TOWARDS THE EXPLOITATION OF LLM-BASED CHATBOT FOR PROVIDING LEGAL SUPPORT TO PALESTINIAN COOPERATIVES

Rabee Al-Qasem
Data scientist
Palestine
Rabee.qasem93@gmail.com

Banan Tantour
Legal Advisor
Palestine
banan.tantour@hotmail.com

Mohammed Maree
Department of Information Technology,
Faculty of Engineering and Information Technology,
Arab American Unievsrity, Jenin, Palestine
mohammed.maree@aaup.edu

June 12, 2023

ABSTRACT

With the ever-increasing utilization of natural language processing (NLP), we started to witness over the past few years a significant transformation in our interaction with legal texts. This technology has advanced the analysis and enhanced the understanding of complex legal terminology and contexts. The development of recent large language models (LLMs), particularly ChatGPT, has also introduced a revolutionary contribution to the way that legal texts can be processed and comprehended. In this paper, we present our work on a cooperative-legal question-answering LLM-based chatbot, where we developed a set of legal questions about Palestinian cooperatives, associated with their regulations and compared the auto-generated answers by the chatbot to their correspondences that are designed by a legal expert. To evaluate the proposed chatbot, we have used 50 queries generated by the legal expert and compared the answers produced by the chart to their relevance judgments. Finding demonstrated that an overall accuracy rate of 82% has been achieved when answering the queries, while exhibiting an F1 score equivalent to 79%.

Keywords Large Language Models · Artificial Intelligence · Chatbots · NLP · Experimental Evaluation · Legal Text

1 Introduction

Natural Language Processing (NLP) has revolutionized the way we interact with legal texts. It has made it easier to analyze and comprehend complex legal texts (Becerra, 2018; Dale, 2019). One of the most recent significant advancements in this field is the development of Large Language Models, and the development of chatbots that are based on such models (Brown et al., 2020; Rae et al., 2021), were ChatGPT is at the forefront of this development (Lee et al., 2022). With its vast training data and powerful capabilities, ChatGPT has had a profound impact on global users. It provides them with intelligent conversational agents capable of understanding and responding to their queries. The integration of LLMs-powered chatbots has extended beyond the legal domain, finding applications in various fields. However, it is in the realm of legal discourse where these chatbots truly shine (Omar et al., 2023). They leverage their expertise to assist users in navigating complex legal terms and processes (Brooks et al., 2020; Fang et al., 2023; Rajendra and S Thuraisingam, 2022).

The huge improvement in LLM-based chatbot technology and the ease of integrating it seamlessly in the context of the legal domain has encouraged us to build a chatbot to provide answers to legal inquiries and questions about Palestinian

cooperative law. We noticed that there have been numerous inquiries from cooperative societies and cooperative unions regarding private legal issues. This is mainly because the law is relatively new, having been issued at the end of 2017 (OGB, 2017) . Additionally, there is an urgent need to provide legal answers at all times, especially considering the need for a labor-intensive effort to answer such queries. Furthermore, considering the large number of cooperative members, which reached 58,883 at the end of 2021 as reported in (Cooperative Work Agency, 2021), there is an urgent need for a chatbot that is available 24/7 to address their legal inquiries and provide timely assistance. The rest of those article is organized as follows. In Section 2, we review the literature and discuss the related works. Section 3, introduces the utilized dataset for testing and evaluating our proposed chatbot. In Section 4, we discuss the proposed methodology. Section 5 present the experimental evaluation and results. In Section 6, we conclude our work and point to the future directions of our research work in Section 7.

2 Literature review

The use of machine learning (ML) techniques in the legal domain has long history with it a lot of research has integrated the two domains in many fields such as Legal document review (Mahoney et al., 2019; Wei et al., 2018), Legal prediction (Sil and Roy, 2020), Legal writing (Phelps and Ashley, 2022) , legal summarization Elaraby and Litman (2022), and Legal compliance (Mandal et al., 2017). Prompting can be used to improve the performance of LLMs in different criteria and explore the effectiveness of using prompts in legal judgment prediction (LJP). (Trautmann et al., 2022) conduct experiments using data from the European Court of Human Rights and the Federal Supreme Court of Switzerland, comparing different prompts with multilingual language models (LLMs) such as mGPT, GPT-J-6B, and GPT-NeoX-20B. The results demonstrate that zero-shot prompt engineering can improve LJP performance with LLMs, yielding better macro-averaged F1 scores, precision, and recall compared to simple baselines. However, the performance of zero-shot learning still falls short of current supervised state-of-the-art results in the field. The paper also highlights the following key findings: prompting can enhance LLM performance in legal judgment prediction, multilingual LLMs can be effective even with training data in a single language, and while zero-shot learning holds promise, further improvements are needed to achieve state-of-the-art outcomes. The authors conclude by emphasizing the potential value of prompting for legal professionals and the accessibility benefits of multilingual LLMs in the field of legal natural language processing (NLP). In an experiment, (Pettinato Oltz, 2023) built a fictitious law professor who had a normal week of duties including teaching and community service planned out for her. then they used ChatGPT prompts for each task to test how well the system worked. For six of the seven tasks given, ChatGPT was able to produce workable first drafts in just 23 minutes. The most common tasks, including making a practice exam question or preparing a class handout, showed ChatGPT to be the most proficient. ChatGPT struggled with more complex tasks, especially those that had to do with education, although it still had the potential to save time in some cases. The experiment's findings indicate that ChatGPT, especially service-related jobs, has a lot of promise for reducing some components of the workload for law faculty. Additionally, ChatGPT may enable law professors to spend less time on specific teaching responsibilities, giving up more time for them to concentrate on pedagogy and create innovative teaching strategies. Finally, (Queudot et al., 2020) design and implementation of two immigration chatbots to advise their users about immigration legal questions and cases. One answers immigration-related questions, and the other answers legal questions from NBC employees. Both chatbots use supervised learning to learn embeddings for their answers.

3 Dataset

In our research work, we used 5 resources to acquire the input for our chatbot. To do this, we used three official documents which are Law No. 20 of 2017 on Cooperatives, Cooperatives Bylaws, and Housing Cooperatives Bylaws. Also, we created two sets of questions and answers datasets which we will discuss more in section 3.1

3.1 Formal Legal Documents

In order to give the Chatbot the legal context that it needs to answer legal questions, we needed to provide it with the legal documents that the lawyers and Legal advisors depend on and use to answer legal questions but we made some reformatting for these documents where we only kept the necessary articles and definitions these legal documents are:

1. Law No. 20 of 2017 on Cooperatives: published in 2017 to govern the cooperative work in Palestine, under which the authority supervising the cooperative work sector in Palestine was established, known as the Cooperative Work Agency (CWA). It also specially regulates cooperative societies and unions. cooperative members and the local community.

2. Cooperatives Bylaws: It is the bylaws that govern the cooperative and the union, which regulate their work and the nature of their activity, based on the provisions of Law Decree No. 20 of 2017.

3. Housing Cooperatives Bylaws: It is the bylaws that govern the housing cooperatives, which regulate their work and the nature of their activity, based on the provisions of Law Decree No. 20 of 2017.

3.2 Question and Answers Dataset

In order to support the chatbot to better understand the legal questions we created two other data set which contains a JSON file of questions about the Decree – Law No. (20) Of 2017 On Cooperative cooperatives the two data set are as following:

3.2.1 Human Generated Question Answer Dataset

We asked the legal advisor on cooperatives to create a dataset containing 40 questions and answers about different articles from the Decree – Law No. (20) Of 2017 On Cooperative. The questions and answers cover the basic topics of the definition of a cooperative, the requirements for forming a cooperative, the rights and responsibilities of cooperative members, and the role of the CWA in regulating cooperatives.

3.2.2 Chatgpt Generated Question and Answers

We used the ChatGPT API to generate 5 questions and their corresponding answers for each article of Law No. (20) of 2017 on Cooperative. However, we needed to customize the answers to simulate the response of a real legal advisor. This involved starting the answer by referring to the article number in the law. To achieve this, we utilized the following prompt structure, as shown in Figure 1: we first requested the generation of the question and answer, then provided the article itself, and finally, to control the output, we asked ChatGPT to create a dictionary with two keys: "question" and "answer". After ChatGPT generated the dictionary, we appended it to another dictionary to collect the data. This process resulted in 350 questions and their corresponding answers.

Figure 1: prompt Structure
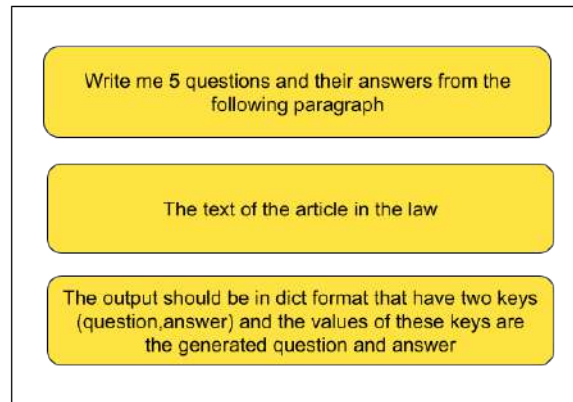
This prompt helped us to control the output of the ChatGPT as we can see from the following code snippet

```
{"question":"ما هي الشروط التي يجب توافرها للجمع في العضوية بين جمعيتين تعاونيتين من
ذات التخصص الواحد في نفس المحافظة؟"ٍ
"answer":– لا يجوز الجمع في العضوية بين جمعيتين تعاونيتين من ذات التخصص الواحد في
"نفس المحافظة
}
```

4 Methodology

In our work, we encountered a vast amount of textual data that exceeded ChatGPT GPT-4's current processing limit at 8,192 tokens (White et al., 2023). In order to take advantage of the ChatGPT API and overcome this obstacle, we utilized a comprehensive solution by employing LlamaIndex(Liu, 2023). This proved to be a strategic decision as it enabled us to index large-scale datasets quickly and efficiently through its tailored features created specifically for Language Models (LLMs). The provided tools excel in generating vectors for every document while keeping them readily available.

To make the text data compatible with the ChatGPT API, we employed LlamaIndex to create an index encompassing all the legal documents and question-answer data at hand. Subsequently, we generated vectors for each document, ensuring that the input size did not exceed 8,192 tokens, while employing a chunk size of 600 tokens. The chosen chunk size of 600 tokens aligned with the requirements of the LLM. Moreover, we configured the maximum chunk overlap to be 50 tokens.

We efficiently stored the generated vectors within the index, enabling their swift retrieval whenever necessary. Leveraging the LlamaIndex query engine, which harnessed the power of ChatGPT in the background, we successfully addressed our legal queries and concerns. Figure 2 represents the comprehensive pipeline that we implemented for our case study, clearly demonstrating the use of LlamaIndex with ChatGPT and the subsequent vector generation and indexing of the legal documents.

Figure 2: ChatGPT-based Chatbot workflow
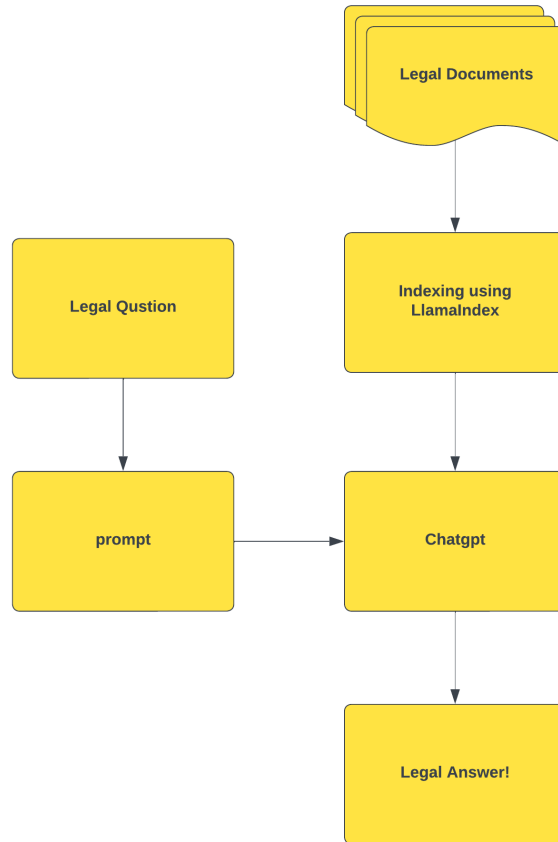
5 Experimental Setup and Evaluation Results

After building the chatbot, we asked the legal advisor to write another 50 questions and their answers for testing purpuses. We then used the written questions to test the chatbot and compared the chatbot's answers to the legal

advisor's answers. The chatbot was able to answer 41 questions in general. For example, we asked the chatbot about membership, financial statements, administrative issues, and how to register a new cooperative. The chatbot not only answered the right answers but in some cases, it also cited the law number and article (see Table 1.)

Table 1: Right answers provided by the chatbot

Question	Legal advisor answer	Chatbot	satisfaction score
هل يمكن للهيئة العمومية الموافقة على استمرار عضو لجنة الإدارة لأكثر من (6) سنوات متتالية؟	نعم يجوز إذا وافقت الهيئة العمومية على بأغلبية ثلثي الأعضاء.	الإجابة: نعم، يمكن للهيئة العمومية الموافقة على استمرار عضو لجنة الإدارة لأكثر من (6) سنوات متتالية إذا وافقت بأغلبية ثلثي الأعضاء.	100
ماذا يحدث إذا انقضت المدة دون اتخاذ قرار بشأن طلب التسجيل؟	تسجيل الجمعية بحكم القانون، حيث يتوجب على المجلس أن يصدر قراره بالتسجيل أو الرفض، مع بيان الأسباب خلال شهر من تاريخ استكمال الطلب، وإذا انقضت هذه المدة دون اتخاذ قرار، تعتبر الجمعية مسجلة بحكم القانون	يجب على المجلس الادارة ان يرد على طلب التسجيل خلال شهر من تاريخ استكمال الطلب. إذا انقضت المدة دون اتخاذ قرار بشأن طلب التسجيل، فستعتبر الجمعية مسجلة بحكم القانون.	100
ما هي المدة التي يجب على لجنة الإدارة أن ترسل فيها نسخة من وقائع الاجتماعات؟	خلال (15) يوماً من تاريخ عقد الاجتماع	المدة المحددة في قرار بقانون رقم (20) لسنة 2017 بشأن لجنة الإدارة لترسيل نسخة من وقائع الاجتماعات في مدة أقصاها 15 يوم.	100

Although the chatbot was able to answer 41 questions correctly, not all of them were answered directly. Eight out of the 41 answers were relevant, but not direct. For example, when we asked the chatbot in Arabic "When does the management committee meet in the cooperative?", the chatbot gave us the correct answers, but it combined two answers. The first answer was the meeting of the management committee, and the second answer was the meeting of the general assembly. We have a lot of cases like this, and this is due to not having enough questions and answers for each article. Some articles of the law are short, so 5 questions and answers were enough to give the chatbot the context of the article. However, some articles are long and needed more questions and answers. 5 questions and answers were not enough to give the chatbot the context it needs to understand the difference, for example, between the meeting of the management committee and the general assembly. But in general, it gave us the right answer (See table 2).

Table 2: Related answers provided by the chatbot

Question	Legal advisor answer	Chatbot	satisfaction score
متى يجب ان يتم تدقيق دفاتر حسابات الجمعية	يجب أن تدقق دفاتر حسابات كل جمعية مسجلة مرة واحدة على الأقل في السنة، وفقاً للأصول المحاسبية	يجب أن يتم تدقيق دفاتر حسابات الجمعية في نهاية السنة المالية بعد تغطية العجز واقتطاع المصروفات واستهلاك الاصول، وفقا لنصوص قرار بقانون رقم (20) لسنة 2017 بشأن الجمعيات التعاونية والأنظمة الصادرة بمقتضاه.	65%
متى تجتمع لجنة الادارة في الجمعية.	تجتمع لجنة الإدارة مرة كل شهرين على الأقل	الهيئة العمومية يجب أن تجتمع على الأقل مرة واحدة في السنة، وذلك خلال المدة المحددة في قرار بقانون رقم (20) لسنة 2017 بشأن الجمعيات التعاونية. وعلى لجنة إدارة الجمعية ان تجتمع مرة كل شهرين في السنة	85%
من هي الهيئة العمومية في الجمعيات التعاونية؟	هي جميع الأعضاء المساهمين في رأس مال الجمعية أو الاتحاد القطاعي أو الاتحاد.	الهيئة العمومية في الجمعيات التعاونية هي الهيئة التي تشمل الأعضاء الذين يشتركون في الجمعية والذين يحق لهم التصويت في الاجتماعات العمومية للجمعية.	70%

Finally, when we analyzed the rest of the wrong answers, we found that most of them were due to two reasons. First, there were not enough questions and answers for long articles, which required more explanation for the chatbot. Second, some articles had bylaws that needed to be provided to the chatbot. For example, when we asked the chatbot "Is it permissible to establish more than one general union?", which is illegal, the chatbot answered yes, which is a totally illegal act. (See table. 3)

5

Table 3: Wrong answers provided by the chatbot

Question	Legal advisor answer	Chatbot	satisfaction score
هل يجوز تأسيس أكثر من اتحاد عام؟	لا يجوز تسجيل أكثر من اتحادا تعاوني عام واحد في فلسطين	نعم، يجوز تأسيس أكثر من اتحاد عام في فلسطين بناءان على قانون الجمعيات التعاونية الفلسطيني رقم 20 لسنة 2017. يحتوي قانون الجمعيات التعاونية على تعليمات وأحكام يجب أن يتم الالتزام بها عند تأسيس أي جمعية تعاونية أو اتحاد عام.	0%
ما هي الحالة التي يتم فيها تولي اللجنة المؤقتة إدارة الجمعية؟	في حالة استنكاف أو شغور أكثر من ثلث مقاعد لجنة الإدارة، أو ظهور مخالفات بعد التحقيق تستدعي اتخاذ هذا الإجراء، أو وجود تنازع بين أعضاء اللجنة بصورة تعرقل تحقيق أغراض الجمعية، وتضر بمصالحها	بناءً على قانون الجمعيات التعاونية الفلسطيني رقم 20 لسنة 2017، يتم تولي اللجنة المؤقتة إدارة الجمعية في الحالة التي يكون فيها غياب الرئيس.	0%
ما يجب على لجنة الإدارة فعله إذا شغر مقعد أو أكثر فيها؟	مادة (36) إذا شغر مقعد أو أكثر في لجنة الإدارة، فعلى لجنة الإدارة أن تشغل المقعد أو المقاعد الخالية من الأعضاء الاحتياط.	لجنة الإدارة يجب عليها أن تنتخب نائباً للرئيس إذا كانت طبيعة عملها تحتاج ذلك ويتولى مهام الرئيس في حالة غيابه، وفقا للمادة 37 من قانون الجمعيات التعاونية الفلسطيني رقم 20 لسنة 2017.	0%

To measure the performance of our chatbot, we used the following metrics:

Overall accuracy: This metric is calculated by dividing the total number of correct answers by the total number of questions asked. The equation for overall accuracy is:

$$\text{Overall accuracy} = \frac{\text{Total number of correct answers}}{\text{Total number of questions}} \times 100 \tag{1}$$

In this case, the chatbot achieved an overall accuracy of 41/50, or 82%. This is a good result, as it means that the model was able to correctly answer 82% of the questions asked. and if we provided more data for it it will increase its accuracy

Overall satisfaction: Many studies used satisfaction scores with other metrics to evaluate their trained chatbots (Casas et al., 2020), but in our case, we used only the satisfaction score. We did this by letting the legal counsel give a mark for how satisfied they were with the answer. For example, in the case of a right answer, the legal counsel was very satisfied, so she gave it a score of 100%. For wrong answers, the score was 0%. For related answers, the score was between 60% and 85%. We calculated this by computing the average of all satisfaction scores and the total number of questions.

$$\text{Average satisfaction score} = \frac{\sum_{i=1}^{n} S_i}{n} \tag{2}$$

where Si is the satisfaction score for the ith question and n is the total number of questions.

In this case, the chatbot achieved an average satisfaction 78.3%. Which is also a good result for our chatbot.

Confusion matrix: To measure the performance of the chatbot, we used the precision, recall, and F1 score. The confusion matrix is used to measure classification models and use the actual value and the predicted value of the model to compute the precision and recall and then the F1 score. However, since we didn't train the chatbot, we assumed that all the answers of the legal counselor are correct. This assumption will affect the precision value, as the chatbot will not be penalized for incorrectly identifying an answer as wrong.

1. **Precision:** Since we made the assumption that there are no wrong actual answers the precision for class (wrong) is 0, and the precision for class 1 (right + related) is 1.0, indicating that our chatbot correctly predicted all instances as "right" or "related."

2. **Recall:** The recall for class (wrong) is 0, indicating that our chatbot did not correctly identify any instances as "wrong." The recall for class (right/related) is 0.79, meaning that our chatbot correctly identified 79% of instances labeled as "right" or "related."

3. **F1-score**: The F1-score for class (wrong) is 0, which aligns with the precision and recall being 0. The F1-score for class (right/related) is 0.88, indicating a relatively good balance between precision and recall for this class.

4. The **accuracy** of our chatbot is reported as 0.79, meaning it correctly predicted the label for 79% of the instances in the questions that we provided.

   For more information and details on the developed chatbot, please refer to our GitHub Repository at the following link: Github

6 Conclusion

In this paper, we introduced our LLM-based legal chatobot that aims to assist Palestinian cooperatives and their members in finding relevant answers to their legal inquiries. Our objective was to provide accurate and reliable support 24/7 by leveraging the vast amount of publicly-available legal documents that we were able to acquire. After utilizing the chatbot on this extensive dataset, we achieved an overall accuracy of 82% and an F1 score of 79%.

However, as we encountered an enormous volume of text data, we faced challenges with the chatbot's processing limit in terms of the maximum amount of text data that can submitted to through ChatGPT's API. To overcome this obstacle, we implemented a technique called 'vectorization' using LlamaIndex. This process converted the text data into a format that the chatbot could effectively utilize.

Whilst serving as a valuable legal aid for diverse cooperative members, there are certain limitations inherent to the chatbot that should be acknowledged. Our study uncovered instances where the chatbot provided incorrect answers, which could potentially lead users to unintentionally violate legal regulations. Consequently, we believe that continuous development and improvement of the chatbot are necessary to enhance its accuracy and reliability. Furthermore, it is crucial to be transparent about the chatbot's limitations and ensure that users have access to comprehensive information. This will enable them to make informed decisions about utilizing the chatbot's services. By refining the chatbot and openly communicating its capabilities, we can harness its potential as an invaluable tool for delivering reliable legal support to a diverse audience.

7 Future Works

In the future work, we plan to address the challenges highlighted in the previous section through the following steps. First, we will focus on increasing the size of the used dataset by formalizing additional question, with their relevance judgments, and also ensuring that the expert questions are close in terms of their number to those generated by the chatbot. Second, we plan to post-process the answers produced bu the chatbot to further enhance and improve the overall quality, i.e. accuracy of the answers. This may also require the exploitation of legal domain knowledge and semantic resources that can be further utilized for reformulating users' questions in a more legal relevant context.

References

Becerra, S. D. (2018). The rise of artificial intelligence in the legal field: where we are and where we are going. *J. Bus. Entrepreneurship & L.*, 11:27.

Brooks, C., Gherhes, C., and Vorley, T. (2020). Artificial intelligence in the legal sector: pressures and challenges of transformation. *Cambridge Journal of Regions, Economy and Society*, 13(1):135–152.

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. *Advances in neural information processing systems*, 33:1877–1901.

Casas, J., Tricot, M.-O., Abou Khaled, O., Mugellini, E., and Cudré-Mauroux, P. (2020). Trends & methods in chatbot evaluation. In *Companion Publication of the 2020 International Conference on Multimodal Interaction*, pages 280–286.

Cooperative Work Agency (2021). Annual Report. https://cwa.pna.ps/publications/201. Cooperative Work Agency.

Dale, R. (2019). Law and word order: Nlp in legal tech. *Natural Language Engineering*, 25(1):211–217.

Elaraby, M. and Litman, D. (2022). Arglegalsumm: Improving abstractive summarization of legal documents with argument mining. *arXiv preprint arXiv:2209.01650*.

Fang, C., Wilkenfeld, J. N., Navick, N., and Gibbs, J. L. (2023). "ai am here to represent you": Understanding how institutional logics shape attitudes toward intelligent technologies in legal work. *Management Communication Quarterly*, page 08933189231158282.

Lee, M., Srivastava, M., Hardy, A., Thickstun, J., Durmus, E., Paranjape, A., Gerard-Ursin, I., Li, X. L., Ladhak, F., Rong, F., et al. (2022). Evaluating human-language model interaction. *arXiv preprint arXiv:2212.09746*.

Liu, J. (2023). Llamaindex (gpt index) is a data framework for your llm applications.

Mahoney, C. J., Zhang, J., Huber-Fliflet, N., Gronvall, P., and Zhao, H. (2019). A framework for explainable text classification in legal document review. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 1858–1867. IEEE.

Mandal, S., Gandhi, R., and Siy, H. (2017). Modular norm models: A lightweight approach for modeling and reasoning about legal compliance. In *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*, pages 657–662. IEEE.

OGB (2017). Official gazette bureau. https://lab.pna.ps/ar/Category/20. 138.

Omar, R., Mangukiya, O., Kalnis, P., and Mansour, E. (2023). Chatgpt versus traditional question answering for knowledge graphs: Current status and future directions towards knowledge graph chatbots. *arXiv preprint arXiv:2302.06466*.

Pettinato Oltz, T. (2023). Chatgpt, professor of law. *Professor of Law (February 4, 2023)*.

Phelps, T. and Ashley, K. (2022). " alexa, write a memo": The promise and challenges of ai and legal writing. *Legal Writing: J. Legal Writing Inst.*, 26:329.

Queudot, M., Charton, É., and Meurs, M.-J. (2020). Improving access to justice with legal chatbots. *Stats*, 3(3):356–375.

Rae, J. W., Borgeaud, S., Cai, T., Millican, K., Hoffmann, J., Song, F., Aslanides, J., Henderson, S., Ring, R., Young, S., et al. (2021). Scaling language models: Methods, analysis & insights from training gopher. *arXiv preprint arXiv:2112.11446*.

Rajendra, J. B. and S Thuraisingam, A. (2022). Artificial intelligence and its impact on the legal fraternity. *UUM Journal of Legal Studies (UUMJLS)*, 13(1):129–161.

Sil, R. and Roy, A. (2020). A novel approach on argument based legal prediction model using machine learning. In *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, pages 487–490. IEEE.

Trautmann, D., Petrova, A., and Schilder, F. (2022). Legal prompt engineering for multilingual legal judgement prediction. *arXiv preprint arXiv:2212.02199*.

Wei, F., Qin, H., Ye, S., and Zhao, H. (2018). Empirical study of deep learning for text classification in legal document review. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 3317–3320. IEEE.

White, J., Hays, S., Fu, Q., Spencer-Smith, J., and Schmidt, D. C. (2023). Chatgpt prompt patterns for improving code quality, refactoring, requirements elicitation, and software design. *arXiv preprint arXiv:2303.07839*.