Prompt engineering and provision of context in domain specific use of GPT

Marton RIBARY [a,1], Paul KRAUSE [b], Miklos ORBAN [b,c], Eugenio VACCARI [a], and Thomas WOOD [d]

[a] *Royal Holloway, University of London, UK*
[b] *University of Surrey, UK*
[c] *OPL gunnercooke*
[d] *Fast Data Science*

Abstract. Large Language Models (LLMs) can appear to generate expert advice on legal matters. However, at closer analysis, some of the advice provided has proven unsound or erroneous. We tested LLMs' performance in the procedural and technical area of insolvency law in which our team has relevant expertise. This paper demonstrates that statistically more accurate results to evaluation questions come from a design which adds a curated knowledge base to produce quality responses when querying LLMs. We implemented a triage system using the natural language processing (NLP) techniques of pattern matching and zero-shot classification based on sentence embeddings in order to process a user query and identify relevant statutes, case law, and HMRC forms. The information was then passed with the user query to gpt-3.5-turbo or gpt-4 in a prompt (prompt engineering) with the intention to improve the response accuracy. We evaluated our bot head-to-head on an unseen test set of twelve questions about insolvency law against the unmodified versions of gpt-3.5-turbo and gpt-4 with a mark scheme similar to those used in examinations in law schools. On the "unseen test set", the insolvency bot based on gpt-3.5-turbo outperformed gpt-3.5-turbo ($p = 1.8\%$), and our gpt-4 based bot outperformed unmodified gpt-4 ($p = 0.05\%$). These promising results can be expanded to cross-jurisdictional queries and be further improved by matching on-point legal information to user queries. Overall, they demonstrate the importance of incorporating trusted knowledge sources into traditional LLMs in answering domain-specific queries.

Keywords. legal tech; LLMs (GPT); prompt engineering; NLP; insolvency law (England); chatbot

1. Introduction

Conversational Large Language Models (LLMs), such as ChatGPT, have generated significant interest in various domains for tasks ranging from giving medical assessments through generating computer code to providing expert advice on legal matters. ChatGPT and the gpt-4 model have demonstrated some significant success in the legal field, in particular when it passed the multistate part of the US bar exam, but according to practitioners, real life legal cases tend to be more complicated than the bar .[1] In addition,

[1] Corresponding Author: Marton Ribary, marton.ribary@rhul.ac.uk.

at closer analysis, some of the legal advice provided by such systems have proven to be unsound, erroneous, and sometimes even absurd. While these systems can seem to replicate the response of an expert in language that is confident and compelling, it is important to keep in mind that the purpose of the text corpora used to train the models is to provide examples of the usage of natural language. Because these items of text inevitably concern specific knowledge domains, the responses of LLMs can appear to demonstrate expert knowledge. However, this is a side-effect of the generation of the LLMs; they have not been developed as knowledge elicitation tools. In this paper, we explore methods by which an LLM can be enhanced to provide a trusted knowledge source with a certain level of professional expertise. Specifically, our goal is to support the triage of potential legal cases for stakeholders involved in insolvency issues for micro, small and medium enterprises (MSMEs) with a level of competency comparable to a Level 6 or 7 Law Student. This is a specific area of law where many solo practitioners and smaller law firms lack sufficient legal expertise, so our system could - if successful enough - provide a helping hand to such practitioners in expanding the scope of their services. Specifically, in this paper we evaluate the hypothesis that query responses from an LLM will be improved if the model is enhanced with a trusted domain specific knowledge base.

2. The Insolvency Context

Micro-, Small- and Medium-Sized Enterprises (MSMEs) are the backbone of modern economies. The COVID-19 pandemic, the evolution of consumer demand, rising costs of debt and the implementation of new technologies have increased insolvency risks for these enterprises. However, traditional insolvency procedures can be overly expensive, complex, long and – ultimately – ineffective for MSMEs, therefore some countries such as the US [2,3], Italy [4, s. IV.], Ireland [5] and Australia [6] have introduced simplified insolvency regimes for MSMEs, while many other countries continue to treat insolvent or financially distressed MSMEs in the same way as they do large corporations.

The UK is one of the few common law countries not to have introduced MSME-specific rules (besides those applicable to people and individual entrepreneurs on the discharge of debt). This may prove to be an unfortunate policy choice as opportunities to rescue distressed yet viable businesses may be lost. The UK's approach also sits at odds with main international trends and recommendations such as the Report on the Treatment of MSME Insolvency, published by the World Bank in 2017, and the European Union's proposal for a directive on Harmonising Certain Aspects of Insolvency Law. This is not to say, however, that the UK's system is hopelessly ill-equipped to deal with MSMEs in distress in an efficient and effective manner. As evidenced elsewhere,[4, s. VI.C.] English law offers a sufficiently flexible and modular [7,8] approach to corporate restructuring. However, there are downsides. Filing costs and legal fees have been, and still are, one of the main barriers preventing cash-stripped debtors from using restructuring procedures. Additionally, no special legal measures have been introduced to promote more timely filings or to ensure that management are more accountable in the way they operate their companies. Such innovations are especially timely because company insolvencies are rising at an increasingly fast pace. The latest statistics show that the number of company insolvencies in Q2 2023 was the highest since Q2 2009, 9% higher than in Q1 2023 and 13% higher than in Q2 2022.[9]

3. Related Works

The release of OpenAI's ChatGPT bot in November 2022 caused an explosion of interest in the potential of Large Language Models (LLMs). The perceived threat to long-held practices in the creative and educational sector was reverberated in the legal sector which has been contemplating threat and potential at equal measure.[10] Riding the public interest, city law firms such as Allen & Overy and Mischcon de Reya announced investments in GPT-based solutions. The reception was probably more sedate in the academic legal tech community which has been experimenting with LLMs for a while with full awareness of the limitations of such models.

As we describe in detail in section 4, our system has two main layers: matching user prompts with relevant legal content called from a structured knowledge base in the triaging layer, and creating a structured prompt to query an LLM in the prompt engineering layer. For the former, submissions to the University of Alberta's annual Competition on Legal Information Extraction/Entailment event (COLIEE) provide an overview of recent trends. In the pre-ChatGPT period of COLIEE 2021 and 2022 [11,12], solutions leveraging LLMs for identifying relevant case or statute law (or a section thereof) for a given input predominantly used BERT-based models. [13] took inspiration from COLIEE 2021 and used their 2021 and 2022 datasets to query gpt-3 for solving legal problems, that is, sitting the Japanese Bar exam. Similar to our system in its prompt engineering layer, [13] solicited gpt-3 to "think step by step" with chain-of-thought prompts.[14] They also used iterative cycles to improve the model output after feeding back its own response.[15] We also experimented with this approach, but noticed that it might only feed noise and throw the model off the task. Similar to our approach, [13] created a hypothesis from the initial prompt which was then fed back for True/False testing. An additional component of [13] was fine-tuning prompts with specific legal reasoning techniques to generate best results, notably the IRAC structure of issue, rule, application, conclusion, something we may look at in the future.

In COLIEE 2023, [16] used a topic-based approach to filter and rank candidate cases for match, [17] used term extraction, vectorization (TF-IDF), ranking (with BM25), year filtering and post-processing, while [18] used simple text similarity measure optimising gpt-4 by OpenAI with structured prompts and FLaN-T5 by Google/HuggingFace with instruction tuning. The latter solution noted the difficulty we encountered in processing both user prompts and legal content, that is, the use of open-textured terms.

The triaging layer of our legal advice tool is most similar to the system by the THUIR team [19] which won the case law retrieval task in COLIEE 2023. THUIR's system incorporated structural knowledge of case law into pre-trained language models which included removing "useless information" and including reference sentences in the pre-processing stage, looking for similar cases based on vector representations (TF-IDF), harnessing the structure of a case (facts $\Rightarrow$ reasoning $\Rightarrow$ decision), and filtering and limiting matches in the post-processing stage. Our work suggests that creating a curated knowledge base successfully supplements these technically sophisticated information retrieval and entailment matching systems for producing quality outputs when querying LLMs.

4. Design

4.1. NLP and Prompt Engineering

The Insolvency Bot (https://fastdatascience.com/insolvency) is written in Python 3.10 [20] and deployed as an API using Microsoft Azure Functions [21], with a simple HTML and Javascript-based front end. The system receives an input query from the user, and a combination of a rule-based keyword matching algorithm, and zero-shot classification [22], is used to identify relevant cases, statutes, and HMRC forms from a domain specific knowledge base. The zero-shot learning uses OpenAI's text-embedding-ada-002 model [23] to convert the query into a sentence embedding vector.[24] The vector is compared to the database of vectors, and the closest vectors (using cosine distance) are chosen.

The domain specific knowledge base discussed in section 4.2 consists of around 6,000 texts which have been converted offline using text-embedding-ada-002 to vector embeddings. Each text, such as "apply for an extension to a moratorium", has been mapped to all relevant sections of statute, HMRC forms, and case law. When a user's query comes in, it is split into sentences. Each sentence is converted to an embedding using text-embedding-ada-002 and the relevant statute sections, forms and cases are retrieved. These are assembled into a prompt which is passed to OpenAI's gpt-4 model.

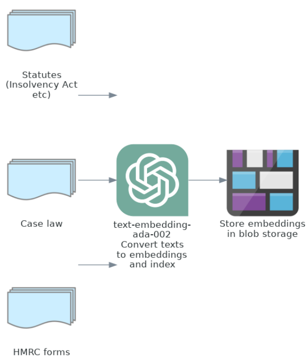

Figure 1. Workflow diagram of how the knowledge base is indexed

The resulting prompt contains the closest matching statutes, case law, and forms, and finally the user's query. gpt-4 is instructed to answer as an insolvency lawyer in England and Wales, taking into account relevant statute and case law, such as the Insolvency Act 1986. The response from gpt-4 is returned to the user. The bot has been deployed to the web using static HTML page, a Javascript front end, an Azure Functions back end. Users can try it on the Fast Data Science website.

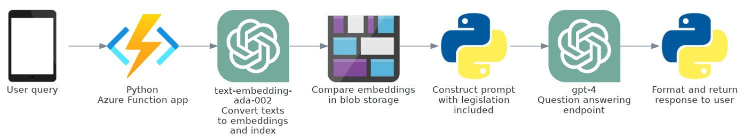

Figure 2. Workflow diagram of the Insolvency Bot when in use

4.2. Knowledge Base

4.2.1. Statute Law

We ingested the entire text of the statutes, excluding appendices, into our knowledge base. Each section of each of the below statutes was included as one row in our structured knowledge base:

- Insolvency Act 1986
- Company Directors Disqualification Act 1986
- Companies Act 2006
- The Insolvency (England and Wales) Rules 2016
- The Cross-Border Insolvency Regulations 2006

4.2.2. HMRC Forms

We made use of two lists of .pdf forms for company owners on HMRC's website, forms for insolvency rules and forms for limited companies. Each form has a little text in the top left corner identifying the relevant section of statute. There are also instructions such as "Use this form to confirm that the company details are up to date." We manually created a table of three columns: form name (e.g. CS01), form instructions (e.g. "Use this form to confirm that the company details are up to date"), and legislation that the form cites (e.g. "In accordance with Section 853A of the Companies Act 2006").

4.2.3. Case Law

We created a custom database of 198 insolvency related cases based on the *English corporate insolvency law* primer by Eugenio Vaccari and Emilie Ghio.[25] We collected information about the cases from the FindCaseLaw service of The National Archives (FCL), Westlaw UK (WL-UK) and the Insolvency Lawyers' Association (ILAUK). We extracted references to sections of statutes from case law (full text as well as summary) which we used as a proxy for identifying the topic discussed within. For the purpose of linking user queries with the relevant case law, we assigned keywords to cases in plain English, but we also recorded the keywords assigned by WL-UK. When it was available, we recorded the summary of the case as found on ILAUK and WL-UK. For select cases, we created our own summary of the case recording its basic facts, the decision reached by the court, and the often quoted sentences from the judgment itself. The knowledge base includes more information than the Insolvency Bot currently uses. As some of this information is proprietary, this part of our work is not made public in the project repository.[26]

5. Methodology

The system described in section 4 was used to evaluate our research hypothesis according to the following methodology. For the purpose of testing and fine-tuning the Insolvency Bot, we relied on user queries on corporate insolvency law matters related to small businesses as posted on the "Legal, Employment and Insolvency" section of the UK Business Forum platform. We took all sixty queries posted between 27 January 2023 and 4 March 2023 and identified twelve of them related to the topic of insolvency to some extent. These queries formed the basis of our experiments in the developmental stage of our Insolvency Bot. As will be discussed later, a further twelve queries were prepared that remained unseen by the model developers until the final testing of the model.

An early prototype (Mickey-bot) [27] enriched queries with a set of prompt engineering instructions related to clarity of the initial prompt, jurisdiction, style ("act as a solicitor providing legal advice") and reasoning method (turning the question into a hypothesis and chain-of-thought). We fed the queries as they were into ChatGPT and Mickey-bot, and saw a promising competitive edge of the latter. While some ChatGPT issues persisted (e.g. hallucinations, false confidence), the most important shortcoming was the lack of legal authority in the provided response. This prototype provided valuable insights that informed the development of our system that is described in section 4, with the main enhancement being the domain specific knowledge base.

For final testing, a new set of twelve queries was prepared by an experienced academic specialising in corporate insolvency and bankruptcy law. The academic had no involvement with the development of the system. A mark scheme was also developed to assess responses to these queries and score the responses at a level commensurate with a Level 6 or 7 UK law student. This mark scheme included a mix of questions (between 7 and 10) to assess the ability of ChatGPT and the Insolvency Bot to provide accurate answers to twelve original queries. Each mark scheme had a total output of approximately 25 points, and the questions were weighted depending on their importance. For instance, much like either a traditional exam question or a legal opinion to a client, omission of key information and/or the provision of unsound or incomplete legal advice was deemed more penalising (in terms of scoring) than not referring to the applicable statute or the binding precedent in the area. Thus, we were able to assess versions of "raw-GPT" and the our system as if they were university level exam candidates.

We used the mark scheme to evaluate our system head-to-head against "raw GPT" answers. We ran the unmodified query on gpt-3.5 turbo and gpt-4 models without the knowledge base and prompt engineering architecture of the Insolvency Bot, and then we ran the same query on the Insolvency Bot using gpt-3.5 turbo and gpt-4 as the underlying LLM. Here is an example of our mark scheme for the following user input.

> I left a position a couple of months ago after a disagreement with the other director on the direction of the company. I signed a stock transfer form a month ago to return the 5% I had in the business and assumed it was dealt with after I was told it would be filed and I would be removed. Today, the other director dropped the keys for the offices rented by the business and a note saying I was never taken off of the business and to sell all of the assets, close down the business and pay off the debts (along with sending the other director some money). It came a bit out of left field and I'm fairly new to business but I assumed when I signed that form it removed all of my permissions from doing anything with company assets or filings? Any advice would help as I currently don't know what to do other than send the keys back recorded delivery so I cannot be accused of theft or anything.

We assessed the output of (1) raw gpt-3.5 turbo, (2) raw gpt-4, (3) the Insolvency Bot wrapping around gpt-3.5 turbo, and (4) the Insolvency Bot wrapping around gpt-4. Here is an example of the custom questions in our marking scheme related to the above example.

1. Does the lawyer mention the key features of a director? (max 3 points) $\Rightarrow$ bot got 3 points ✓

2. Does the lawyer refer to sections 154-155 CA 2006? (max 2 points) ⇒ bot got 0 points ✗
3. Does the lawyer identify the criteria for being a director under the law? (max 3 points) ⇒ bot got 0 points ✗

The evaluation itself was also automated. We fed the four different outputs along with the simple yes-no question to gpt-4, and parsed the generated answer for words like "yes", "no", or "however", to work out whether to give the bot 0%, 50%, or 100% of the points available for that question. In this way, gpt-4 was simulating a human examiner.

To create the test mark scheme, we ran the test questions through all four bots, shuffled the responses, and passed them to the domain expert on our team, who was then able to generate a mark scheme. The full mark scheme can be accessed in our project's GitHub repository [28] and on Zenodo.[26] The Insolvency Bot available on the Fast Data Science website runs on gpt-4. All test questions and bot answers are available on GitHub.

6. Results

One component of our system is the use of zero-shot classification and some keyword matching to recognise which cases are relevant for the user's question. We evaluated the precision and recall of this component and found that on the training questions, our system identified the correct cases with 49% precision and 57% recall. On the test questions, the bot performed slightly worse, with 24% precision and 33% recall.

We report the results of all four bots (gpt-3.5-turbo, gpt-4, and the Insolvency Bot based on gpt-3.5-turbo and gpt-4) for all questions in the test dataset. We used a two-sided paired t-test to compare each GPT variant with and without the Insolvency Bot. The average score of gpt-3.5-turbo on the test questions was 20% and that of the Insolvency Bot modification of gpt-3.5-turbo was 29%. This difference was significant t(-4.322) = 0.0012, $p < .05$. The average score of gpt-4 on the training questions was 21% and that of the Insolvency Bot modification of gpt-4 was 47%. This difference was also significant t(-4.832) = 0.00053, $p < .05$. We report the results of these experiments in Table 1 and Figure 3.

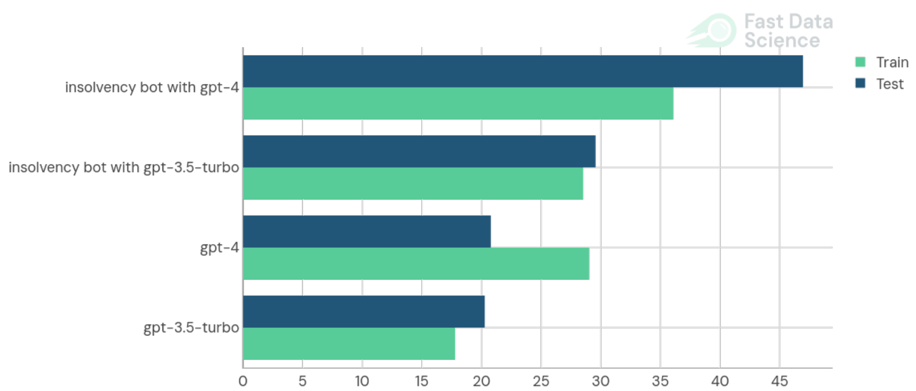

Figure 3. Average percentage score of the four outputs according to the mark scheme

Table 1. Scores of unmodified GPT bots and those enhanced by the Insolvency Bot (IB) according to our marking scheme

Question no.	Points available	gpt-3.5-turbo	gpt-4	IB (gpt-3.5-turbo)	IB (gpt-4)
Q1	25	6	12	12.5	15.5
Q2	24	3	3	3	4.5
Q3	25	3	3	3	10
Q4	25	3	3	9	5
Q5	22	3	3	3	12
Q6	25	6	6	9	14
Q7	25	3	6.5	1.5	9.5
Q8	25	11	3	16.5	19.5
Q9	25	3	3	6	15
Q10	25	11	11	16	19.5
Q11	25	3	3	3	4.5
Q12	25	5	5	5	10
Total	**296**	**60**	**61.5**	**87.5**	**139**
Percent	**-**	**20%**	**21%**	**30%**	**47%**

7. Discussion

Whilst LLMs have received significant publicity as appearing to show capability to emulate humans in the precision of their responses, there has been little in the way of formal evaluation of this. Our results show that although GPT is improving in its capability, it is still not at a level where it could pass a university level examination in a specific domain - insolvency law in our study. However, by enhancing gpt-4 with a significant level of additional domain specific knowledge, we have been able to use it to answer queries at a level that would pass an exam, with a statistically significant improvement in capability over the unmodified LLM.

It should be noted, however, that this is not a strong pass and so further work is needed in order to move our system up to a level where it can demonstrate professional-level competence. One of the key issues we found was that it was especially difficult to match user queries to case law. Since GPT and the vector embedding are good at matching semantically similar texts from the same domain, we found that queries from the internet are very different from case law transcripts. What helped with the case law triage was the step of including our training questions in the zero-shot classification process, so that an incoming query which is similar to any of our initial training questions could be triaged to the relevant case law from those questions. We believe that by improving the accuracy of this triage, we will be able to generate even more informative prompts to submit to the LLM, and hence significantly increase the score of the returned responses.

These results are not unexpected. LLMs are trained to be conversational, and not for information retrieval. Hence our need to enhance the LLM with access to relevant statutes, institutions and case law. In addition, addressing the information retrieval challenge of matching queries in the language of a lay person to the pertinent legal information was a further necessary enhancement. Our architecture of using queries to initially identify relevant information from a professionally curated domain specific resource in order to generate a prompt to be sent to a LLM could provide a widely applicable approach to producing trustworthy responses from LLMs. In addition, the assessment of capability by testing the resulting system against a professional-level exam sets the direction for formal validation of AI systems.

8. Conclusion

We have tested the performance of LLMs with a legal-specific prompt engineering tool, then presented a system that uses a curated knowledge base to improve the performance of LLMs in answering insolvency queries. Our system outperforms both the prompt engineering tool and the unmodified LLMs on an unseen test set of 12 questions, and it has the potential to be expanded to other jurisdictions and cross-jurisdictional queries.

Insolvency law is a fairly stable area of law, where legislative changes are rare, thus it may be more challenging to implement such a system in areas of law which are subject to more rapid changes in legislation, such as immigration law.

Our next steps are to expand our system to other legal areas such as employment law. We believe that it is a promising route to greater access to justice if we could provide AI-powered tools to citizens as well as legal practitioners. We will also explore the possibility of registering as an alternative legal services provider with the Solicitors Regulation Authority. This would allow us to provide advice in a regulated environment in the UK.

References

[1] Katz DM, Bommarito MJ, Gao S, Arredondo P. GPT-4 passes the Bar Exam. SSRN. 2023. Available from: `https://ssrn.com/abstract=4389233`.

[2] Norton III WL, Bailey JB. The pros and cons of the Small Business Reorganization Act of 2019. Emory Bankruptcy Developments Journal. 2020;36(2):383-93. Available from: `https://scholarlycommons.law.emory.edu/ebdj/vol36/iss2/2`.

[3] Walters A. The Small Business Reorganization Act: America's new tool for SME restructuring for the COVID and post-COVID era. The Company Lawyer. 2020;(10):324-5.

[4] Vaccari E, Ehmke D, Burigo F. MSMEs in Distress: Regulatory Costs and Efficiency Considerations in the Implementation of Preventive Restructuring Mechanisms: An Anglo-German-Italian Perspective. Journal of International and Comparative Law. 2023. Accepted for publication.

[5] Hutchinson GB. The Small Companies Rescue Act – false hope for failing companies? Company Law Practice? 2021;(7).

[6] Corporations Amendment (Corporate Insolvency Reforms) Act 2020 (Cth) (Act) (Australia); 2020. Available from: `http://classic.austlii.edu.au/au/legis/cth/num_reg/cairr2020202001654694/`.

[7] Mokal RJ, Davis R, Madaus S, Mazzoni A, Mevorach I, Romaine B, et al. Micro, small, and medium enterprise insolvency: A modular approach. Oxford: Oxford University Press; 2018.

[8] Vaccari E. A Modular Approach to Restructuring and Insolvency Law: Executory Contracts and Onerous Property in England and Italy. Norton Journal of Bankruptcy Law and Practice. 2022;(5).

[9] National Statistics. Company Insolvency Statistics: April to June 2023; 2023. Available from: `https://www.gov.uk/government/statistics/company-insolvency-statistics-april-to-june-2023`.

[10] Thomson Reuters Institute. ChatGPT and Generative AI within Law Firms Law firms see potential, eye practical use cases and more knowledge around risks; 2023. Available from: `https://www.thomsonreuters.com/en-us/posts/wp-content/uploads/sites/20/2023/04/2023-Chat-GPT-Generative-AI-in-Law-Firms.pdf`.

[11] Rabelo J, Goebel R, Kano Y, Kim MY, Satoh K, Yoshioka M. Overview and Discussion of the Competition on Legal Information Extraction/Entailment (COLIEE) 2021. The Review of Socionetwork Strategies. 2022;16:111-33. Available from: `https://doi.org/10.1007/s12626-022-00105-z`.

[12] Kim MY, Rabelo J, Goebel R, Kano Y, Satoh K, Yoshioka M. COLIEE 2022 Summary: Methods for Legal Document Retrieval and Entailment. In: Takama Y, Yada K, Satoh K, Arai S, editors. New Frontiers in Artificial Intelligence. JSAI-isAI 2022. No. 13859 in Lecture Notes in Computer Science. Springer; 2023. p. 51-67. Available from: https://doi.org/10.1007/978-3-031-29168-5_4.

[13] Yu F, Quartey L, Schilder F. Legal prompting: Teaching a Language Model to think like a lawyer. arXiv. 2022. Available from: https://doi.org/10.48550/arXiv.2212.01326.

[14] Kojima T, Shane Gu S, Reid M, Matsuo Y, Iwasawa Y. Large Language Models are Zero-Shot Reasoners. arXiv. 2023. Available from: https://arxiv.org/abs/2205.11916v4.

[15] Zelikman E, Wu Y, Mu J, Goodman ND. STaR: Bootstrapping Reasoning with Reasoning. arXiv. 2022. Available from: https://arxiv.org/abs/2203.14465v2.

[16] Novaes LP, Vianna D, da Silva A. A Topic-Based Approach for the Legal Case Retrieval Task. In: Rabelo J, Goebel R, Kano Y, Kim MY, Satoh K, Yoshioka M, editors. Proceedings of the Tenth International Competition on Legal Information Extraction/Entailment (COLIEE 2023) in association with the 19th International Conference on Artificial Intelligence and Law; 2023. p. 27-31.

[17] Debbarma R, Prawar P, Chakraborty A, Bedathur S. IITDLI: Legal Case Retrieval Based on Lexical Models. In: Rabelo J, Goebel R, Kano Y, Kim MY, Satoh K, Yoshioka M, editors. Proceedings of the Tenth International Competition on Legal Information Extraction/Entailment (COLIEE 2023) in association with the 19th International Conference on Artificial Intelligence and Law; 2023. p. 40-7.

[18] Bilgin O, Fields L, Laverghetta AJ, Marji Z, Nighojkar A, Steinle S, et al. AMHR Lab 2023 COLIEE Competition Approach. In: Rabelo J, Goebel R, Kano Y, Kim MY, Satoh K, Yoshioka M, editors. Proceedings of the Tenth International Competition on Legal Information Extraction/Entailment (COLIEE 2023) in association with the 19th International Conference on Artificial Intelligence and Law; 2023. p. 77-86.

[19] Liu Y, Li H, Su W, Wang C, Wu Y, Ai Q. THUIR@COLIEE 2023: Incorporating Structural Knowledge into Pre-trained Language Models for Legal Case Retrieval. In: Rabelo J, Goebel R, Kano Y, Kim MY, Satoh K, Yoshioka M, editors. Proceedings of the Tenth International Competition on Legal Information Extraction/Entailment (COLIEE 2023) in association with the 19th International Conference on Artificial Intelligence and Law; 2023. p. 1-6.

[20] Van Rossum G, Drake FL. Python 3 Reference Manual: (Python Documentation Manual Part 2). Documentation for Python. Scotts Valley, CA: CreateSpace; 2009.

[21] Microsoft. Azure Functions; 2023. Computer software. Available from: https://azure.microsoft.com/en-gb/products/functions.

[22] Xian Y, Akata Z, Sharma G, Nguyen Q, Hein M, Schiele B. Latent embeddings for zero-shot classification. 2016. Available from: https://doi.org/10.48550/arXiv.1603.08895.

[23] OpenAI. New and improved embedding model; 2022. Blog post. Available from: https://openai.com/blog/new-and-improved-embedding-model.

[24] Reimers N, Gurevych I. Sentence-BERT: Sentence embeddings using siamese BERT-networks. In: Inui K, Jiang J, Ng V, Wan X, editors. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics; 2019. p. 3982–3992. Available from: https://aclanthology.org/D19-1410.

[25] Vaccari E, Ghio E. English corporate insolvency law: A primer. Cheltenham: Edward Elgar; 2022.

[26] Wood T. Evaluation script for insolvency bot. Zenodo; 2023. Dataset. Available from: https://doi.org/10.5281/zenodo.8292105.

[27] Miklos O. Mickey-bot; 2023. Computer software. Available from: https://play.google.com/store/apps/details?id=com.omui.lgprompt.

[28] Wood T. Evaluate insolvency. GitHub; 2023. Code repository. Available from: https://github.com/fastdatascience/evaluate_insolvency.