chinese legal LLM with 700 million parameters. The model is pre-trained using 60G of cases, regulations, indictments, and legal news (2K tokens). It fine-tunes 53k Chinese general instructions, 41k Chinese legal instructions, and so on. It is assessed through legal issues spanning nine modules, including labor, marriage, and so on. This model has 8x NVIDIA AI00/A800 computing power. It can provide legal question and answer, multi-round dialogue, article generation, search, and other functions.

ChatLaw: A series of open-source legal LLMs developed by Beijing University (Cui et al., 2023). It includes models such as ChatLaw-13B and ChatLaw-33B, which are trained on a large dataset of legal news, forums, and judicial interpretations. This model's computing power is made up of multiple NVIDIA V100 GPUs. ChatLaw-Text2Vec uses a dataset of 930,000 court cases to train a similarity-matching model. In addition, it can match user question information with corresponding legal provisions.

Lychee: A fine-tuned Chinese legal LLM based on the Law-GLM-10B architecture and fine-tuned on 30 GB of Chinese legal data. It offers better performance and professionalism in legal consultation and support.

WisdomInterrogatory: A legal LLM is based on Baichuan-7B, developed by Zhejiang University, Alibaba, and Hua Research. The pre-training data, which totals 40G, includes legal documents, judicial cases, and legal question-and-answer data. It uses a pre-training model and domain-specific data to perform legal question-answering and reasoning.

JurisLMs: A collection of AI judges trained on Chinese legal datasets. AI Judge is an explainable legal judgment prediction model that combines a GPT-2 model with a legal applicability model. It can provide not only judgment results but also corresponding court opinions. AI Lawyer is an intelligent legal consultation model that can answer questions and provide relevant legal regulations.

Fuzi.mingcha: A model based on the ChatGLM architecture, trained on a large dataset of unsupervised Chinese legal texts and supervised legal fine-tuning data. It supports functions such as legal article search, case analysis, deductive reasoning, and legal dialogue.

4.2. Evaluation indicators for legal LLMs

In August 2023, a joint proposal was released by the Smart Judicial Technology Chief System, Zhejiang University, Shanghai Jiao Tong University, and Alibaba Cloud.[2] The proposal suggests a "pretrain-prompt-predict" learning model, which utilizes prompt words to fine-tune pre-training models and assess their performance.

4.2.1. Composition of evaluation indicators

The proposal suggests a comprehensive evaluation system that combines subjective and objective indicators (Xu et al., 2024). The subjective indicators are evaluated by legal experts, whereas the objective indicators are weighted and scored using a combination of weights and ratings. The evaluation system for legal LLMs consists of two levels, with the first level including four primary indicators: functional indicators, performance indicators, safety indicators, and quality indicators, as shown in Fig. 5. The second level of evaluation indicators further breaks down each primary indicator into more detailed sub-indicators.

4.2.2. Calculation of evaluation indicators

When we talk about how to specifically quantify the evaluation indicators of legal LLMs, the proposal provides us with calculation formulas for specific indicators.[3] The following are examples of each type of indicator:

(i) **Quality Index**: It is divided into two components: the reliability index and the maintainability index (Alavian et al., 2020). The reliability index is calculated as follows: $MTBF = \frac{T}{F}$, where T represents time and F represents the number of failures of legal LLMs within this period of time. The maintainability index is calculated as $MTBR = \frac{1}{N}\sum_{i=1}^{N} t_i$, where N represents the number of legal LLM failures included in the statistics and t_i represents the recovery time of legal LLMs after the ith failure. If $N = 0$, $MTBR = 0$.

(ii) **Safety Indicators**: They must count whether there is prohibited content or questionable item content. Among them, prohibited content counts whether there is any content marked as "prohibited" in each security category set. The specific calculation is as $F = H$, where H indicates whether there is content manually marked as "prohibited". If so, the H value is 1; otherwise, it is 0.
Question item content counts the question items in each security category and then calculates the proportion as $P_f = \frac{H}{N}$, where H represents the number of question items in each security category, and N represents the number of results in each security test set.

(iii) **Performance Index**: It includes the calculation of eight indicators, including F1, which measures accuracy, initial response time, which measures time characteristics, and processing efficiency. For example, the F1 indicator is specifically calculated as follows: $P = \frac{TP}{TP+FP}$, $R = \frac{TP}{TP+FN}$, and $F1 = \frac{2PR}{P+R}$, where TP represents the number of positive samples that predict correctly; FP predicts positive samples, but the actual number is negative samples; FN predicts negative samples, but the actual number is positive samples.

(iv) **Functional Indicators**: They include four types: language understanding, content generation, knowledge question-and-answer, and logical reasoning. Specifically, it includes 12 functions such as legal text inspection, legal element extraction, and judicial decision-making reasoning. In this proposal, the total number of functional indicators for legal LLMs is calculated as $FuncScore = \sum iC_i$, where C_i represents the existence of the ith function, 1 if it exists, 0 if it does not exist.

[2] https://github.com/liuchengyuan123/LegalLLMEvaluation

[3] https://github.com/liuchengyuan123/LegalLLMEvaluation

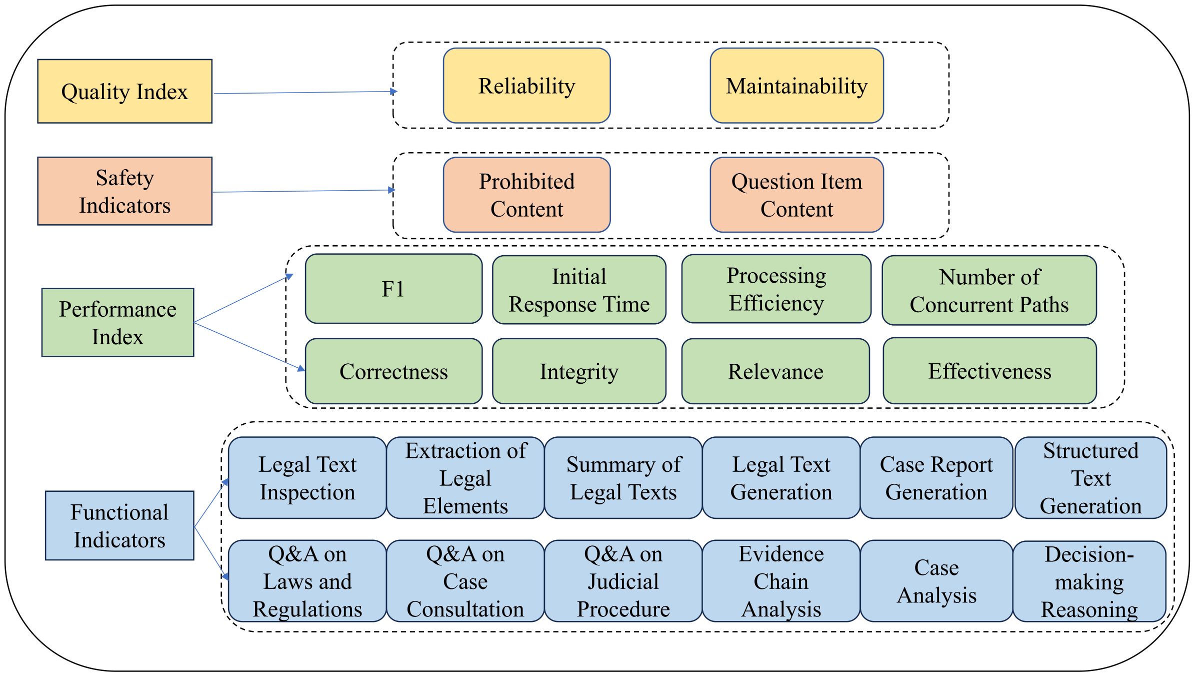

Fig. 5. There are various indicators for AI and law systems.

4.3. The cases of Law+AI

With the increasing application of AI in the judicial field, judicial AI systems are gradually becoming popular. Talking about the cases of Law+AI, the specific applications of legal LLMs can be divided into effectiveness and flaws.

4.3.1. Effectiveness of legal LLMs

Effectiveness is an important issue. For example, applications in China include the "206" criminal case assistance system in Shanghai (Cui, 2020), the "Mobile Micro-Court" in Zhejiang (Shi et al., 2021), the Hangzhou Internet Court's smart judgment system (Sung, 2020) and the ChatLaw legal LLM released by Beijing University in 2023. In addition, AI is more commonly used to assist judges in making decisions and assessing the risk of criminal behavior. For example, in 2023, a court in Colombia used the ChatGPT decision-making result to assist judges in sentencing a medical insurance case (Aydin and Karaarslan, 2023). The HART intelligent system in the UK assists judges in criminal convictions (Greenstein, 2022), and the ProKid 12-SI system in the Netherlands assesses the possibility of recidivism (La Fors, 2020).

Legal LLMs have also been used to predict judges' decisions. For example, the Illinois Institute of Technology and South Texas College of Law jointly developed an algorithm based on the U.S. Supreme Court database and analyzed 28,000 decisions of U.S. Supreme Court justices from 1816 to 2015, 240,000 votes were used for prediction, and the accuracy rates reached 70.2% and 71.9%, respectively, which was higher than the 66% prediction accuracy rate of legal experts (Morison and Harkens, 2019).

Legal LLMs can complete some simple and repetitive tasks (Aini et al., 2020; Nowotko, 2021), allowing more human judges to focus on handling more complex cases. For example, Luminance, a British start-up company, has developed an artificial intelligence platform called Luminance AI, which can automatically process large amounts of documents and quickly generate contract content, thus saving a lot of time and labor costs. A Danish startup called Legal Robot has developed an artificial intelligence platform called "Contract Mill" that can automate the review and revision of contract documents through natural language processing technology. Alibaba's legal service platform, "Faxin", uses artificial intelligence technology to provide citizens with intelligent legal consulting services (Stern et al., 2020). Users only need to enter relevant information to get professional answers to specific questions.

4.3.2. Flaws in applications of legal LLMs

In some cases of judicial application of legal LLMs, legal LLMs have also shown some shortcomings. For example, an AI lawyer in the UK called DoNotPay was able to automatically draft simple appeal applications (Sparkes, 2023), but when it came to more complex cases, it made errors and inaccuracies. In recent years, some states in the United States have conducted research on cases involving police violence and found that the results of artificial intelligence judgments often involve racial discrimination (Zuiderveen Borgesius, 2020). In addition, online legal service platforms developed by enterprises may leak customers' personal information and sensitive data.

As a result, legal LLMs are effective in the judicial field, but they also exhibit some shortcomings. We will discuss the challenges faced by legal LLMs in detail in the following sections.

5. Challenges

Legal LLMs still face many challenges, including the defects of the datasets and shortcomings in the models/algorithms. The impact on the traditional legal industry and issues arising in specific judicial practices are also significant challenges for legal LLMs. Let us introduce more knowledge about this aspect.

5.1. Defects in the datasets

The development level of judicial AI is determined by the legal dataset used for training, especially the annotated dataset. In terms of the dataset, judicial AI primarily entails obtaining more representative, comprehensive, and high-quality annotated datasets. This includes addressing unfair elements in the dataset and potential privacy leaks, ensuring the timeliness of legal datasets, and emphasizing the dataset's scalability.

Data Inadequate Acquisition: The success of legal LLMs depends heavily on data. Legal big data is characterized by variability and non-structured data. Therefore, the construction of successful legal

LLMs relies on the acquisition, organization, and deep learning of legal big data. Given this, the inadequate acquisition of legal data poses a challenge to legal large models.

(i) Insufficient sources of judicial data and documents: In the field of AI and law, the current large-scale legal datasets are insufficient (Chalkidis and Kampas, 2019). To prevent the leakage of legal data, courts tend to be conservative in terms of the openness of judicial data and legal documents, resulting in a limited variety of legal procedural documents. Consequently, deep learning by legal LLMs cannot cover all legal datasets, which may lead to judicial errors such as misjudgment in practical judicial applications.

(ii) Insufficient sharing of legal data (Tenopir et al., 2011): Due to inconsistent sharing permissions for legal big data among different levels of courts, the legal data used by LLMs is not standardized. Generally, higher-level courts have stricter sharing permissions compared to their subordinate branches, and lower-level courts cannot access data from higher-level courts. Moreover, constrained by sharing technologies, difficulties in data integration, incompleteness, inaccuracies, and outdated legal data can lead to decision-making errors in legal LLMs.

(iii) Non-standard legal documents: Furthermore, due to the lack of standardized review or requirements for legal documents in some courts and the insufficient responsibility of some legal professionals, some legal documents do not meet the standards. Inaccurate and non-standardized legal documents are difficult for algorithms within legal LLMs to recognize, resulting in incomplete data acquisition.

Inaccurate Interpretation of Legal Concepts: Legal concepts can be classified into evaluative concepts, descriptive concepts, and discursive concepts (Hage and von der Pfordten, 2009). However, legal concepts have inherent uncertainty in their connotations and extensions. When dealing with legal concepts, current AI systems have recognition deficiencies, which can lead to inappropriate assumptions and deficiencies in concept identification. For example, big data may derive unknown conclusions or misinterpret the boundaries of legal concepts. When applying specific legal concepts, judges typically consider the specific context and value assessment, which is difficult for legal LLMs to achieve. Recognizing and structuring fuzzy boundary concepts in law poses a significant challenge to the development of legal LLMs.

Characteristics of the Datasets: Legal data is semantically rich and has unique characteristics that make it different from data in other fields.

(i) *Timeliness*: In the legal system, the specific application and meaning of legal concepts may evolve and improve over time, depending on the location. When the judicial datasets used for training legal LLMs do not include these subtle changes, the results generated by AI may not be suitable for the latest legal environment, leading to judicial misjudgments.

(ii) *Credibility*: Due to the variety and complexity of laws, the large number of judicial cases and documents, and the complexity of the training process for legal datasets, biased elements in data acquisition, inadequate selection of inconsistent, incomplete, or erroneous data, improper privacy protection in the datasets (De Capitani Di Vimercati et al., 2012), and imbalanced training of the datasets may exist. For example, civil litigation cases may be more common than criminal cases or other types, resulting in imbalanced training datasets and poor performance of judicial AI in certain cases.

(iii) *Scalability*: Currently, the scale of the datasets is limited, and some legal datasets only include cases from specific time periods or the application of individual legal articles at specific levels, making it difficult to extend to other aspects.

5.2. Shortcomings in algorithms

LLMs in the field of judiciary have garnered significant attention regarding algorithmic interpretability, ethics, bias and fairness, and algorithmic optimizability. In this subsection, we investigate the shortcomings of LLMs in these areas.

Interpretability: The use of deep learning, neural networks, and other algorithms in LLMs makes their structures complex, and the results of their decisions are difficult to predict. Current LLMs cannot achieve complete value neutrality, and the judiciary's authority is based, to some extent, on the public's acceptance of case-handling outcomes. The insufficient interpretability of large models undoubtedly reduces people's trust in the judicial application of AI (Atkinson et al., 2020). Therefore, interpretability is a key challenge for legal LLMs, and it is crucial to reduce the black-box nature of legal LLMs and improve their interpretability.

Ethics, Bias, and Fairness: The fairness and security of algorithms behind legal LLMs have received significant attention (Wachter et al., 2021).

(i) Algorithms may contain elements of inequality. Some data may contain elements of gender or racial discrimination, which may also be influenced by historical factors. For example, the COMPAS algorithm made different prediction errors for white and black individuals (Flores et al., 2016). Black individuals were erroneously predicted as high-risk, while white individuals were erroneously predicted as low-risk. Such algorithmic discrimination may exacerbate racial bias. The National Institute of Standards and Technology categorizes these biases into statistical bias, human bias, and systemic bias (Schwartz et al., 2022). However, when AI learns from this data, it cannot judge the fairness of the data and therefore cannot eliminate unfair factors. In practical judicial applications, these unequal elements may unintentionally be amplified by LLMs, resulting in unfair decisions.

(ii) Insufficient security in algorithm outsourcing: Due to a lack of legal-technical hybrid talent, the algorithmic part of legal LLMs is often outsourced to algorithm companies, which unintentionally reduces transparency in legal interpretations. The algorithms may be exposed to "biased" elements, or elements that do not align with human values, which are then "technically cleansed" by AI, hiding these unjust elements. This leads to issues such as "algorithmic black boxes" (Yu and Alì, 2019). For example, when an AI system provides a judgment in a significant case, the party being judged may find it difficult to be convinced by the decision if the system cannot provide a reasonable explanation for the underlying reasons or algorithmic principles behind the ruling.

(iii) The reduced transparency of LLMs in law may lead to significant issues of judicial unfairness (Rai, 2020), and trust in the judiciary may decrease. During the development stage of legal LLMs, the handling of most legal concepts and unstructured data such as legal documents, including collection, cleaning, annotation, and processing, may introduce elements of inequality. For example, if AI does not have a comprehensive understanding of sensitive information (such as race), it may be challenging to completely remove sensitive elements during the cleaning process of legal datasets, incorporating biased elements into the decision-making model. During the testing stage of the model, developers may make adjustments to the model's decision results based on their own cultural background and scope of understanding. However, developers may not be legal professionals, resulting in biases in the understanding of legal concepts and specific judicial discretion. This can lead to "blind spot bias", such as racial bias in the Gangs Matrix's prediction of potential offenders (Densley and Pyrooz, 2020).

(iv) If the algorithmic bias or lack of interpretability behind legal LLMs cannot be fixed and becomes a major cause of AI judicial errors, people will not trust the use of legal LLMs in court, which will stop the development of legal LLMs.

5.3. Challenges of traditional legal industry

The rapid development of the digital age and AI has led to a shift towards proactive justice, creating a challenge for the traditional legal industry known as "activist justice". The specific challenges are as follows:

Neglecting Judicial Independence: Currently, legal LLMs are not sufficient to replace judges in making decisions (Završnik, 2020). Judicial independence encompasses aspects of legal enforcement and fact-finding. In terms of legal enforcement, independence includes interpreting civil law, explaining uncertain concepts, and evaluating disputes over the rights and interests of parties involved in a case. In terms of fact-finding, independence includes the use of discretion, subjective judgment, experiential judgment, and weighing pros and cons in decision-making. For example, judges need to make judgments based on the evidence and testimony presented by both parties during litigation. In these aspects of judicial independence, legal LLMs often play an auxiliary role. But if legal LLMs get too involved in case decisions, judges might rely too much on AI to find relevant literature, establish facts, and even form preconceived notions before the proceedings, ignoring what the people involved in the case want. This weakens the exercise of judicial discretion. The decision-making solutions generated by legal LLMs based on case characteristics may deprive judges of their discretionary powers in the details of a case (Fagan and Levmore, 2019). Judges have the freedom to exercise discretion in aspects such as assessing the credibility of evidence based on the law's applicability. They can also make reasonable decisions by considering factors such as the extent of the victim's property loss and the defendant's compensation ability, based on judicial experience or the perspectives of the parties involved. For example, in assessing the compensation amount in civil litigation, judges can comprehensively consider factors such as the extent of the victim's financial loss and the defendant's ability to compensate. In contrast, the algorithms of legal LLMs struggle to measure the extent of loss and evaluate a person's ability to pay, so decisions based solely on case characteristics weaken the judge's discretionary powers to a certain extent.

The position of legal LLMs is unclear. The unclear positioning of legal LLMs undermines the role and judicial power of judges in cases. Legal LLMs should assist judges in decision-making (Rissland et al., 2003), provide advice, automatically generate legal texts, etc., but they do not possess professional judicial experience and cannot independently make judgments in cases. Therefore, users should fully understand the position of legal LLMs within the legal system when utilizing them.

Impact on the Traditional Court System: The concept of trial centrism emphasizes the principle of equality between prosecution and defense, with the judge serving as the main authority and the trial playing a decisive role. With the development of AI technologies such as AIGC, legal LLMs have alleviated the problem of manpower shortages in the legal field, but they have also restrained judges' subjective initiative and the development of traditional trial systems. This is mainly reflected in the following aspects:

(i) Court idleness: Trial centrism emphasizes the central role of judges in the judicial process (Re and Solow-Niederman, 2019), and a fair trial allows equal confrontation between the parties involved, enabling them to have confidence in the judicial process and reach a just judgment. However, the popularization of legal AI systems may lead to court idleness, which would diminish the solemnity of the legal process and the subtle influence it brings. The idleness or weakness of court trials will restrict judges' subjective initiative in the judicial process and reduce public trust in the judiciary.

(ii) Crisis in the hierarchy of trials: The hierarchical system establishes a relationship between higher and lower courts, ensuring people's litigation rights and achieving fair and prudent judicial outcomes by setting different levels of adjudication. Legal AI systems and other AI technologies may impact the judicial process in the hierarchical system (Contini, 2020). For example, if a party is dissatisfied with a lower court's judgment and appeals to a higher court for a second trial, if most courts use the same legal AI system, the second trial will be no different from the first trial. This goes against the purpose of safeguarding litigation rights, higher courts' supervision of lower courts, and achieving judicial fairness through the hierarchical system. Therefore, the application of legal LLMs in the judicial field may have certain impacts on the trial system.

5.4. Issues arising from specific judicial practice

The Lack of Universality in Applications. Legal LLMs, when assisting in decision-making (Zhong et al., 2020b), often extract feature values from cases and search for similar cases within existing multidimensional datasets to find the "optimal solution". However, because of the differences between cases, the "optimal solution" proposed by the LLMs may not be applicable to a specific case. Furthermore, legal regulations may vary across countries or regions, leading to inconsistent decision outcomes for the same case under different legal rules. Legal LLMs struggle to address the issue of case diversity and cannot be applied to all legal cases.

The Lack of Subjective Thinking, Emotions, and Experience. Legal LLMs, compared to legal professionals, lack autonomous thinking abilities and professional experience, among other things. Legal LLMs can process cases through factor identification, but the judicial experience is difficult to quantify accurately (e.g., the standard of proof for "reasonable doubt" in criminal proceedings), and AI struggles to subjectively assess the truthfulness of case statements. Furthermore, legal AI systems lack the ability to blend law and empathy, which undoubtedly results in a lack of human touch in legal regulations and undermines people's trust in the judiciary. The judicial decision-making process is not merely a logical reasoning process on a single layer but also involves moral, ethical, and practical considerations in the legal system (Xu, 2022). Given these limitations, the application of legal LLMs in the field of justice is still lacking.

Contradiction with the Presumption of Innocence Principle. In the past few years, AI systems have been used to help stop crime. Some examples are the COMPAS system for crime prediction and risk assessment (Beriain, 2018), PredPol for iterative calculation of potential crime locations through the analysis of criminal history data (Rosser et al., 2017), and the PRECOBS system in Germany for burglary prevention and violence crime prediction (Egbert and Krasmann, 2020). This has led to a shift in policing from being 'service-oriented' and 'reactive' to 'proactive prediction' (Hardyns and Rummens, 2018). However, this shift contradicts the principle of presumption of innocence. Judicial decisions should be based on known cases, and predictive modeling algorithms based on personal privacy data may deprive individuals identified as "future potential criminals" of basic public services such as education or social welfare, leading to discrimination and restrictions. However, these future-oriented "judgments" fundamentally undermine human rights. Moreover, as predictive policing becomes more prevalent and the number of registered "tagged offenders" increases, a particular area may be perceived as a "high crime area" due to these model-driven decisions, inadvertently promoting regional discrimination and bias. This contradicts both human rights and the presumption of innocence principle.

(i) The imbalance of prosecution and defense. The application of AI technologies in the judiciary, such as legal LLMs, may lead to an imbalance of public and private powers. The application of AI in the judiciary will result in the problem of "triangular imbalance", causing an imbalance in the relationship between prosecution and defense, and the excessive exercise of public power is accompanied by a decrease in public trust.

(ii) Unequal control over data. Controlling over the big data used by LLMs is in the hands of public authorities or large corporations, making it difficult for individuals to access the operational mode of big data (Prescott, 2017; Zuiderveen Borgesius, 2020). This is likely to lead to the expansion of public power and the contraction of private rights. For example, during the investigation process, public authorities can access case data and collect criminal evidence about suspects through methods such as software or mobile phone tracking. However, the defense side, due to its power limitations, is in a disadvantaged position when collecting more relevant data. The unequal control over data severely restricts the trial's fairness.

(iii) Differences in the ability to analyze case data. Some research (Wexler, 2021) believes that even if both the prosecution and defense have the same access rights to obtain data, there is still a significant gap in their ability to analyze case data (Wexler, 2021). The prosecution can extract and analyze legal case data based on national regulations, utilizing national resources such as a large number of professional data analysts and advanced data analysis equipment. Most defense lawyers lack this kind of professional data analysis capability, and a large amount of data analysis increases litigation costs for citizens.

(iv) Inconsistent attention to legal LLMs. The application of legal LLMs reflects issues of policy attention, investment imbalance, and unequal exploration results between public power departments such as public security agencies, judicial organs, and individuals. This will reduce the public's trust in the judiciary.

(v) Administrative intervention leads to the abuse of legal LLMs. Individual performance assessment is an important evaluation criterion for judges. With the gradual application of legal LLMs, some courts incorporate indicators of AI-assisted decision-making into the evaluation criteria for judges. To some extent, this has resulted in the abuse of AI in the judiciary and has also undermined judges' status in the trial process.

Privacy Infringement. In practical applications, legal LLMs are prone to privacy infringement issues (Rodrigues, 2020; Završnik, 2020). Firstly, legal LLMs may collect excessive user data, which can lead to privacy breaches when the input information is sensitive. Secondly, the underlying algorithms of legal LLMs may be inadequate, resulting in improper data handling. For example, if user inputs contain sensitive labels (such as birthdates or home addresses) in the legal dataset used for deep learning and the model lacks proper anonymization procedures, user privacy could be amplified, leading to privacy infringement issues.

Issues of Intellectual Property Identification and Protection. With the continuous development of AI technology, the identification and protection of intellectual property rights in the legal field have attracted attention (Rodrigues, 2020). Because AI is involved in creative processes, it is difficult to determine intellectual property ownership. It becomes unclear whether credit should be given to the human users who utilize AI or to the AI system itself. Since AI lacks independent thinking capabilities and its creations are derived from deep learning on existing works within the field, it may not meet the requirement of originality. Therefore, it can be inferred that legal LLMs and other AI technologies may amplify the deficiencies in the identification and protection of intellectual property rights in the judicial domain.

5.5. Ethical views impacting human society

Disregard for Human Subjectivity. During the training process of LLMs on datasets, to ensure the values of the LLMs themselves, negative comments in the dataset need to be labeled and filtered. However, due to the low cost of labor, some workers who are hired at very low prices may experience psychological problems when screening a large number of negative comments, and human subjectivity is susceptible to algorithmic bullying. Furthermore, the human labor contribution to this process is ignored. Legal LLMs are challenging the ethical views of human society (Raso et al., 2018).

Misleading User Comments. AI has displayed behaviors such as inducing users to divorce, making inappropriate comments, and even encouraging users to disclose personal privacy or engage in illegal activities (Riveiro and Thill, 2021). The reason is that the AI training dataset contains incorrect or discriminatory comments, and through deep learning, AI acquires these characteristics. This undoubtedly has an impact on the ethics of human society.

Ethical Value Consistency. The development of LLMs should first and foremost adhere to the ethical values of all humanity. If AI's values do not align with those of humans, there may be situations where AI misleads or harms human interests, posing challenges to national governance systems and global cooperation.

6. Future directions

The rapid development of legal LLMs is continuously changing the landscape and practices in the legal field. With the ongoing advancements in AI and natural language processing technologies, we are paying attention to more issues related to the applications of legal LLMs, and we recognize that they hold broader prospects for development. We propose several potential directions for the future development of legal LLMs. Details are discussed below.

6.1. Data and infrastructure

Obtaining More Comprehensive Legal Big Data: Firstly, we should broaden the scope of obtaining legal big data (Zhong et al., 2020c). For example, we can collect data not only from specific court databases but also enhance the protection techniques for legal data, such as encrypting data transmission and increasing access control, in order to promote greater openness of data by various courts. Secondly, to address the issue of insufficient sharing permissions, we can improve legal regulations and clearly define the sharing permissions between different levels of courts, stipulating their rights and obligations, thus facilitating more sharing of legal data. Leveraging cloud computing to enhance new sharing technologies and establish an open and secure big data sharing platform. Lastly, courts can standardize legal document formats and conduct preprocessing operations, such as evaluating and cleaning large amounts of legal data using unified filtering criteria. Moreover, optimizing pre-training models (e.g., Lawformer) for legal documents can enhance training on long legal documents (Xiao et al., 2021).

Defining the Boundaries of Legal Concepts and Limiting the Scope of Application: To address the challenge of legal LLMs struggling to convert certain ambiguous legal concepts into programming symbols, we have summarized some solutions. Firstly, it involves defining the boundaries of legal concepts and restricting fuzzy legal concepts based on criteria like societal impact (Rissland et al., 2003). Secondly, it entails limiting the applicability of legal LLMs in legal cases and reducing their decision-making proportion in cases involving relevant legal concepts while upholding the judicial authority of judges.

Data Transparency: Whether it is ChatLaw or Legal BERT, both legal LLMs require a series of operations, such as collecting, screening, extracting, classifying, and training legal datasets. These processes may introduce bias or adopt unreasonable and incomplete methods. So, we need to set standards for datasets and AI mechanisms and make sure they are properly shared. These standards should cover things like dataset content, usage limits, licenses, data collection methods, information on data quality and uncertainty, as well as data features, structure, and classification schemes (Zhang et al., 2023). Openly sharing data promotes more comprehensive and accurate datasets for large-scale model training, while also enhancing public trust in the judiciary.

Building a Legal Knowledge Graph: This enables the connection of legal concepts, legal cases, and applicable legal rules, establishing a model that is easily understandable by AI. Firstly, legal concepts such as "criminal law" and "intellectual property rights" are classified. Then, logical relationships are established for applicable legal rules, such as "criteria for determining illegality". Specific legal cases, such as "robbery cases", are classified. Lastly, by using legal documents as references in the legal knowledge base, an in-depth legal knowledge graph is created (Yang et al., 2021), which can optimize the legal model's functionality.

Optimizing the Foundational Infrastructure for Model Training: In general, this involves various opportunities, including but not limited to:

(i) High-performance computing resources: Considering the wide variety and different kinds of legal datasets, it is suggested to utilize high-performance computing resources such as Graphics Processing Units (GPUs) (Brodtkorb et al., 2013) and Tensor Processing Units (TPUs) (Jouppi et al., 2017) to significantly improve the training and inference speed of legal LLMs (Scao et al., 2022).

(ii) Distributed computing frameworks: Single-model training is no longer sufficient to meet specific requirements. Consider using distributed computing frameworks, such as TensorFlow (Abadi et al., 2016) and PyTorch's distributed training capabilities (Paszke et al., 2019), to achieve parallel processing and accelerate model training (Li et al., 2020). By distributing computing tasks across multiple nodes and devices, the training process can be completed more quickly.

(iii) Storage and data management: For handling datasets of legal LLMs, appropriate storage and data management solutions should be considered. For example, high-performance storage systems like SSDs (Solid State Drives) can provide fast data read and write speeds (Chen et al., 2011). Moreover, distributed file systems or object storage systems can be used to effectively manage large-scale legal datasets.

(iv) Data preprocessing and cleaning: Proper preprocessing and cleaning of data are necessary before using legal LLMs. This includes removing noise and redundant information, standardizing text formats, and ensuring the quality and consistency of the dataset. Utilizing professional data cleaning tools and techniques can help improve the performance and accuracy of legal LLMs.

(v) Model scaling and deployment: When deploying legal LLMs into practical applications, considerations must be given to model scaling and deployment issues. Model compression and pruning techniques can reduce the model's size and computational requirements, resulting in improved deployment efficiency on edge devices. Moreover, containerization technologies such as Docker and Kubernetes (Bernstein, 2014) can be used to simplify model deployment and management (Kozhirbayev and Sinnott, 2017).

6.2. Algorithm level

Strategy Adjustment and Optimized Algorithms: To address the challenges of modeling long texts, we can employ external recall methods, model optimization (Hoffmann et al., 2022), and optimization of the attention mechanism. These approaches can help us tackle the difficulties of long text modeling and improve the capability and effectiveness of legal LLMs in handling long texts. External recall methods involve utilizing external tools or external memory to assist in processing long texts. In the legal domain, a legal knowledge base can be established, containing legal documents, precedents, regulations, and other information. The model can query this knowledge base to obtain relevant legal background and explanations, enabling a better understanding and processing of long texts. Model optimization involves improving the modeling capability for long texts by optimizing the model itself. Specialized legal models can be designed and trained to address specific tasks and requirements in the legal domain. These models can be trained using legal data during the pre-training phase to enhance their understanding and processing capabilities for legal texts. The attention mechanism, which is a critical component of large models, can be optimized to improve the effectiveness of long-text modeling (Niu et al., 2021). More complex and refined Attention mechanisms can be designed to enable the model to better focus on key information and context within the text.

Limiting Algorithmic Biases and "Black Box" Operations: When elements of inequality, such as racial or gender biases, are incorporated into the algorithmic decision-making of LLMs, it can lead to different legal judgments in criminal law, resulting in unjust judicial outcomes. Moreover, the opacity of certain legal LLMs makes it difficult to prove the absence of "black box" operations (Pedreschi et al., 2019). Making algorithms public and subjecting them to evaluation can enhance the transparency and openness of legal LLMs, thereby facilitating their use in assisting judicial decision-making (Liang et al., 2021), including but not limited to:

(i) Implement protective algorithms: In the process of handling judicial data at the underlying algorithmic level of LLMs, algorithms can be introduced to protect against biased elements. For example, random noise can be added to the original data, and algorithms based on differential privacy can be applied to protect the data (Wei et al., 2020).

(ii) Implement algorithmic explainability: With the application of explainable artificial intelligence (XAI) (Das and Rad, 2020; Došilović et al., 2018), we need to consider how to better integrate XAI and legal requirements and how to further enhance the ability of XAI algorithms to identify biases (Deeks, 2019). By increasing the explainability of AI decision-making algorithms, we can promote the transparency of LLM judgments and restrict the occurrence of unfair operations.

(iii) Incorporate legal indicators for evaluation: Solaiman (Solaiman and Dennison, 2021) proposes a Process for Adapting Language Models to Society (PALMS) with Values-Targeted Datasets (Solaiman and Dennison, 2021). The method is designed to ensure that the models are fair and do not discriminate against certain groups. The evaluations are based on quantitative metrics such as similarity between output and target results, toxicity scoring, and qualitative metrics that analyze the most common words associated with a given social category. By using legal metrics to evaluate the decision-making results of the LLMs, the method can effectively address potential bias issues. Additionally, some countries have enacted laws to prevent algorithmic bias, such as the European Union's laws protecting the rights of individuals against algorithmic discrimination, including data protection laws and non-discrimination laws (Zuiderveen Borgesius, 2020).

Promote Limited Algorithmic Transparency: Making decision algorithms public allows people to gain a better understanding of the underlying principles, whether it is the weighing of judgments in common civil litigation or the criminal discretion of public authorities. Even small algorithmic changes, or the definition of a biased element, can introduce difficult-to-assess risks. Decision-making bodies should focus on promoting the transparency of decision algorithms, allowing legal LLMs' algorithms to be supervised by all parties involved, promoting judicial transparency, and achieving fairer judgments. For example, in Wisconsin v. Loomis (Beriain, 2018), the refusal to disclose the judgment results based on trade secrets undermines public trust. However, algorithmic transparency should be subject to certain constraints. For instance, in cases involving potential trade secrets or national security information, flexibility should be allowed. Limited algorithmic transparency can be achieved by signing confidentiality agreements with parties involved or holding closed pretrial hearings. In addition, algorithmic transparency enables ordinary citizens with legal consulting needs to input keywords that better match the algorithmic procedures, thereby obtaining legal advice that is more tailored to the characteristics of their cases.

6.3. Dealing with traditional judiciary

Clarifying the Positioning of Large Models: legal LLMs often intervene in judicial decision-making and even influence the discretionary power of judges. This is mainly due to the unclear positioning of legal models within the legal system (Re and Solow-Niederman, 2019). To address this issue, we should uphold the independence and autonomy of judges and improve relevant laws and regulations to clearly define the role, functions, and scope of legal AI in the judiciary. We should use large models to assist judges in decision-making, such as providing precise references to legal documents, rather than relying solely on LLMs. Moreover, with the assistance of AI, judges should strive to become more empathetic, rational, and professional decision-makers.

Defining the Thinking Capability of LLMs: Legal LLMs process a vast amount of legal datasets through algorithms like deep learning to simulate judges' information extraction and decision-making in cases. LLMs do not, however, have independent thinking abilities due to computational limitations. As a result, we need to define large models' thinking capability, and here are some evaluation indicators. Firstly, we can assess the predictive accuracy of legal LLMs by comparing the predictions for a certain type of case with the outcomes of traditional judicial systems (McKay, 2020). Secondly, we can evaluate the extent of resource consumption. For example, we can compare and evaluate the time, resource, and manpower consumption of processing a given case using a legal LLM versus traditional judicial methods. Furthermore, the legal LLMs should provide explanations regarding the connection between the case and the judgment results, and these explanations can be compared with those provided by judges to determine the degree of alignment between AI and human judicial thinking.

Ensuring Parties' Access to Data: Parties involved in judicial cases are the main subjects of the cases and the primary sources of data used by AI. They should have the right to access, question, and update their data (Zuiderveen Borgesius, 2020). Several studies argue that "data access rights" are essential in a technology-based judiciary (Alzain and Pardede, 2011; Prescott, 2017). In the process of AI-assisted judicial decision-making, the parties' legal access rights should be protected. For example, they should have the right to question the digitized evidence of their case and the right to update outdated data used by AI. In addition, the defense should have the right to contact technical personnel related to AI, ensuring their ability to review and question the data used by legal LLMs.

Expanding and Optimizing the Consulting Function of Judicial LLMs: AI technologies such as ChatGPT have interactive question-and-answer capabilities. In the field of judicial judgments, expanding AI consulting capabilities may provide more opportunities for individuals who are not familiar with the law to receive legal advice. Optimizing the LLM's legal consulting capabilities involves specific applications and criminal law constraints, for example:

(i) Fine-tuning the LLMs and introducing multimodal capabilities. The model can be fine-tuned for different subfields of law. For example, to optimize the consulting capabilities in the field of labor law, the model can be fine-tuned using labor law-related cases and materials to enhance its performance in that subfield. Additionally, to better understand user needs, multimodal inputs can be introduced (Joshi et al., 2021), such as allowing users to provide legal questions through voice or images. Similarly, the model can provide multimodal outputs based on user needs and scenarios, such as speech, text, charts, etc.

(ii) Involvement of legal experts and practitioners. Inviting legal experts, judges, prosecutors, and lawyers to participate in the development and evaluation process of judicial AI. They can provide valuable legal knowledge and practical experience to help guide the design and training of AI systems, ensuring compliance with legal regulations and judicial practices. Furthermore, legal experts can evaluate and refine the results of the LLMs (Wachter et al., 2021), fine-tuning them to improve the accuracy and reliability of the model in legal consulting.

(iii) User feedback and continuous optimization: In the training process of the legal LLMs, the Reinforcement Learning from Human Feedback (RLHF) machine learning method can be adopted (Liu et al., 2023). First, provide users with a feedback mechanism (Ouyang et al., 2022), encourage them to actively participate in the use of legal LLMs, and collect user feedback. According to the annotated data containing user feedback, train a reward model (RM) to rank the content generated by the large language model. Finally, update the LLM parameters through deep learning, making the legal LLMs more in line with the actual needs of users. This method reduces the need for extensive trial and error in traditional reinforcement learning, and it optimizes and adjusts the model based on user needs.

6.4. Judicial practice

Improve Accountability Mechanisms to Prevent Political Interference: Establishing a sound accountability mechanism is beneficial for regulating the use of AI in the judiciary and preventing legal errors caused by legal LLMs (Contini, 2020). For example, the responsibility for judicial errors would discourage judges from completely delegating decision-making to legal LLMs, instead prompting them to carefully consider the decision recommendations provided by legal LLMs. Furthermore, we should guard against political interference in the utilization of AI in the judiciary. For instance, to promote the application of AI in the judicial field, courts at all levels may receive performance assessment indicators related to the use of AI technologies. In order to meet these indicators, courts may inadvertently overlook AI's limitations, resulting in judicial errors.

Foster the Development of Interdisciplinary Talents: One major reason for issues such as "algorithmic bias" and "algorithmic black boxes" in legal LLMs is the lack of interdisciplinary talents combining legal and technological expertise. We should encourage the cultivation of talents that possess both legal and technological knowledge (Wei and Fengru, 2021). This would enable us to address the problem of legal concepts that are difficult to translate into algorithmic programs. Moreover, interdisciplinary talents with legal experience would be able to identify discriminatory elements in large datasets and eliminate them. This also ensures the security of algorithms in legal LLMs and greatly reduces the probability of "algorithmic black boxes" occurring.

Collaboration and Sharing of Experiences: AI technology has developed rapidly internationally. The research on reasoning capabilities can be traced back to the Mycin system in the 1970s, followed by the establishment of IBM's Watson model, the study of LLMs, and the release of Open AI's ChatGPT. There is rich experience in the field of AI. It is important to collaborate with foreign research institutions, companies, and experts to share experiences and technological achievements. International cooperation can facilitate exchanges and collaborations among different countries in the field of AI in the judiciary, enabling them to address technical and ethical issues jointly. It can also avoid duplication of effort and errors.

7. Conclusions

This paper is dedicated to synthesizing various technologies and ideas regarding the opportunities, challenges, and recommendations for the application of AI in the judicial field. We hope that this paper can give some potential inspiration or research directions to those who are researching legal LLMs or legal practitioners. This paper reviews the opportunities and challenges of legal LLMs in the judicial field. With the rapid development of AI technology, LLMs based on AIGC and other AI technologies have attracted widespread attention in the legal field. We first provide an overview of the development and related research of legal LLMs, as well as explore their opportunities in judge-assisted trials and legal consultation. Next, we discuss the shortcomings of legal LLMs at the algorithmic level and in specific practice. Legal models have huge potential in the judicial field. They can provide efficient legal advice, assist judges in making decisions, speed up the processing of judicial cases, reduce the workload of judges, and improve the accuracy and consistency of judicial decision-making.

When analyzing the current status of research on legal LLMs, as mentioned in Sections 5 and 4.3, firstly, they can equip users with efficient knowledge of question-and-answer functions and legal document generation. For example, ChatLaw, proposed by Peking University in 2023, can facilitate user information interaction and match the most similar cases for knowledge questions and answers based on the user's legal consultation. However, legal LLMs also face problems, such as insufficient processing capabilities for long texts or data set processing defects. In addition, legal LLMs are used to assist judges in logical reasoning and decision-making, but they also face problems such as user privacy, legal ethics, and the "decision-making black box". For example, COMPAS (Beriain, 2018) is widely used in judicial sentencing in the United States. It attempts to predict the probability of a defendant being rearrested based on existing criminal records. Black defendants are more likely to receive higher COMPAS scores than white defendants, which leads to bias. More importantly, legal LLMs can make legal judgment predictions based on similar facts of existing legal cases, but they also face problems such as limited understanding and adaptability to individual cases and unexplainable judgment predictions. Although there are evaluation systems for fine-tuning legal LLMs, there are also challenges, such as unreasonable and incomplete evaluation index settings and difficulty in the numerical calculation of some evaluation indicators.

To give full play to the advantages of legal LLMs and meet the challenges of the current research status, we put forward the following suggestions and future development directions: Pay attention to the data quality of legal LLMs and the privacy protection of users; improve the long text processing ability of legal LLMs; improve the adaptability and intelligence level of the model; establish a reasonable ethical framework and regulatory mechanism to promote the interpretability of legal LLMs' decisions; establish a complete and systematic evaluation index system for legal LLMs; explore the application of legal models in other judicial fields; strengthen international cooperation and knowledge sharing; and establish a multi-party cooperation model to achieve the sustainable development and social benefits of legal LLMs.

CRediT authorship contribution statement

Wensheng Gan: Writing – original draft. **Jiayang Wu:** Investigation, Resources. **Zhenlian Qi:** Investigation, Writing – original draft. **Philip S. Yu:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported in part by Guangzhou Basic and Applied Basic Research Foundation (No. 2024A04J9971), National Natural Science Foundation of China (No. 62272196), Natural Science Foundation of Guangdong Province (No. 2022A1515011861), Engineering Research Center of Trustworthy AI, Ministry of Education (Jinan University), and Guangdong Key Laboratory of Data Security and Privacy Preserving.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al., 2016. TensorFlow: a system for large-scale machine learning. In: The 12th USENIX Symposium on Operating Systems Design and Implementation. pp. 265–283.

Adnan, K., Akbar, R., 2019. An analytical study of information extraction from unstructured and multidimensional big data. J. Big Data 6 (1), 1–38.

Aini, G., et al., 2020. A summary of the research on the judicial application of artificial intelligence. Chin. Stud. 9 (01), 14.

Alavian, P., Eun, Y., Liu, K., Meerkov, S.M., Zhang, L., 2020. The (α, β)-precise estimates of MTBF and MTTR: Definition, calculation, and observation time. IEEE Trans. Autom. Sci. Eng. 18 (3), 1469–1477.

Alzain, M.A., Pardede, E., 2011. Using multi shares for ensuring privacy in database-as-a-service. In: The 44th Hawaii International Conference on System Sciences. IEEE, pp. 1–9.

Ashley, K.D., 1991. Reasoning with cases and hypotheticals in HYPO. Int. J. Man-Mach. Stud. 34 (6), 753–796.

Ashley, K.D., 1992. Case-based reasoning and its implications for legal expert systems. Artif. Intell. Law 1 (2), 113–208.

Atkinson, K., Bench-Capon, T., Bollegala, D., 2020. Explanation in AI and law: Past, present and future. Artificial Intelligence 289, 103387.

Aydin, Ö., Karaarslan, E., 2023. Is ChatGPT leading generative AI? What is beyond expectations? Acad. Platform J. Eng. Smart Syst. 11 (3), 118–134.

Baude, W., 2020. Precedent and discretion. Supreme Court Rev. 2019 (1), 313–334.

Bengio, Y., Ducharme, R., Vincent, P., Jauvin, C., 2003. A neural probabilistic language model. J. Mach. Learn. Res. 3, 1137–1155.

Beriain, I.D.M., 2018. Does the use of risk assessments in sentences respect the right to due process? A critical analysis of the Wisconsin v. Loomis ruling. Law, Probab. Risk 17 (1), 45–53.

Bernstein, D., 2014. Containers and cloud: From LXC to docker to kubernetes. IEEE Cloud Comput. 1 (3), 81–84.

Bhaskar, A., Fabbri, A., Durrett, G., 2023. Prompted opinion summarization with GPT-3.5. In: Findings of the Association for Computational Linguistics. pp. 9282–9300.

Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M.S., Bohg, J., Bosselut, A., Brunskill, E., et al., 2021. On the opportunities and risks of foundation models. arXiv preprint, arXiv:2108.07258.

Brants, T., Popat, A.C., Xu, P., Och, F.J., Dean, J., 2007. Large language models in machine translation. EMNLP-CoNLL 858.

Brennan, T., Dieterich, W., Ehret, B., 2009. Evaluating the predictive validity of the COMPAS risk and needs assessment system. Crim. Justice Behav. 36 (1), 21–40.

Brodtkorb, A.R., Hagen, T.R., Sætra, M.L., 2013. Graphics processing unit (GPU) programming strategies and trends in GPU computing. J. Parallel Distrib. Comput. 73 (1), 4–13.

Brown, P.F., Della Pietra, V.J., Desouza, P.V., Lai, J.C., Mercer, R.L., 1992. Class-based n-gram models of natural language. Comput. Linguist. 18 (4), 467–480.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al., 2020. Language models are few-shot learners. Adv. Neural Inf. Process. Syst. 33, 1877–1901.

Brynjolfsson, E., Mcafee, A., 2017. Artificial intelligence, for real. Harv. Bus. Rev. 1, 1–31.

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P.S., Sun, L., 2023. A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT. arXiv preprint, arXiv:2303.04226.

Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., Roberts, A., Brown, T., Song, D., Erlingsson, U., et al., 2021. Extracting training data from large language models. In: 30th USENIX Security Symposium. pp. 2633–2650.

Chalkidis, I., Kampas, D., 2019. Deep learning in law: early adaptation and legal word embeddings trained on large corpora. Artif. Intell. Law 27 (2), 171–198.

Chen, F., Lee, R., Zhang, X., 2011. Essential roles of exploiting internal parallelism of flash memory based solid state drives in high-speed data processing. In: IEEE 17th International Symposium on High Performance Computer Architecture. IEEE, pp. 266–277.

Chen, B., Li, Y., Zhang, S., Lian, H., He, T., 2019. A deep learning method for judicial decision support. In: IEEE 19th International Conference on Software Quality, Reliability and Security Companion. IEEE, pp. 145–149.

Chowdhary, K., Chowdhary, K., 2020. Natural language processing. Fundam. Artif. Intell. 603–649.

Church, K.W., 2017. Word2Vec. Nat. Lang. Eng. 23 (1), 155–162.

Contini, F., 2020. Artificial intelligence and the transformation of humans, law and technology interactions in judicial proceedings. Law, Technol. Hum. 2 (1), 4–18.

Croitoru, F.A., Hondru, V., Ionescu, R.T., Shah, M., 2023. Diffusion models in vision: A survey. IEEE Trans. Pattern Anal. Mach. Intell. (01), 1–20.

Cui, Y., 2020. Artificial Intelligence and Judicial Modernization. Springer.

Cui, J., Li, Z., Yan, Y., Chen, B., Yuan, L., 2023. ChatLaw: Open-source legal large language model with integrated external knowledge bases. arXiv preprint, arXiv:2306.16092.

Das, A., Rad, P., 2020. Opportunities and challenges in explainable artificial intelligence (XAI): A survey. arXiv preprint, arXiv:2006.11371.

De Capitani Di Vimercati, S., Foresti, S., Livraga, G., Samarati, P., 2012. Data privacy: Definitions and techniques. Internat. J. Uncertain. Fuzziness Knowledge-Based Systems 20 (06), 793–817.

de Sousa, W.G., Fidelis, R.A., de Souza Bermejo, P.H., da Silva Gonçalo, A.G., de Souza Melo, B., 2022. Artificial intelligence and speedy trial in the judiciary: Myth, reality or need? A case study in the Brazilian supreme court (STF). Gov. Inf. Q. 39 (1), 101660.

Deeks, A., 2019. The judicial demand for explainable artificial intelligence. Columbia Law Rev. 119 (7), 1829–1850.

Densley, J.A., Pyrooz, D.C., 2020. The matrix in context: Taking stock of police gang databases in London and beyond. Youth Justice 20 (1–2), 11–30.

Devlin, J., Chang, M.W.C., Lee, K., Toutanova, K., 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In: NAACL-HLT. pp. 4171–4186.

Došilović, F.K., Brčić, M., Hlupić, N., 2018. Explainable artificial intelligence: A survey. In: The 41st International Convention on Information and Communication Technology, Electronics and Microelectronics. IEEE, pp. 0210–0215.

Du, M., He, F., Zou, N., Tao, D., Hu, X., 2022. Shortcut learning of large language models in natural language understanding: A survey. arXiv preprint, arXiv:2208.11857.

Easterbrook, F.H., 1987. Stability and reliability in judicial decisions. Cornell Law Rev. 73, 422.

Egbert, S., Krasmann, S., 2020. Predictive policing: not yet, but soon preemptive? Policing Soc. 30 (8), 905–919.

Etulle, R.D., Moslares, F.R., Pacad, E.Z.J., Odullo, J.L., Nacionales, J., Claridad, N., 2023. Investigating the listening and transcription performance in court: experiences from stenographers in Philippine courtrooms. J. Lang. Pragmatics Stud. 2 (2), 100–111.

Fagan, F., Levmore, S., 2019. The impact of artificial intelligence on rules, standards, and judicial discretion. South. Calif. Law Rev. 93, 1.

Fedus, W., Zoph, B., Shazeer, N., 2022. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. J. Mach. Learn. Res. 23 (1), 5232–5270.

Fetzer, J.H., Fetzer, J.H., 1990. Minds, bodies, and machines. Artif. Intell.: Scope Limits 269–303.

Flores, A.W., Bechtel, K., Lowenkamp, C.T., 2016. False positives, false negatives, and false analyses: A rejoinder to machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. Fed. Probat. 80, 38.

Gan, W., Qi, Z., Wu, J., Lin, J.C.-W., 2023a. Large language models in education: Vision and opportunities. In: IEEE International Conference on Big Data. IEEE, pp. 4776–4785.

Gan, W., Wan, S., Yu, P.S., 2023b. Model-as-a-service (MaaS): A survey. In: IEEE International Conference on Big Data. IEEE, pp. 1–10.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Greenleaf, G., Mowbray, A., Chung, P., 2018. Building sustainable free legal advisory systems: Experiences from the history of AI & law. Comput. Law Secur. Rev. 34 (2), 314–326.

Greenstein, S., 2022. Preserving the rule of law in the era of artificial intelligence (AI). Artif. Intell. Law 30 (3), 291–323.

Hage, J.C., von der Pfordten, D., 2009. Concepts in Law, vol. 88, Springer Science & Business Media.

Hamet, P., Tremblay, J., 2017. Artificial intelligence in medicine. Metabolism 69, S36–S40.

Han, K., Xiao, A., Wu, E., Guo, J., Xu, C., Wang, Y., 2021. Transformer in transformer. Adv. Neural Inf. Process. Syst. 34, 15908–15919.

Hanretty, C., 2020. A Court of Specialists: Judicial Behavior on the UK Supreme Court. Oxford University Press, USA.

Hardyns, W., Rummens, A., 2018. Predictive policing as a new tool for law enforcement? Recent developments and challenges. Eur. J. Crim. Policy Res. 24, 201–218.

Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D.d.L., Hendricks, L.A., Welbl, J., Clark, A., et al., 2022. Training compute-optimal large language models. arXiv preprint, arXiv:2203.15556.

Hoy, M.B., 2018. Alexa, Siri, Cortana, and more: an introduction to voice assistants. Med. Ref. Serv. Q. 37 (1), 81–88.

Huang, Q., Tao, M., An, Z., Zhang, C., Jiang, C., Chen, Z., Wu, Z., Feng, Y., 2023. Lawyer LLaMA technical report. arXiv preprint, arXiv:2305.15062.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe: Convolutional architecture for fast feature embedding. In: The 22nd ACM International Conference on Multimedia. pp. 675–678.

Joshi, G., Walambe, R., Kotecha, K., 2021. A review on explainability in multimodal deep neural nets. IEEE Access 9, 59800–59821.

Jouppi, N.P., Young, C., Patil, N., Patterson, D., Agrawal, G., Bajwa, R., Bates, S., Bhatia, S., Boden, N., Borchers, A., et al., 2017. In-datacenter performance analysis of a tensor processing unit. In: The 44th Annual International Symposium on Computer Architecture. pp. 1–12.

Kallus, N., 2020. Generalized optimal matching methods for causal inference. J. Mach. Learn. Res. 21 (1), 2300–2353.

Kanapala, A., Pal, S., Pamula, R., 2019. Text summarization from legal documents: a survey. Artif. Intell. Rev. 51, 371–402.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., et al., 2023. ChatGPT for good? On opportunities and challenges of large language models for education. Learn. Individ. Differ. 103, 102274.

Kepuska, V., Bohouta, G., 2018. Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home). In: IEEE 8th Annual Computing and Communication Workshop and Conference. IEEE, pp. 99–103.

Kozhirbayev, Z., Sinnott, R.O., 2017. A performance comparison of container-based technologies for the cloud. Future Gener. Comput. Syst. 68, 175–182.

La Fors, K., 2020. Legal remedies for a forgiving society: Children's rights, data protection rights and the value of forgiveness in AI-mediated risk profiling of children by Dutch authorities. Comput. Law Secur. Rev. 38, 105430.

Laghari, A.A., Wu, K., Laghari, R.A., Ali, M., Khan, A.A., 2021. A review and state of art of internet of things (IoT). Arch. Comput. Methods Eng. 1–19.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Li, S., Zhao, Y., Varma, R., Salpekar, O., Noordhuis, P., Li, T., Paszke, A., Smith, J., Vaughan, B., Damania, P., et al., 2020. PyTorch distributed: Experiences on accelerating data parallel training. VLDB Endow. 13 (12), 3005–3018.

Liang, P.P., Wu, C., Morency, L.-P., Salakhutdinov, R., 2021. Towards understanding and mitigating social biases in language models. In: International Conference on Machine Learning. PMLR, pp. 6565–6576.

Liu, B., Ding, M., Shaham, S., Rahayu, W., Farokhi, F., Lin, Z., 2021a. When machine learning meets privacy: A survey and outlook. ACM Comput. Surv. 54 (2), 1–36.

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., et al., 2023. Summary of ChatGPT-Related research and perspective towards the future of large language models. Meta-Radiology 100017.

Liu, J., Snodgrass, S., Khalifa, A., Risi, S., Yannakakis, G.N., Togelius, J., 2021b. Deep learning for procedural content generation. Neural Comput. Appl. 33 (1), 19–37.

McKay, C., 2020. Predicting risk in criminal procedure: actuarial tools, algorithms, AI and judicial decision-making. Curr. Issues Crim. Justice 32 (1), 22–39.

Min, B., Ross, H., Sulem, E., Veyseh, A.P.B., Nguyen, T.H., Sainz, O., Agirre, E., Heintz, I., Roth, D., 2023. Recent advances in natural language processing via large pre-trained language models: A survey. ACM Comput. Surv. 56 (2), 1–40.

Morison, J., Harkens, A., 2019. Re-engineering justice? Robot judges, computerised courts and (semi) automated legal decision-making. Legal Stud. 39 (4), 618–635.

Nadkarni, P.M., Ohno-Machado, L., Chapman, W.W., 2011. Natural language processing: an introduction. J. Am. Med. Inform. Assoc. 18 (5), 544–551.

Nguyen, H.T., 2023. A brief report on LawGPT 1.0: A virtual legal assistant based on GPT-3. arXiv preprint, arXiv:2302.05729.

Ni, J., Abrego, G.H., Constant, N., Ma, J., Hall, K., Cer, D., Yang, Y., 2022. Sentence-T5: Scalable sentence encoders from pre-trained text-to-text models. In: Findings of the Association for Computational Linguistics. pp. 1864–1874.

Niu, Z., Zhong, G., Yu, H., 2021. A review on the attention mechanism of deep learning. Neurocomputing 452, 48–62.

Nowotko, P.M., 2021. AI in judicial application of law and the right to a court. Procedia Comput. Sci. 192, 2220–2228.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., et al., 2022. Training language models to follow instructions with human feedback. Adv. Neural Inf. Process. Syst. 35, 27730–27744.

Pan, Z., Yu, W., Yi, X., Khan, A., Yuan, F., Zheng, Y., 2019. Recent progress on generative adversarial networks (GANs): A survey. IEEE Access 7, 36322–36333.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. PyTorch: An imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 32.

Pedreschi, D., Giannotti, F., Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., 2019. Meaningful explanations of black box AI decision systems. In: The AAAI Conference on Artificial Intelligence. pp. 9780–9784.

Prescott, J.J., 2017. Improving access to justice in state courts with platform technology. Vanderbilt Law Rev. 70, 1993.

Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., et al., 2018. Improving language understanding by generative pre-training.

Rai, A., 2020. Explainable AI: From black box to glass box. J. Acad. Market. Sci. 48, 137–141.

Raso, F.A., Hilligoss, H., Krishnamurthy, V., Bavitz, C., Kim, L., 2018. Artificial Intelligence & Human Rights: opportunities & Risks. Berkman Klein Center Research Publication.

Re, R.M., Solow-Niederman, A., 2019. Developing artificially intelligent justice. Stanf. Technol. Law Rev. 22, 242.

Reddy, B.R., Mahender, E., 2013. Speech to text conversion using Android platform. Int. J. Eng. Res. Appl. 3 (1), 253–258.

Rissland, E.L., Ashley, K.D., Loui, R.P., 2003. AI and law: A fruitful synergy. Artificial Intelligence 150 (1–2), 1–15.

Riveiro, M., Thill, S., 2021. "That's (not) the output I expected!" On the role of end user expectations in creating explanations of AI systems. Artificial Intelligence 298, 103507.

Rodrigues, R., 2020. Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. J. Responsib. Technol. 4, 100005.

Rosser, G., Davies, T., Bowers, K.J., Johnson, S.D., Cheng, T., 2017. Predictive crime mapping: Arbitrary grids or street networks? J. Quant. Criminol. 33, 569–594.

Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., Aberman, K., 2023. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In: The IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22500–22510.

Saon, G., Kuo, H.K.J., Rennie, S., Picheny, M., 2015. The IBM 2015 English conversational telephone speech recognition system. In: Annual Conference of the International Speech Communication Association. pp. 3140–3144.

Šarčević, S., 2016. Basic principles of term formation in the multilingual and multicultural context of EU law. In: Language and Culture in EU Law. Routledge, pp. 183–206.

Scao, T.L., Wang, T., Hesslow, D., Saulnier, L., Bekman, S., Bari, M.S., Bideman, S., Elsahar, H., Muennighoff, N., Phang, J., et al., 2022. What language model to train if you have one million GPU hours? arXiv preprint, arXiv:2210.15424.

Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., Hall, P., et al., 2022. Towards a standard for identifying and managing bias in artificial intelligence. NIST Spec. Publ. 1270 (10.6028).

Shaw, P., Uszkoreit, J., Vaswani, A., 2018. Self-attention with relative position representations. In: Proceedings of NAACL-HLT. pp. 464–468.

Shi, C., Sourdin, T., Li, B., 2021. The smart court-a new pathway to justice in China? Int. J. Court Admin. 12, 1.

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. J. Big Data 6 (1), 1–48.

Sleimi, A., Sannier, N., Sabetzadeh, M., Briand, L., Dann, J., 2018. Automated extraction of semantic legal metadata using natural language processing. In: IEEE 26th International Requirements Engineering Conference. IEEE, pp. 124–135.

Solaiman, I., Dennison, C., 2021. Process for adapting language models to society (PALMS) with values-targeted datasets. Adv. Neural Inf. Process. Syst. 34, 5861–5873.

Sourdin, T., 2018. Judge v robot?: Artificial intelligence and judicial decision-making. Univ. New South Wales Law J. 41 (4), 1114–1133.

Sparkes, M., 2023. AI will advise a defendant in court.

Stern, R.E., Liebman, B.L., Roberts, M.E., Wang, A.Z., 2020. Automating fairness? Artificial intelligence in the Chinese courts. Colum. J. Transnat'l L. 59, 515.

Sun, J., Gan, W., Chao, H.C., Yu, P.S., Ding, W., 2023. Internet of behaviors: A survey. IEEE Internet Things J. 10 (13), 11117–11134.

Sung, H.C., 2020. Can online courts promote access to justice? A case study of the internet courts in China. Comput. Law Secur. Rev. 39, 105461.

Surden, H., 2019. Artificial intelligence and law: An overview. Georgia State Univ. Law Rev. 35, 19–22.

Sutskever, I., Vinyals, O., Le, Q.V., 2014. Sequence to sequence learning with neural networks. Adv. Neural Inf. Process. Syst. 27.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M., Frame, M., 2011. Data sharing by scientists: practices and perceptions. PLoS One 6 (6), e21101.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al., 2023. LLaMA: Open and efficient foundation language models. arXiv preprint, arXiv:2302.13971.

Trivedi, A., Pant, N., Shah, P., Sonik, S., Agrawal, S., 2018. Speech to text and text to speech recognition systems-Areview. IOSR J. Comput. Eng. 20 (2), 36–43.

Tu, Z., Wang, Y., Birkbeck, N., Adsumilli, B., Bovik, A.C., 2021. UGC-VQA: Benchmarking blind video quality assessment for user generated content. IEEE Trans. Image Process. 30, 4449–4464.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.

Venugopalan, S., Rohrbach, M., Donahue, J., Mooney, R., Darrell, T., Saenko, K., 2015. Sequence to sequence-video to text. In: IEEE International Conference on Computer Vision. pp. 4534–4542.

Voita, E., Sennrich, R., Titov, I., 2019. The bottom-up evolution of representations in the transformer: A study with machine translation and language modeling objectives. In: EMNLP-IJCNLP. pp. 4396–4406.

Wachter, S., Mittelstadt, B., Russell, C., 2021. Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. Comput. Law Secur. Rev. 41, 105567.

Waterman, D.A., Paul, J., Peterson, M., 1986. Expert systems for legal decision making. Expert Syst. 3 (4), 212–226.

Wei, L., Fengru, Z., 2021. Innovative research on legal talents training model in the era of artificial intelligence. In: 16th International Conference on Computer Science & Education. IEEE, pp. 257–262.

Wei, K., Li, J., Ding, M., Ma, C., Yang, H.H., Farokhi, F., Jin, S., Quek, T.Q., Poor, H.V., 2020. Federated learning with differential privacy: Algorithms and performance analysis. IEEE Trans. Inf. Forensics Secur. 15, 3454–3469.

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., et al., 2022. Emergent abilities of large language models. arXiv preprint, arXiv:2206.07682.

Wexler, R., 2021. Privacy asymmetries: Access to data in criminal defense investigations. UCLA Law Rev. 68, 212.

Wu, J., Gan, W., Chen, Z., Wan, S., Lin, H., 2023. AI-generated content (AIGC): A survey. arXiv preprint, arXiv:2304.06632.

Wu, Y., Schuster, M., Chen, Z., Le, Q.V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., et al., 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint, arXiv:1609.08144.

Xiao, C., Hu, X., Liu, Z., Tu, C., Sun, M., 2021. Lawformer: A pre-trained language model for chinese legal long documents. AI Open 2, 79–84.

Xu, Z., 2022. Human judges in the era of artificial intelligence: challenges and opportunities. Appl. Artif. Intell. 36 (1), 2013652.

Xu, J., Liu, C., Kuang, K., et al., 2024. Legal large model evaluation indicators and methods. CAAI Communications 16–23.

Yang, J., Ma, W., Zhang, M., Zhou, X., Liu, Y., Ma, S., 2021. LegalGNN: Legal information enhanced graph neural network for recommendation. ACM Trans. Inf. Syst. 40 (2), 1–29.

Yang, Y., Wipf, D.P., et al., 2022. Transformers from an optimization perspective. Adv. Neural Inf. Process. Syst. 35, 36958–36971.

Yu, R., Alì, G.S., 2019. What's inside the black box? AI challenges for lawyers and researchers. Legal Inf. Manag. 19 (1), 2–13.

Završnik, A., 2020. Criminal justice, artificial intelligence systems, and human rights. In: ERA Forum, vol. 20, (no. 4), Springer, pp. 567–583.

Zeng, F., Gan, W., Wang, Y., Liu, N., Yu, P.S., 2023. Large language models for robotics: A survey. arXiv preprint, arXiv:2311.07226.

Zhang, Y., Jiang, J., Li, Y., 2023. Intelligent analysis and application of judicial big data sharing based on blockchain. In: 6th International Conference on Artificial Intelligence and Big Data. IEEE, pp. 592–596.

Zhang, Y., Jin, R., Zhou, Z.-H., 2010. Understanding bag-of-words model: a statistical framework. Int. J. Mach. Learn. Cybern. 1, 43–52.

Zhang, C., Lu, Y., 2021. Study on artificial intelligence: The state of the art and future prospects. J. Ind. Inf. Integr. 23, 100224.

Zhang, Y., Sun, S., Galley, M., Chen, Y.-C., Brockett, C., Gao, X., Gao, J., Liu, J., Dolan, B., 2020. DIALOGPT: Large-scale generative pre-training for conversational response generation. In: The 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. pp. 270–278.

Zhao, W.X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., et al., 2023. A survey of large language models. arXiv preprint, arXiv:2303.18223.

Zhong, H., Wang, Y., Tu, C., Zhang, T., Liu, Z., Sun, M., 2020a. Iteratively questioning and answering for interpretable legal judgment prediction. In: The AAAI Conference on Artificial Intelligence. pp. 1250–1257.

Zhong, H., Xiao, C., Tu, C., Zhang, T., Liu, Z., Sun, M., 2020b. How does NLP benefit legal system: A summary of legal artificial intelligence. arXiv preprint, arXiv:2004.12158.

Zhong, H., Xiao, C., Tu, C., Zhang, T., Liu, Z., Sun, M., 2020c. JEC-QA: a legal-domain question answering dataset. In: The AAAI Conference on Artificial Intelligence. pp. 9701–9708.

Zuiderveen Borgesius, F.J., 2020. Strengthening legal protection against discrimination by algorithms and artificial intelligence. Int. J. Hum. Rights 24 (10), 1572–1593.