2023 15th International Conference on Knowledge and Systems Engineering (KSE) | 979-8-3503-2974-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/KSE59128.2023.10299488

The Impact of Large Language Modeling on Natural Language Processing in Legal Texts: A Comprehensive Survey

Dang Hoang Anh[1], Dinh-Truong Do[1], Vu Tran[2], and Nguyen Le Minh[1]
[1]*Japan Advanced Institute of Science and Technology*
{s2220409,truongdo,nguyenml}@jaist.ac.jp
[2]*The Institute of Statistical Mathematics, Japan*
vutran@ism.ac.jp

***Abstract*—Natural Language Processing (NLP) has witnessed significant advancements in recent years, particularly with the emergence of large language models. These models, such as GPT-3.5 and its variants, have revolutionized various domains, including legal text processing (LTP). This survey explores the impact of large language modeling on NLP in the context of legal texts. By analyzing the latest research and developments, we seek to understand the benefits, challenges, and potential applications of large language models in the field of legal language processing.**

***Index Terms*—Large language modeling, natural language modeling, legal texts**

I. Overview of Large Language Models

A. Definition and Characteristics

Large language models (LLMs) are deep learning models designed to process natural language. They are characterized by their massive size, typically comprising hundreds of millions to billions of parameters. These models are trained on vast amounts of text data from diverse sources, which enables them to capture complex linguistic patterns and semantic relationships [27].

TABLE I
Comparision with previous surveys

Surveys	**Topics**		
	LLMs	***Multi-domain***	***Latest year***
Chalkidis et al. [4]	no	yes	2019
Cui et al. [8]	no	no	2022
Dias et al. [10]	no	yes	2022
Katz et al. [14]	no	yes	2023
Ours	yes	yes	2023

We summarize the main topics and problem statements to survey

B. Architecture

LLMs are often built upon transformer-based architectures, which have proven to be highly effective in processing sequential data, such as sentences and paragraphs. The transformer architecture was first introduced by Vaswani et al. in 2017 [27]. It replaces the traditional recurrent neural network (RNN) layers with attention mechanisms, allowing the model to process input tokens in parallel, significantly speeding up training and inference.

Transformers consist of an encoder-decoder architecture, but in the case of LLMs used for tasks like text generation and understanding, only the decoder part is typically employed. The decoder comprises multiple layers of self-attention and feed-forward neural networks. These layers work together to encode the input text into rich contextual embeddings, capturing both local and global dependencies within the text.

C. Pre-training Process

The pre-training of LLMs is a crucial step that enables their exceptional performance. It involves training the model on a vast corpus of unsupervised text data. Casual Language Modeling (CLM) and Masked Language Modeling (MLM) are two popular approaches used in pre-training LLMs.

In CLM, a LLM is trained to predict the next word in a sentence, given the context of the preceding words. During pre-training, the LLM processes a large corpus of text data and tries to learn the underlying grammar, syntax, and semantic relationships, becoming proficient in generating coherent and contextually appropriate text. Decoder-only models, such as GPT-3 [3], employ this objective function during pre-training. In contrast to CLM, MLM is a slightly different approach that was popularized by BERT [9]. During pre-training, MLM randomly masks certain words or tokens in the input text, and the model is tasked with predicting those masked tokens. This approach differs from CLM as it requires the model to make use of the context both before and after the masked word to make accurate predictions.

After pre-training, these models can be fine-tuned to follow instructions [6] or perform specific downstream tasks, such as question-answering or text summarization [29]. Overall, LLMs represent a breakthrough in NLP, enabling machines to understand and generate human-like language with unprecedented accuracy and versatility. In the legal domain, these models can potentially revolutionize LTP, making it more efficient and accessible for legal professionals and the general public alike.

II. Legal Text Processing

A. Challenges in Legal Language

Legal texts are notorious for their complexity, formal language, and specialized terminology. Understanding legal documents requires not only a strong grasp of language but also an in-depth knowledge of legal concepts and principles. Some of the specific challenges in LTP include [14, 13]:

1) Ambiguity: Legal language often contains ambiguous terms and phrases, which can lead to different interpretations and outcomes. Resolving these ambiguities requires context awareness and domain-specific knowledge [11].

2) Long and Complex Sentences: Legal documents are characterized by lengthy and convoluted sentences. Parsing and understanding such complex structures is challenging for traditional NLP systems. While LLMs have access to vast amounts of data during pre-training, they lack true world knowledge. Unlike humans, who possess general knowledge about the world and can reason using this knowledge, LLMs rely solely on patterns learned from the training data. This limitation can hinder their ability to understand complex scenarios requiring in-depth background knowledge. Especially if the data contain highly nested structures or excessive repetitions. In such cases, the model might encounter difficulties in fully understanding the sentence or may produce outputs with reduced coherence.

3) Context Sensitivity: Legal language heavily relies on context to derive meaning. The interpretation of a word or phrase may vary depending on the surrounding text, making context-sensitive processing crucial.

B. Traditional Approaches in Processing Legal Text

Traditionally, legal text processing has relied on some machine-learning approaches, including:

1) Rule-based Method: Rule-based systems use manually defined rules to extract relevant information from legal texts. While effective for simple tasks, they often struggle with the complexity and ambiguity present in legal language.

2) Information Retrieval: Information retrieval systems enable the efficient and accurate retrieval of relevant legal information from a variety of legal sources. However, they usually do not capture the context and semantic relationships between documents.

3) Named Entity Recognition (NER): NER systems identify and classify entities such as names, dates, and locations in legal texts. However, they may struggle with recognizing domain-specific entities and uncommon legal terms.

C. How LLMs Overcome Challenges in Legal Text Processing

1) Ambiguity Resolution: LLMs excel at understanding context, which allows them to disambiguate complex legal terms and phrases. The extensive pre-training on diverse data helps them grasp the various ways certain legal terms are used, enhancing their ability to provide more accurate interpretations [5]. The self-attention mechanism in transformers allows the model to weigh the importance of each token relative to other tokens in the sequence. When processing long sentences, we can assign higher attention to the relevant parts of the sentence, capturing dependencies across long distances effectively. This enables them to understand the connections between different parts of the sentence, even in complex structures.

The parallel nature of the transformer architecture allows LLMs to process long sentences more efficiently. While RNNs process sequences sequentially, transformers can process all tokens in parallel. This parallel computation significantly speeds up the processing of long sentences and avoids the limitations associated with sequential models. During pre-training, the models are exposed to a wide variety of texts from the internet. This diverse training data includes long and complex sentences, which helps the models learn to understand and generate such sentences effectively. After pre-training, fine-tuning on domain-specific data enables them to adapt their language understanding to the particular characteristics and complexity of the target domain, improving their performance on long and complex sentences within that domain. Despite their capabilities, they may struggle with certain types of ambiguity, especially when the context is limited or when encountering rare or domain-specific words. Additionally, LLMs might not always provide the correct disambiguation, as their performance heavily depends on the quality and diversity of the training data.

2) Long and Complex Sentence Comprehension: Transformers, the underlying architecture of LLMs, are designed to handle long-range dependencies in sentences. This enables them to comprehend and generate lengthy legal texts more effectively than traditional NLP models.

LLMs can be sensitive to the phrasing and ordering of input sentences. Slight changes in input formulation might lead to different model responses, making them less robust to variations in how the same information is presented to avoid the sensitivity to the phrasing and ordering of input sentences. The model may struggle with negation and uncertainty in context, leading to potential misunderstandings. For example, a negative statement may be interpreted as positive or vice versa. If the training data contains biased information, the model's responses might reflect and perpetuate these biases. LLMs may provide plausible-sounding responses even when the content is implausible or nonsensical in the real world. This limitation stems from the lack of genuine understanding of common sense and real-world reasoning.

3) Contextual Understanding: LLMs' contextual embeddings capture the meaning of words based on their surroundings, leading to better context-sensitive language processing. This feature proves crucial in disambiguating legal language and understanding the subtle nuances of legal texts. During pre-training, LLMs are exposed to a vast and diverse corpus of text data, which includes legal texts such as court opinions, statutes, contracts, and legal articles. This exposure allows the models to learn legal terminology, principles, and the context of legal language. By training on legal text, they could develop a foundational understanding of legal concepts and jargon, which helps them process and generate text related to the legal domain [27].

After pre-training, the models can be fine-tuned on domain-specific legal data and tasks. Fine-tuning involves training the model on labeled legal datasets that are relevant to the target legal task. Fine-tuning on legal text helps to adapt their language understanding to the nuances and specific requirements of the legal domain, improving their performance in processing legal documents and generating legal text.
LLMs have a limited context window, which means they can only consider a fixed number of tokens in their input sequence. In legal text, where sentences and documents can be lengthy and complex, this limitation poses a challenge. Long sentences might be truncated or split into smaller parts, potentially leading to the loss of crucial context. Despite this limitation, we still can leverage their attention mechanisms to attend to relevant parts of the sentence and capture important dependencies for context understanding.

In legal text, understanding the context of a sentence or case often involves referencing previous cases or legal precedents. The ability to handle context and generate text-based responses allows them to perform legal reasoning tasks, drawing connections to similar past cases and their outcomes. Their broad knowledge of legal documents enables them to provide contextually appropriate responses for legal questions. However, the problem of "hallucinations" can arise when using LLMs for legal text generation [7]. Hallucinations, in this context, refer to the generation of false or inaccurate information that may seem contextually relevant but is not based on actual legal precedents or cases. These inaccuracies can potentially lead to erroneous legal conclusions and misguided advice. Therefore, ensuring that LLMs are well-trained and regularly updated with accurate legal information is crucial to mitigate the risk of generating such hallucinations and to maintain the integrity of legal text generation for reliable legal reasoning and guidance.

4) Domain-Specific Fine-Tuning: LLMs can be fine-tuned on legal corpora, enabling them to specialize in legal language and terminology. This fine-tuning process enhances their performance on legal tasks and makes them more adept at handling domain-specific challenges [1]. Before fine-tuning, LLMs undergo a pre-training phase on a massive corpus of diverse text data, which includes general language from the internet. This pre-training process enables the models to acquire a broad understanding of natural language, syntactic structures, and semantic patterns. It provides a solid foundation for language understanding, which is beneficial across various domains.

To fine-tune for a specific domain, a dataset is collected from the target domain. This dataset contains labeled examples relevant to the domain-specific task the model is intended to perform. For instance, if the aim is to use the LLM for LTP, a dataset of labeled legal texts (e.g., court opinions, legal articles, contracts) is compiled. Fine-tuning helps it adapt to the nuances and conventions of the target domain. The model learns to recognize domain-specific terminology, context, and language usage. For example, in LTP, the fine-tuned model gains familiarity with legal jargon, case law, and contract language. Even after fine-tuning, LLMs retain their general language understanding from the pre-training phase. This enables them to be versatile and potentially handle multiple domains effectively. By combining pre-trained knowledge with domain-specific adaptation, we can provide contextually appropriate responses within their fine-tuned domain while still being capable of handling general language tasks.

By undergoing domain-specific fine-tuning, they can be tailored to excel in a wide range of applications across various domains, empowering them to provide more accurate and contextually appropriate responses within specific tasks and industries. Fine-tuning is a crucial step in realizing the full potential of LLMs for domain-specific NLP.

5) Document Summarization and Information Extraction: LLMs' capabilities extend beyond individual sentence understanding. They can summarize lengthy legal documents, extract key information, and identify relevant entities, facilitating efficient legal research and analysis.

Overall, the flexibility, context awareness, and domain adaptation capabilities of LLMs make them invaluable tools for LTP. By leveraging these models, legal professionals can streamline their workflows, extract insights from vast volumes of legal documents, and gain a deeper understanding of legal language and concepts. As advancements continue in the field of LLMs, their impact on the legal domain is likely to grow, further revolutionizing the way legal texts are analyzed, processed, and understood.

III. Applications of LLMs in Legal Text Processing

A. *Legal Document Summarization*

Legal documents, such as court opinions, contracts, and legislation, are often lengthy and dense, making it time-consuming for legal professionals to extract essential information. LLMs can be employed for legal document summarization, where they generate concise and coherent summaries of lengthy texts. By identifying the key arguments, rulings, and crucial points, these models can assist legal professionals in quickly understanding the content of complex legal documents. Legal document summarization can significantly improve efficiency in legal research and enable better decision-making for lawyers, judges, and other legal practitioners [18].

B. *Contract Analysis and Generation*

Contract analysis is a critical task in legal domain, where lawyers need to review and understand the terms and conditions of legal agreements. LLMs can aid in contract analysis by automatically extracting and highlighting essential clauses, potential risks, and legal obligations from contracts. Additionally, these models can be utilized to generate contract templates based on specific requirements. This capability streamlines the contract drafting process and reduces the risk of errors, leading to more accurate and standardized contracts [15].

C. Legal Question Answering

Legal professionals often encounter complex legal queries that require extensive research to find relevant information from a vast corpus of legal texts and case law. LLMs can be used for legal question-answering (QA), where they analyze the query, search through legal documents, and provide precise and contextually relevant answers. By efficiently retrieving legal information, legal QA systems powered by LLMs can save valuable time for lawyers and enhance the accessibility of legal knowledge for the general public [17].

D. Legal Text Classification

LLMs can significantly improve the accuracy of legal text classification tasks. Legal texts cover a wide range of topics, from civil law to criminal law, environmental law, and intellectual property law. By fine-tuning LLMs on labeled legal datasets, these models can classify legal documents into specific categories or topics. This application has multiple uses, such as organizing legal document repositories, routing cases to appropriate legal teams, and facilitating legal research on specific legal issues [25].

E. Legal Information Extraction

Legal information extraction involves identifying and extracting specific information from legal texts, such as entities (e.g., names, dates, organizations), relationships between entities, and events (e.g., court decisions). With their contextual understanding and ability to generate embeddings, LLMs can excel at this task. Legal information extraction can enhance legal data analysis, support the creation of legal knowledge graphs, and contribute to the development of AI-driven legal research platforms [24]. The applications of LLMs in LTP have the potential to revolutionize the legal profession by streamlining various tasks, improving decision-making, and enhancing the accessibility of legal information. However, there are also challenges to address, such as ensuring the ethical and unbiased use of these models, addressing data privacy concerns, and providing transparency in the decision-making process.

F. Legal reasoning

While LLMs cannot engage in formal logical reasoning or manipulate Abstract Meaning Representation (AMR) graphs directly, they can perform legal reasoning in a more implicit and data-driven manner. Legal reasoning often involves understanding legal concepts, applying laws to specific cases, and deriving conclusions based on legal principles and precedents [19]. The models can handle legal reasoning tasks by leveraging their pre-trained language understanding and pattern recognition capabilities: *Legal Language Understanding*, *Precedent and Case Law Analysis*, *Text Generation and Argumentation*, or *Common-Sense Reasoning*, which is crucial for understanding implied meanings and context in legal language.

However, without exhibiting impressive reasoning capabilities, they still have certain limitations, including:

- *Lack of Explicit Logic:* These models lack a built-in logical reasoning mechanism. Hence, their reasoning is based on learned patterns and associations in the training data rather than formal logical rules.
- *Legal Context and Specificity:* Although LLMs have been exposed to legal texts, they may not have the same level of expertise and contextual understanding as legal professionals, struggling with highly specialized legal concepts and nuances.
- *Biases and Interpretability:* The biases present in the training data, leading to potential biases in legal reasoning outcomes. Additionally, their lack of interpretability makes it challenging to understand the reasoning process behind their decisions.

In summary, while LLMs do not perform legal reasoning using explicit logic or AMR, they can demonstrate sophisticated language understanding and generate text-based responses that resemble legal reasoning. Their strength lies in their ability to process vast amounts of natural language data, but they also have limitations that need to be considered when applying them to legal tasks. Legal professionals should use LLMs as valuable tools to support legal reasoning and not as replacements for human expertise and judgment.

IV. Ethical and Legal Considerations

A. Bias and Fairness

LLMs are trained on vast amounts of data from the internet, which may contain biased information and reflect societal prejudices. As a result, these models can inadvertently learn and perpetuate biases in their responses, including gender, race, and socio-economic biases. In the context of LTP, bias can have significant implications, leading to unfair legal decisions and reinforcing existing disparities in the legal system. Addressing bias in LLMs is a crucial ethical consideration. Research and development efforts should focus on reducing bias during model training and fine-tuning. Techniques such as data augmentation, adversarial training, and fairness-aware learning can help mitigate biases. Additionally, incorporating diverse and representative training data can lead to more equitable and unbiased language models for legal applications.

B. Privacy and Confidentiality

Legal texts often contain sensitive information, including personally identifiable information (PII) and private case details. Using LLMs to process such data raises privacy concerns. Legal professionals and users of these models must ensure that appropriate measures are in place to protect the privacy and confidentiality of the individuals involved in legal cases. Privacy-preserving techniques, such as data anonymization, differential privacy, and federated learning, can help safeguard sensitive legal information while benefiting from LLMs' capabilities. Compliance with data protection regulations, such as the General Data Protection Regulation (GDPR), is essential when handling personal data in LTP applications.

C. Intellectual Property

LLMs are typically trained on massive amounts of publicly available data, which might include copyrighted material. This raises concerns about potential intellectual property infringement. Using copyrighted content without permission or proper licensing can lead to legal issues for both the developers and users of these models. Researchers and organizations working with LLMs should carefully consider copyright and licensing implications. Utilizing publicly available and open-access datasets can help mitigate copyright concerns. Additionally, collaborations with content providers and rights holders can pave the way for responsible and legally compliant usage of copyrighted materials in LLMs.

D. Explainability and Transparency

LLMs, particularly those with billions of parameters, can be perceived as black boxes due to their complexity. The lack of transparency in model decision-making raises challenges for legal professionals who need to understand the reasoning behind model outputs, especially in critical legal contexts. Explainable AI techniques, such as attention visualization, saliency maps, and rule-based post-hoc explanations, can provide insights into how LLMs arrive at their decisions. Ensuring transparency in model behavior is crucial for gaining the trust of legal professionals and ensuring the accountability of AI-powered legal applications.

E. Responsible Use and Regulatory Framework

The increasing adoption of LLMs in LTP necessitates a responsible and ethical approach to their development and deployment. It is crucial to establish clear guidelines and regulatory frameworks for using these models in legal domain. A comprehensive regulatory framework should address issues such as data privacy, bias mitigation, transparency, and intellectual property rights. Collaboration between AI researchers, legal experts, policymakers, and stakeholders is essential to develop guidelines that ensure the responsible and ethical use of LLMs in LTP.As LLMs continue to advance and become integral to LTP, it is imperative to address the ethical and legal considerations associated with their usage. By prioritizing fairness, privacy, transparency, and responsible development, we can harness the potential of LLMs to revolutionize legal language processing while upholding the values of justice, equality, and accountability in the legal system.

V. Empirical Study

In this section, our goal aims to gain deeper insights into the impact of LLMs in the legal domain by conducting experiments on the legal question-answering task.

A. Dataset and Evaluation Metric

For our experiments, we utilize the legal textual entailment (yes/no question-answering) dataset from COLIEE 2022[1], a well-known legal competition for researchers. This dataset includes both English and Japanese versions. The objective of the task is to determine whether a given question Q can be answered ”Yes” or ”No” based on a set of relevant articles denoted as $A = \{A_1, A_2, ..., A_n\}$. The dataset comprises 887 training samples and 109 test samples. For evaluating the performance of our method, we use accuracy as the metric, which is calculated as the number of correctly predicted questions divided by the total number of questions.

B. Method

The proposed method for tackling the legal textual entailment task in a zero-shot setting using LLMs is outlined in Figure 1. The process begins with creating a set of prompts (instructions). These prompts are obtained by collecting prompts for the GLUE tasks available in the PromptSource library [2]. After converting these prompts to JSON format, we end up with 56 prompts that can serve as input instructions for LLMs.

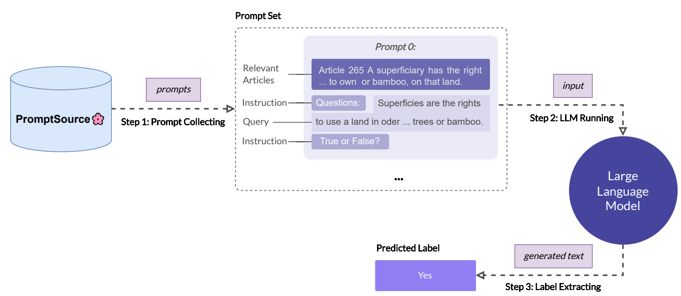

Fig. 1. Overview of the proposed method.

Next, we load the LLMs using the Huggingface library[2] and provide them with the collected prompts, along with the legal questions and the relevant articles. The LLMs then generate output text, from which we extract the answer by searching for negative labels such as ”false” or ”incorrect” in the generated text. If any of these labels are found, the function returns 0; otherwise, it returns 1.

C. Experimental Settings

The experiments involve testing several LLMs. Two models, namely Flan-t5-xxl and Flan-Ul2, are initialized with the model checkpoints T5-xxl [20] and Ul2 [26], respectively. They were then fine-tuned on the Flan dataset [6], which focuses on thinking and reasoning tasks. T0pp and Bloomz-7b1 are instruction-tuned versions of T5-xxl [20] and Bloom-7b1 [23], respectively, on the P3 instruction set [16]. Additionally, Mt0-xxl is instruction-adjusted versions of the Mt5 model [28], trained on the xP3 multilingual task mixed dataset [16]. In this experiment, the Mt0-xxl model is specifically used for the Japanese version of the dataset, while the other models are employed for the English version.

[1] https://sites.ualberta.ca/~rabelo/COLIEE2022/

[2] https://huggingface.co/

D. Experimental Results

The experimental results are presented in Table II. We compared our findings with two top-performing methods [12, 30] from the COLIEE 2022 legal textual entailment task. The outcomes demonstrate that LLMs outperform non-LLM methods that rely on medium-sized pre-trained language models (shown in rows 1, 2, and 3). This highlights the remarkable performance of LLMs when appropriately instructed for legal tasks. Interestingly, larger model parameter sizes do not always translate to improved performance, as indicated in rows 3 and 4, consistent with previous research [21]. Additionally, we compared the performance of the Flan-t5-xxl (4) and T0pp (5) models, both have a similar base model however fine-tuned on different instruction data (Flan [6] and P3 [22], respectively). The results showed that choosing the right fine-tuning instruction data significantly impacts the performance of LLMs in legal tasks.

TABLE II
EXPERIMENTAL RESULTS ON THE LEGAL TEXTUAL ENTAILMENT TASK.

Description	Base model	Params	Accuracy
(1) Best of 2022 [12]	BERT-based model	$<$ 1B	0.6931
(2) 2nd best of 2022 [30]	BERT-based model	$<$ 1B	0.6697
(3) flan-ul2-zero-shot (ours)	ul2	20B	**0.7889**
(4) flan-t5-xxl-zero-shot (ours)	t5-xxl	11B	**0.7889**
(5) t0pp-zero-shot (ours)	t5-xxl	11B	0.7339
(6) bloomz-7b1-zero-shot (ours)	bloom-7b1	7B	0.6422
(7) mt0-xxl-zero-shot (ours)	mt5-xxl	13B	0.7155

Furthermore, we observed that the choice of prompt significantly affects LLMs' performance. For instance, using the most effective prompt resulted in 78.89% accuracy, while the least effective prompt achieved only 55.04% accuracy. This underscores the importance of selecting appropriate prompts for LLMs or exploring methods for automatic prompt generation, which are potential areas of future research.

VI. Future Directions and Open Challenges

A. Future direction

1) Multilingual Legal Text Processing: Future research and development efforts should focus on advancing LLMs to handle legal texts in multiple languages effectively. Multilingual LTP can be beneficial for cross-border legal analysis, international legal research, and legal translation. Training language models on diverse legal corpora from different jurisdictions and legal systems can help improve their performance in handling legal texts in various languages.

2) Domain-Specific Fine-Tuning: One of the promising future directions is fine-tuning LLMs on more specific legal domains. While pre-training on a vast corpus provides a general understanding of language, fine-tuning domain-specific legal data can enhance the model's performance and domain-specific knowledge. Legal professionals can fine-tune the models on their specific legal documents or cases, tailoring the models to their precise needs and improving their accuracy in LTP tasks.

3) Integration with Legal Research Tools: To fully leverage the capabilities of LLMs, integrating them with existing legal research platforms and tools is a logical future direction. By integrating language models with legal databases, case law repositories, and legal search engines, legal professionals can access advanced NLP capabilities directly within their existing workflows. This integration can streamline legal research, improve information retrieval, and enhance the efficiency of legal professionals.

B. Challenges

1) Interpretability and Explainability: As LLMs become more complex and powerful, interpreting their decisions becomes increasingly challenging. Ensuring the interpretability and explainability of model outputs in legal contexts is essential for legal professionals to trust and adopt these models. Addressing this challenge involves developing effective explainable AI techniques that shed light on how LLMs arrive at their conclusions, especially in complex legal cases.

2) Limited Access to Legal Data: Obtaining high-quality labeled legal datasets for training and fine-tuning LLMs can be a challenge. Legal data is often sensitive and subject to copyright restrictions, making it difficult to access and use in research. Efforts to create publicly available legal datasets and promote data sharing among researchers can help mitigate this challenge and facilitate progress in LTP research.

3) Handling Rare and Domain-Specific Legal Concepts: LLMs may struggle with rare or domain-specific legal concepts that are not well-represented in the training data. Developing strategies to address out-of-vocabulary (OOV) terms and low-resource legal domains will be crucial to ensuring the accuracy and effectiveness of LLMs in LTP.

4) Ethical Concerns and Bias Mitigation: As mentioned in the previous part, addressing ethical concerns and mitigating bias is a significant challenge in deploying LLMs in LTP. Ensuring fairness and avoiding unjust legal decisions requires continuous research and development of bias mitigation techniques tailored to legal applications' specific needs.

By acknowledging and addressing these future directions and challenges, researchers and practitioners can unlock the full potential of LLMs in LTP, transforming how legal information is analyzed, interpreted, and made accessible to legal professionals and the general public.

VII. Conclusion

LLMs have transformed the landscape of NLP, offering immense potential for legal text processing. This survey has provided a comprehensive overview of the impact of large language modeling on NLP in legal texts. While highlighting the benefits and applications, we also addressed the ethical, legal, and technical challenges that arise. By understanding these implications, we can harness the power of LLMs to revolutionize legal language processing, making it more efficient, accurate, and accessible.

References

[1] Al-qurishi Muhammad et al. *AraLegal-BERT: A pre-trained language model for Arabic Legal text*. Abu Dhabi, United Arab Emirates (Hybrid), Dec. 2022. URL: https://aclanthology.org/2022.nllp-1.31.

[2] Stephen Bach et al. "PromptSource: An Integrated Development Environment and Repository for Natural Language Prompts". In: *Proceedings of ACL 2022: System Demonstrations*. Dublin, Ireland, May 2022, pp. 93–104. DOI: 10.18653/v1/2022.acl-demo.9.

[3] Tom Brown et al. "Language models are few-shot learners". In: *Advances in neural information processing systems* 33 (2020), pp. 1877–1901.

[4] Ilias Chalkidis and Dimitrios Kampas. "Deep learning in law: early adaptation and legal word embeddings trained on large corpora". In: *Artificial Intelligence and Law* 27 (June 2019), pp. 1–28. DOI: 10.1007/s10506-018-9238-9.

[5] Jonathan H. Choi et al. *ChatGPT Goes to Law School (January 23, 2023). Journal of Legal Education (Forthcoming), Available at SSRN: http://dx.doi.org/10.2139/ssrn.4335905*. 2023.

[6] Hyung Won Chung et al. "Scaling instruction-finetuned language models". In: *arXiv preprint arXiv:2210.11416* (2022).

[7] Jiaxi Cui et al. "Chatlaw: Open-source legal large language model with integrated external knowledge bases". In: *arXiv preprint arXiv:2306.16092* (2023).

[8] Junyun Cui et al. *A Survey on Legal Judgment Prediction: Datasets, Metrics, Models and Challenges*. 2022. arXiv: 2204.04859 [cs.CL].

[9] Jacob Devlin et al. "Bert: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805* (2018).

[10] João Dias et al. *State of the Art in Artificial Intelligence applied to the Legal Domain*. 2022. arXiv: 2204.07047.

[11] Virginia K. Felkner et al. *Towards WinoQueer: Developing a Benchmark for Anti-Queer Bias in Large Language Models*. 2022. arXiv: 2206.11484 [cs.CL].

[12] Masaki Fujita et al. "Legal Textual Entailment Using Ensemble of Rule-Based and BERT-Based Method with Data Augmentation by Related Article Generation". In: *New Frontiers in Artificial Intelligence: JSAI-isAI 2022 Workshop, JURISIN 2022, Kyoto, Japan*. Springer. 2023, pp. 138–153.

[13] Kwan Yuen Iu and Vanessa Man-Yi Wong. *ChatGPT by OpenAI: The End of Litigation Lawyers? Available at SSRN: http://dx.doi.org/10.2139/ssrn.4339839*. 2023.

[14] Daniel Martin Katz et al. *Natural Language Processing in the Legal Domain*. 2023. arXiv: 2302.12039.

[15] Rupert Macey-Dare. *ChatGPT Generative AI Systems as Quasi-Expert Legal Advice Lawyers - Case Study Considering Potential Appeal Against Conviction of Tom Hayes*. 2023.

[16] Niklas Muennighoff et al. *Crosslingual Generalization through Multitask Finetuning*. 2022. arXiv: 2211.01786.

[17] John J. Nay. *Law Informs Code: A Legal Informatics Approach to Aligning Artificial Intelligence with Humans*. 2023. arXiv: 2209.13020 [cs.CY].

[18] Alejandro Peña et al. *Leveraging Large Language Models for Topic Classification in the Domain of Public Affairs*. 2023. arXiv: 2306.02864.

[19] Alec Radford et al. *Learning Transferable Visual Models From Natural Language Supervision*. 2021. arXiv: 2103.00020 [cs.CV].

[20] Colin Raffel et al. "Exploring the limits of transfer learning with a unified text-to-text transformer". In: *The Journal of Machine Learning Research* 21.1 (2020), pp. 5485–5551.

[21] Guilherme Moraes Rosa et al. "Billions of parameters are worth more than in-domain training data: A case study in the legal case entailment task". In: *arXiv preprint arXiv:2205.15172* (2022).

[22] Victor Sanh et al. "Multitask Prompted Training Enables Zero-Shot Task Generalization". In: *International Conference on Learning Representations*. 2022. URL: https://openreview.net/forum?id=9Vrb9D0WI4.

[23] Teven Le Scao et al. "Bloom: A 176b-parameter open-access multilingual language model". In: *arXiv preprint arXiv:2211.05100* (2022).

[24] Zhenwei Shao et al. *Prompting Large Language Models with Answer Heuristics for Knowledge-based Visual Question Answering*. 2023. arXiv: 2303.01903.

[25] Xiaofei Sun et al. *Text Classification via Large Language Models*. 2023. arXiv: 2305.08377 [cs.CL].

[26] Yi Tay et al. "Unifying language learning paradigms". In: *arXiv preprint arXiv:2205.05131* (2022).

[27] Ashish Vaswani et al. *Attention Is All You Need*. 2017. arXiv: 1706.03762 [cs.CL].

[28] Linting Xue et al. "mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer". In: *Proceedings of NAACL 2021: Human Language Technologies*. Online, June 2021, pp. 483–498. DOI: 10.18653/v1/2021.naacl-main.41.

[29] Arun Kumar Yadav et al. "Extractive text summarization using deep learning approach". In: *International Journal of Information Technology* 14.5 (2022), pp. 2407–2415.

[30] Masaharu Yoshioka, Youta Suzuki, and Yasuhiro Aoki. "HUKB at the COLIEE 2022 statute law task". In: *New Frontiers in Artificial Intelligence: JSAI-isAI 2022 Workshop, JURISIN 2022, Kyoto, Japan*. Springer. 2023, pp. 109–124.