An Investigation into the Understanding and Reasoning Capabilities of LLMs for Legal Statute Identification

Shounak Paul[1,*], Rohit Kumar Prajapati[1], Pawan Goyal[1] and Saptarshi Ghosh[1]

1Indian Institute of Technology, Kharagpur, 721302, India

Abstract

In this study, we analyze the performance of Large Language Models (LLMs) for the Legal Statute Identification (LSI) task – which entails identifying the relevant legal statutes (articles of law) given the fact description of a legal case or situation. We analyze three specific capabilities that are required for LSI – (i) fact understanding ability, (ii) pre-instilled legal knowledge, and (iii) legal reasoning ability – and especially focus on these capabilities of LLMs in disambiguating between semantically similar (or confusing) statutes. We conduct different experiments and analyses using two state-of-the-art LLMs – GPT-4o-mini and DeepSeek – over a standard LSI dataset over statutes from the Indian judiciary. Our experiments establish that modern LLMs are still limited in these abilities for a complex domain such as the legal domain. However, providing some assistance to the LLMs with regard to fact understanding or legal knowledge can improve LSI performance significantly.

Keywords

Legal Statute Identification, Large Language Models, Legal Fact Understanding, Legal Knowledge, Legal Reasoning

1. Introduction

The legal domain is one of the most crucial public services for any country, which provides a huge scope of intervention through Artificial Intelligence (AI) based methods. Recent advances in AI and Natural Language Processing (NLP), especially generative AI, has led to adoption by legal professionals, law students for conducting legal research and analyses, and even for lay persons in gaining access to legal knowledge. One of the first steps in the legal or judicial process in any country that follows a notion of civil law is to identify the relevant legal statutes (written articles of law) given the description of the facts of a legal case or situation. An automated *Legal Statute Identification* (LSI) system – that identifies the legal statutes relevant to the description of a situation – would not only assist legal professionals with instant statute recommendations, but is also a key requirement for any AI-based legal chatbot or assistant.

Although traditional LSI approaches have involved statistical/mathematical techniques [1, 2, 3], or classifier based approaches employing RNNs or transformers [4, 5, 6, 7, 8, 9, 10], recent advances in generative AI have prompted researchers to use powerful Large Language Models (LLMs) for statute prediction [11, 12]. These models have billions of parameters and can use their superior pre-trained knowledge and reasoning skills to identify relevant statutes without any LSI-specific fine-tuning. From this perspective, certain capabilities/skills of the LLM come into play, and determine the final statute prediction efficacy.

In this study, we particularly focus on three capabilities that are required for the task of LSI – (i) understanding the nuances of the events/actions described in the input facts, (ii) pre-instilled legal knowledge about the statutory texts involved, and (iii) legal reasoning ability, i.e., the ability to reason with both the aforementioned skills to decide on the set of relevant statutes. We conduct experiments

✉ shounakpaul95@kgpian.iitkgp.ac.in (S. Paul); rohit44774112@kgpian.iitkgp.ac.in (R. K. Prajapati); pawang@cse.iitkgp.ac.in (P. Goyal); saptarshi@cse.iitkgp.ac.in (S. Ghosh)

🌐 https://sites.google.com/view/shounakpaul95 (S. Paul); https://cse.iitkgp.ac.in/~pawang/ (P. Goyal); https://cse.iitkgp.ac.in/~saptarshi/ (S. Ghosh)

🆔 0000-0001-8124-4338 (S. Paul); 0000-0002-9414-8166 (P. Goyal); 0000-0002-2306-300X (S. Ghosh)

to investigate to what extent modern LLMs have these aforementioned capabilities. In particular, we investigate how well modern LLMs can utilize the above capabilities to *disambiguate statutes that are highly semantically similar (or, confusing)*. Prior research has shown that this aspect is critical in determining the overall LSI performance [8, 13].

We choose two popular LLMs – GPT-4o-mini and DeepSeek – compare their performance based on the above factors. We observe the performance of LLMs under three setups – completely unassisted, and providing some assistance with regard to fact understanding and legal knowledge respectively. We chose the ILSI dataset [14] for our experiments, which is large-scale LSI dataset for Indian criminal laws, where the statutes are frequently cited Sections from the Indian Penal Code. Specifically, in this study, we have worked with a subset of 150 documents targeting 36 statutes from the Indian Penal Code. This set of 36 statutes comprise many pairs of statutes which are known to be semantically confusing with each other (more details in Section 3).

Our experiments establish that LLMs already have appreciable capability in understanding and interpreting factual events, but relatively struggle more in recalling their pre-trained legal knowledge when performing the LSI task. They especially lack in reasoning capabilities over a complex domain such as the legal domain. However, we also show that providing some 'assistance' with regard to legal knowledge provides a significant boost to performance (for both LLMs). Whereas, providing assistance towards fact understanding also improves performance of one of the models.

2. Objectives

The objective of this study is to understand the capability of LLMs for the Legal Statute Identification (LSI) Task – which entails identifying the legal statutes/articles of law given the facts of a case or a legal situation. In this study, we particularly focus on the efficacy of the LLMs to disambiguate between semantically similar (or confusing) statutes which is known to be a core aspect of the LSI task [13, 8]. We consider the following skills/capabilities that are critical for the LSI task:

(i) **Fact Understanding Capability:** Given the description of the facts of a situation, the first step in LSI is to understand the sequence of events and actions that have been described. Nuances such as whether multiple people were involved in the crime, whether dangerous weapons were used, or whether any death occurred determine the statutes that will be applicable (from a set of confusing statutes). Thus, before identifying the statutes, the model needs to have a sound understanding of the events/actions that took place.

(ii) **Legal Knowledge:** To identify which statutes are relevant, the model must also have a concrete knowledge about all the statutes in the legal system. Specifically, the model needs to be able to recall from its memory the exact language of the statute descriptions, especially with regard to the confusing statutes. This is important because small differences in the language of two confusing statutes pertain to different events and thus determine which can be applicable.

(iii) **Legal Reasoning Capability:** The final step in LSI is to connect the events understood via the fact understanding capability, with the legal knowledge of the LLM, to reason about which of the confusing statutes is applicable for the current case. This reasoning capability needs to leverage all the understanding of the prior two capabilities, to be able to correctly correlate the events of the case with the exact statute to be applied.

The key to understanding the performance of LLMs on the complex task of LSI involves analyzing how they perform across these aspects of the task. In our experiments, we also provide assistance to the models to enhance their Fact Understanding capability and Legal Knowledge, and analyze the performance of the models vis-a-vis with and without the assistance.

Table 1

Some confusing pairs of IPC statutes, along with the disambiguating question(s) identified by legal experts and a brief explanation of which statute applies in which situation.

Statute 1	Statute 2	Disambiguating Questions
IPC 392 (Punishment for robbery)	IPC 397 (Robbery with attempt to cause death)	Was any deadly weapon used? (If yes, then IPC 397 also applies; otherwise IPC 392 only)
IPC 392 (Punishment for robbery)	IPC 394 (Causing hurt in robbery)	Was the victim actually hurt or injured? (If yes, then IPC 394 applies; otherwise IPC 392)
IPC 307 (Attempt to murder)	IPC 304 (Culpable homicide)	Did the victim actually die? Was there any clear intention to kill? (If the victim died, then IPC 304 applies; if the victim did not die but there was clear intention to kill, then IPC 307 applies)
IPC 323 (Causing hurt)	IPC 307 (Attempt to murder)	Was there an intention or knowledge that the act can cause death? (If yes, then IPC 307 applies; otherwise only IPC 323 applies)
IPC 365 (Kidnapping to wrongfully confine)	IPC 364 (Kidnapping to murder)	Was the intent to kill the person? (If yes, IPC 364 applies; else, IPC 365 applies)

3. Dataset

We choose the ILSI dataset [14] for our experiments. This is a large scale dataset (~66k samples) for the task of LSI over 100 Indian criminal laws. Specifically, the class labels / statutes are the most frequently-cited 100 Sections of the Indian Penal Code (IPC). The dataset is multi-label in nature, i.e., each sample (fact description) can be relevant to multiple statutes (see [14] for details of the dataset).

As described in Section 2, we wish to focus our efforts on the ability of LLMs to disambiguate confusing statutes. We consulted with legal experts, and asked them to identify some pairs of statutes (out of the 100 IPC target statutes) that can be semantically confusing. The experts came up with a set of 30 such pairs of confusing statutes, which span a total of 36 IPC sections. We also asked the experts to design questions that can be used to disambiguate between each pair. Some of these confusing statute pairs and the corresponding questions are listed in Table 1. Subsequently, we chose 150 documents from the ILSI test set at random, such that they cite at least one among the 36 aforementioned statutes. All experiments and analyses in this study are conducted over this subset of 150 documents, spanning a total of 36 statutes.

4. Experimental Setup

To analyze the capabilities of LLMs described in Section 2 pertaining to different aspects of the LSI task, we conduct experiments with two LLMs – GPT-4o-mini (https://platform.openai.com/docs/models/gpt-4o-mini) and DeepSeek (https://huggingface.co/deepseek-ai/DeepSeek-R1-Distill-Llama-70B). For both these models, we asked them to quote verbatim the full statute texts of all IPC sections, which they were able to do, demonstrating that both models have prior knowledge of all the IPC statutes that we wish to identify. We devise the following experimental setups.

4.1. Prompting with no assistance (NA)

In this approach, we present the model with case facts and directly ask it to identify relevant IPC sections without any additional context or assistance. In other words, in this approach, the model must rely entirely on its pre-trained knowledge to understand the legal context and make predictions. This technique serves as our baseline, revealing each model's inherent capability to perform LSI without specialized assistance. The prompt is provided in Table 2.

> **System Prompt:** You are a legal expert specializing in Indian Penal Code (IPC). Your task is to analyze case facts and identify the applicable IPC sections.
> **User Prompt:** I will provide you with case facts and ask you to identify which Indian Penal Code (IPC) sections apply.
> *Case facts:* <Facts of the case>
> Based on the case facts provided above, please determine the applicable IPC section(s) that best fit this case.

Table 2
Prompt used for the No Assistance (NA) Setup

4.2. Prompting with Legal Fact Understanding Assistance (FA)

Building on the basic approach, we incorporated specific questions designed by legal experts to help models distinguish between the top 30 frequently confused statute pairs (described in Section 3). For each case, we — (i) identified potentially confusing statute pairs relevant to the case, (ii) included targeted questions that highlight the key distinctions between these statutes, and then (iii) asked the LLM to answer these questions along with statute prediction, expecting that the LLM can consider the pertinent questions for the given case as well as the answers when identifying statutes (especially between confusing statutes). We base our idea on prior works that have found that identifying these critical distinguishing conditions in fact descriptions can lead to more efficient statute identification [5].

A critical component of our disambiguation techniques is determining which statute pairs are potentially confusing in a given case. We implemented two distinct approaches for this identification:

(i) **Ground Truth-Based Identification:** For each case in our dataset, we examined the ground truth labels (the statutes that actually apply) and cross-referenced them with our pre-identified list of 30 frequently confused statute pairs. When a statute from our ground truth appeared in one of these pairs, we flagged that pair as potentially relevant for disambiguation. This analysis provides an upper bound in performance for any pipeline that uses automated method of predicting initial statutes before disambiguation.

(ii) **Model Prediction-Based Identification:** For developing a fully automated pipeline, we used the predictions of the Longformer-LADAN model – which was seen to perform the best on the ILSI dataset in the prior work [13] – instead of the gold standard statutes. We then repeated the same process as the ground-truth based identification to mark the pair of confusing statutes based on the prediction of the aforementioned model.

After the potentially confusing statute pairs for a fact had been identified, we pick the corresponding disambiguating questions (for examples see Table 1) and provide them in the prompt. We ask the LLM to answer the questions as well as identify the relevant statutes, with the expectation that the question-answering process will guide the LLM in focusing on the key events and outcomes in the fact, and help it distinguish between the confusing statutes. The full prompt is provided in Table 3.

4.3. Prompting with Legal Knowledge Assistance (KA)

Similar to the FA setup, we first consider the top 30 confusing statute pairs and identify the potentially confusing statute pairs for each query (either based on ground truth or predictions of the Longformer-LADAN model). Then, we provide the statute names and text descriptions (in full) of all the confusing pairs. Unlike the FA step which tries to assist the LLM's reasoning capability toward understanding key events in the fact, the KA step rather tries to assist the LLM in recalling the full statute texts. As mentioned, in this setup also, we perform the analysis with both gold standard and model-predicted statutes. The prompt is described in Table 4.

> **System Prompt:** You are a legal expert specializing in Indian Penal Code (IPC). Your task is to analyze case facts and identify the applicable IPC sections.
> **User Prompt:** I will provide you with case facts and ask you to identify which Indian Penal Code (IPC) sections apply. Additionally, I'll ask some specific questions to help clarify between potentially confusing statute pairs.
> *Case Facts:* <Facts of the case>
> *To determine the applicable IPC sections, please consider the following questions for potentially confusing statute pairs:* <Question 1, Question 2, ...>
> Based on the case facts and your answers to these questions, please provide:
> 1. The applicable IPC section(s) that best fit this case.
> 2. Your answers to each of the questions listed above.

Table 3
Prompt used for the Fact Understanding Assistance (FA) Setup

> **System Prompt:** You are a legal expert specializing in Indian Penal Code (IPC). Your task is to analyze case facts and identify the applicable IPC sections.
> **User Prompt:** I will provide you with case facts and ask you to identify which Indian Penal Code (IPC) sections apply. Additionally, I'll provide the statute descriptions for some potentially confusing statutes to consider.
> *Case Facts:* <Facts of the case>
> *Relevant Statute Descriptions to consider:* <Text of Statute 1, Text of Statute 2, ...>
> Based on the case facts and the text of IPC sections provided above, please determine the applicable IPC section(s) that best fit this case.

Table 4
Prompt used for the Legal Knowledge Assistance (KA) Setup

4.4. Prompting with Fact Understanding and Knowledge Assistance (FA + KA)

In this setup, we provide the LLM with both Fact Understanding Assistance (FA) as well as Legal Knowledge Assistance (KA), thereby testing the legal reasoning skills of the LLM in isolation. For each pair of confusing statutes, we provide both the disambiguation questions as well as the full statute texts. We follow similar settings as independent FA and KA, including experiments with both gold-standard as well as predicted statutes. The prompt is described in Table 5.

> **System Prompt:** You are a legal expert specializing in Indian Penal Code (IPC). Your task is to analyze case facts and identify the applicable IPC sections.
> **User Prompt:** I will provide you with case facts and ask you to identify which Indian Penal Code (IPC) sections apply. Additionally, I'll ask some specific questions to help clarify between potentially confusing statute pairs, as well as provide the text descriptions of some relevant IPC sections.
> *Case Facts:* <Facts of the case>
> *To determine the applicable IPC sections, please consider the following questions for potentially confusing statute pairs:* <Question 1, Question 2, ...>
> *Also consider the text of relevant IPC sections:* <Text of Statute 1, Text of Statute 2, ...>
> Based on the case facts, the IPC sections provided above, and your answers to these questions, please provide:
> 1. The applicable IPC section(s) that best fit this case.
> 2. Your answers to each of the questions listed above.

Table 5
Prompt used for the Fact Understanding and Knowledge Assistance (FA+KA) Setup

4.5. Hyper-parameters and other Settings

For the all LLM prompting experiments, we used temperature of 0.0, utilized the maximum context length available for each model, and asked the models to return the outputs in a consistent JSON format.

Facts of the case: …The case of the prosecution is that it was the appellant [ENTITY] who knocked at the door of the house of the complainant carrying a battery in his hand and claimed to be an Electrician who was to note the meter reading. When the wife of the complainant opened the door, he along with three other persons entered the house and pressed her mouth. One of them put a knife on her neck and asked for the keys of the almirahs, threatening to kill her in case she raised alarm. One of them caught hold of her grand-daughter, pressed her mouth and put a knife on her neck, whereas the person who was carrying a pistol with him bolted the door from inside. Threatening the wife of the complainant, the aforesaid person had just entered the rooms inside when there was a knock at the door and the aforesaid persons came out and started running away. …
Gold-standard statutes: IPC 392 (Punishment for robbery), IPC 397 (Robbery or dacoity with an attempt to cause death or grievous hurt)

Table 6
A portion of a sample fact along with the gold-standard statutes from our dataset

For evaluating, we use macro-F1 scores by matching the model predictions to the actual statute names in our label set.

4.6. Demonstration of prompts in the three setups

The prompts used for NA, FA and KA and FA+KA setups are provided in Tables 2, 3, 4 and 5 respectively. To demonstrate clearly how we used the different setups, consider the example in Table 6. This fact describes a situation where multiple people, armed with weapons, entered a house to commit robbery. Now, among the gold-standard statutes, we have IPC 392 (Punishment for robbery) and IPC 397 (Robbery or dacoity with an attempt to cause death or grievous hurt). From Table 1, we know the IPC 392 and IPC 397 are ambiguous with each other. Additionally, IPC 392 is ambiguous with IPC 394 (Causing hurt in robbery).

In the FA setup prompt (Table 3), we provide the corresponding questions "Was any deadly weapon used?" (to distinguish between IPC 392 and IPC 397) and "Was the victim actually hurt or injured?" (for IPC 392 and IPC 394) to the LLM (refer to Table 1). Similarly, in the KA setup prompt (Table 4), we provide the complete texts of IPC 392, IPC 394 and IPC 397 to the LLM, and expect that the complete statutory texts can help the model to disambiguate between the confusing statutes.

5. Results and Analysis

The results of all experiments are reported in Table 7. We have considered macro-averaged Precision, Recall and F1 scores to measure model performance, in line with prior works [14, 13].

Firstly, we observe that across all settings, GPT-4o-mini outperforms DeepSeek, across all three metrics – Precision, Recall and F1. For GPT, all types of assistance improves performance compared to the baseline setup (NA), with the improvement being greater for KA compared to FA. However, for DeepSeek, only KA is able to improve the performance, while FA actually leads to deteriorating performance. The combined assistance (FA+KA) setup is sensitive to the initial set of statutes. While using gold-standard statutes lead to significant improvements, using the predicted statutes do not have the same effect. Among all experiments, GPT in the FA+KA setup with gold-standard statutes shows the best performance of 30.22% F1 (77% improvement over the baseline NA setup with 17.07% F1). If we consider fully automated pipelines, then GPT under the KA setup performs the best with 21.75% F1 (27% improvement over baseline NA setup).

5.1. Impact of Fact Understanding Assistance

The FA setup provides insight into the LLMs' capabilities of understanding the events described in the facts. We observe that this setup provides some boost to the final F1 performance of GPT but fails to do the same for DeepSeek.

Table 7
Results of different experimental setups on the LSI Task. All values are macro-averaged and in terms of percentages. Best values of each metric for each LLM are marked in boldface. The best values among fully automated pipelines (not relying on the ground truth) are underlined.

Experimental Setup	GPT-4o-mini			DeepSeek		
	mP	mR	mF1	mP	mR	mF1
Baseline (Fully automated)						
No Assistance (NA)	22.5	15.87	17.07	22.56	12.54	14.81
Using Ground Truth Statutes						
Fact Understanding Assistance (FA)	29.31	20.26	22.3	15.92	10.76	12.84
Legal Knowledge Assistance (KA)	36.45	24.24	26.54	23.50	**15.9**	**17.9**
Fact and Knowledge Assistance (FA + KA)	**43.11**	**27.58**	**30.22**	**29.30**	11.30	15.09
Using Predicted Statutes (Fully automated)						
Fact Understanding Assistance (FA)	24.41	17.72	19.04	15.2	11.0	12.5
Legal Knowledge Assistance (KA)	<u>30.61</u>	<u>19.51</u>	<u>21.75</u>	20.8	<u>13.4</u>	<u>15.2</u>
Fact and Knowledge Assistance (FA + KA)	27.38	15.51	18.58	<u>26.10</u>	9.19	12.0

To further analyze the gap in performance, we manually annotated a set of 60 documents (out of 150) with yes/no answers to the all the disambiguating questions. We used these gold-standard answers to evaluate a LLM's ability to correctly answer these questions under the FA setup. It is interesting to note that GPT obtained a 72.7% accuracy in answering the questions, while DeepSeek obtained 77.7% accuracy. This shows that DeepSeek actually has a better factual understanding of the case facts. However, this translates to a lower F1 score for statute prediction compared to GPT (in fact, for DeepSeek, FA setup reduces performance compared to the NA setup), demonstrating that DeepSeek is not effectively able to reason with the answers to the questions, while GPT can. In fact, including the factual questions in the prompt is leading to further confusion for the DeepSeek model, rather than assisting it to better identify the relevant statutes.

5.2. Impact of Legal Knowledge Assistance

Compared to the indirect strategy employed for the FA setup, the KA setup employs a more direct approach by asking the LLMs to directly disambiguate between the confusing statutes by providing their entire text. This setup seems to be more beneficial for both GPT and DeepSeek, resulting in higher gains. We can perhaps thus conclude that LLMs inherently have strong fact understanding capabilities, but struggle in recalling the necessary legal knowledge, i.e., the full statute texts. It should also be pointed out that the disambiguating questions designed for the FA setup are created by carefully considering the difference in full texts of confusing statute pairs. In the KA setup, by directly providing the full statute texts, we incentivize the LLMs to use their own reasoning ability to correlate the fact events with the conditions described in the statute texts. This seems to work more in favour of the LLMs toward identifying the relevant statutes.

5.3. Impact of Combined Assistance

The impact of combined (FA+KA) assistance does not follow a singular trend across models and setups. In most cases, FA+KA improves over the baseline NA setup, validating the efficacy of assistance setups. When using the gold-standard statutes, there is significant improvement over the FA setup as well, for both models. However, while the combined FA+KA setup performs the best for GPT, in the case of

DeepSeek the performance decreases compared to KA. The disambiguating questions might actually cause greater confusion for DeepSeek, as observed with the independent FA setup as well. When considering the predicted statutes, the performance of the combined setup is inferior to both FA and KA. It is possible that the overload of information (both questions and statute texts) is detrimental to the performance of the model when the information is based on the predicted statutes, which could differ significantly from the gold-standard.

6. Conclusion

In this study, we have tried to understand how different inherent capabilities of LLMs determine their performance in the Legal Statute Identification task. We focused our efforts in analyzing these capabilities with regard to how well LLMs can effectively disambiguate confusing statutes. Our experimental results broadly show that providing legal knowledge assistance to an LLM is more effective compared to providing fact-understanding assistance. This is true for both the LLMs tested, namely GPT-4o-mini and DeepSeek. We postulate that the direct approach of providing the statute texts to the model helps it invoke all its reasoning capabilities to better determine the final statute set, as compared to the indirect method of asking disambiguating questions and nudging the LLM in the direction of those questions. Despite having a good success rate in terms of answering the factual understanding questions, the DeepSeek model was not able to reason effectively which event leads to which statute being applied. Rather, providing the direct texts of the statutes allow both the models to fully exploit their reasoning capabilities. Finally, using a combined assistance setup has mixed results, working better for GPT compared to DeepSeek. This setup also underperforms when using the predicted statutes, hinting that the models are relying heavily on the assistance. Since the predicted statutes are not as accurate as the gold-standard statutes, the combined assistance FA+KA might be misleading the model as compared to independent assistance setups FA and KA.

In future, we would like to experiment with a larger number of model types, ranging across different sizes in terms of number of parameters. We also want to experiment with more confusing statutes, as well as datasets from other jurisdictions.

Declaration on Generative AI

The core methodology of this research involved the use of GPT-4o-mini (via API access) and DeepSeek (via HuggingFace) for comparing the reasoning capabilities of the said LLMs for the task of legal statute identification, as described in the Paper (see Section 4). No generative AI tools were employed for the writing or editing of this manuscript beyond the research itself. The authors take full responsibility for the content of this paper.

References

[1] D. M. Katz, M. J. Bommarito, II, J. Blackman, A general approach for predicting the behavior of the supreme court of the united states, PLOS ONE (2017). doi:`10.1371/journal.pone.0174698`.

[2] W.-C. Lin, T.-T. Kuo, T.-J. Chang, C.-A. Yen, C.-J. Chen, S.-d. Lin, Exploiting machine learning models for Chinese legal documents labeling, case classification, and sentencing prediction, in: Proceedings of the 24th Conference on Computational Linguistics and Speech Processing (ROCLING 2012), 2012.

[3] Y.-H. Liu, Y.-L. Chen, W.-L. Ho, Predicting associated statutes for legal problems, Information Processing & Management (2015). doi:`https://doi.org/10.1016/j.ipm.2014.07.003`.

[4] B. Luo, Y. Feng, J. Xu, X. Zhang, D. Zhao, Learning to predict charges for criminal cases with legal basis, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017. doi:10.18653/v1/D17-1289.

[5] Z. Hu, X. Li, C. Tu, Z. Liu, M. Sun, Few-shot charge prediction with discriminative legal attributes, in: Proceedings of the 27th International Conference on Computational Linguistics, 2018.

[6] P. Wang, Z. Yang, S. Niu, Y. Zhang, L. Zhang, S. Niu, Modeling dynamic pairwise attention for crime classification over legal articles, in: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 2018. doi:10.1145/3209978.3210057.

[7] P. Wang, Y. Fan, S. Niu, Z. Yang, Y. Zhang, J. Guo, Hierarchical matching network for crime classification, in: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019. doi:10.1145/3331184.3331223.

[8] N. Xu, P. Wang, L. Chen, L. Pan, X. Wang, J. Zhao, Distinguish confusing law articles for legal judgment prediction, in: D. Jurafsky, J. Chai, N. Schluter, J. Tetreault (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 3086–3095. URL: https://aclanthology.org/2020.acl-main.280/. doi:10.18653/v1/2020.acl-main.280.

[9] I. Chalkidis, I. Androutsopoulos, N. Aletras, Neural legal judgment prediction in English, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019. doi:10.18653/v1/P19-1424.

[10] Y. Le, Y. Zhao, M. Chen, Z. Quan, X. He, K. Li, Legal charge prediction via bilinear attention network, in: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, 2022. doi:10.1145/3511808.3557379.

[11] I. Chalkidis, Chatgpt may pass the bar exam soon, but has a long way to go for the lexglue benchmark, 2023. arXiv:2304.12202.

[12] D. Bernsohn, G. Semo, Y. Vazana, G. Hayat, B. Hagag, J. Niklaus, R. Saha, K. Truskovskyi, Legal-Lens: Leveraging LLMs for legal violation identification in unstructured text, in: Y. Graham, M. Purver (Eds.), Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, St. Julian's, Malta, 2024, pp. 2129–2145. URL: https://aclanthology.org/2024.eacl-long.130/.

[13] S. Paul, R. Bhatt, P. Goyal, S. Ghosh, Legal statute identification: A case study using state-of-the-art datasets and methods, in: Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '24, Association for Computing Machinery, New York, NY, USA, 2024, p. 2231–2240. URL: https://doi.org/10.1145/3626772.3657879. doi:10.1145/3626772.3657879.

[14] S. Paul, P. Goyal, S. Ghosh, Lesicin: A heterogeneous graph-based approach for automatic legal statute identification from indian legal documents, volume 36, 2022, pp. 11139–11146. URL: https://ojs.aaai.org/index.php/AAAI/article/view/21363. doi:10.1609/aaai.v36i10.21363.