

Lunch and Learn
Cyberjustice Laboratory
Wednesday, February 8, 2023.

Large Language Models: Applications in the Legal Field

Hannes Westermann and Jinzhe Tan

Introduction

Large language models (LLMs) are an important new development in the field of machine learning. They are likely to have a big influence on society as their performance continues to increase. In this Lunch & Learn, we explored how large language models can be applied in the legal field.

What are large language models?

Typically, machine learning algorithms are trained on a dataset of labelled examples. The resulting model can then be used to predict or classify new, unseen examples. To train machine learning algorithms, we need a large amount of data, which can sometimes make it challenging to apply it in a new domain.

Language models are trained to predict the next word in a sequence. To train these algorithms, we can use millions of textual documents available on the internet. By learning to predict the next word, these models can thus learn sophisticated patterns in language. Large language models can be used in different tasks that require language understanding (e.g., text generation, text classification, etc.). Thanks to pretrained embeddings, such as BERT, it is possible to use large language models for different tasks without the need for a lot of additional data.

Some models, such as GPT-3, rely on prompting. An initial text sequence is fed to the model, which then prompts the model to generate some other text, that hopefully corresponds to the task one wishes to accomplish. For example, to summarize text, one might give the model the text itself, and then write “To summarize,”. The model will aim to generate the next words in the sequence, which should correspond to a summary of the previous sections. With prompting, we don’t need to train the algorithms on new data, but we need to give prompts that induce the model to carry out the tasks we want it to. Drafting the appropriate prompt is thus an important part of using such models.

ChatGPT, the model that has attracted the most attention recently, is easier to prompt. ChatGPT is trained on human feedback to give helpful answers to questions. Thus, asking it to do something is often sufficient to get it to perform many different tasks. ChatGPT further includes filters to increase the safety of the outputs.

Potential use cases

Large language models have a number of important application areas, including:

- LLMs can be used to enhance search results to be more helpful and specific (as currently explored by both Microsoft and Google)
- LLMs can be used to assist programmers, by suggesting the next lines of code and explaining code.
- LLMs can be used in Research (e.g. to generate research idea, summarize literature, etc.)
- LLMs can be used for content creation (e.g., to draft documents, summarize documents, etc.)
- LLMs can be used to create highly responsive chatbots that can tackle many different tasks.
- Domain Knowledge Based LLMs can help with providing information, for example in science with BiotGPT-Large, and in philosophy with Philosopher GPT.
- LLMs can power new applications that are able to leverage natural language, i.e. to classify files. Different frameworks have been created to enable the integration of models in a toolchain, such as LangChain and GPT Index.

Potential risks of large language models

Of course, there are also a few potential risks in using LLMs. It is important to understand these risks in using the models.

Such risks include:

- Bias – The language that LLMs are trained on may include bias, that could then be reflected in the output of the model.
- Hallucination – Sometimes, language models give answers that seem correct but contain factual errors.
- Ethical issues – If not properly filtered, LLMs could be asked for how to conduct criminal activities.
- Transparency and Explainability – It may be difficult to understand why LLMs give specific answers in certain cases.

Concrete steps in using, training, deploying, and evaluating LLM

Training LLMs internally could be very expensive. Companies such as OpenAI offer a paid API that can be used to interact with their sophisticated language models, such as GPT-3. This API can be used to explore the models, and in production. Of course, one has to be cognizant of the costs involved.

LLM in the legal field

LLMs have many potential applications in the legal field. We explored a few such applications.

Using LLM to provide legal information

The JusticeBot project, which aims to build tools to give legal information to laypeople, uses legal guided pathways. Creating such legal pathways can be costly and time consuming, so it could be interesting to use language models to create such pathways. We performed some initial experiments by asking a GPT model to turn a piece of legislation into a structured representation, with promising results. The output of such a model could be given to a human annotator to verify, which could mitigate risks of an inaccurate understanding by the model.

Another way of using LLM in the context of Justicebot is to help users interact with legal pathways. Potentially, the user could describe their situation in natural language, and JusticeBot + GPT could explain whether their situation may fulfill certain criteria or not. We conducted some initial experiments on using GPT models in conjunction with the JusticeBot in this way, which showed the potential for LLMs to support users in this way. Of course, this is a research project, and many important questions would need to be answered before such a model could be used in practice.

Another way of using ChatGPT for Justicebot could be to automatize the summarization of legal decisions. Predictice, a legal tech company in France, has started to use GPT so summarize French court decisions.

Some may consider asking GPT models legal questions directly. A test in this direction has been conducted by Linklaters. They found that the answers were impressive for a LLM, but not great for a lawyer, as they e.g. sometimes contained factual errors. This shows the need to be careful in how LLMs are used in the legal domain.

Using LLMs to generate legal documents

There are many types of legal documents, but most types have a fixed format, and the same legal document is written in a relatively consistent manner. Analyzing and preparing these documents is an important task in the legal domain. Potentially, LLMs could be used to distinguish each section of legal documents (e.g., facts, issues, ratio decidendi, conclusion, etc.) and to extract information from them.

Some language models have already proven efficient in summarizing 15-page PDFs and providing the key takeaways of the document. An example of such an approach is filechat.io, which can answer questions about a specific document. We tested this feature by uploading a 99 pages PDF with legal information on filechat.io. Then, we asked the question “What is access to justice?” to test the chatbot. The answer provided by filechat.io corresponded well to the information in the document.

Using LLMs in Online Dispute Resolution (ODR)

We also explored whether GPT models could support an Online Dispute Resolution platform.

This could be done, for example, by asking the model to guide the user toward a specific procedure in an ODR platform. We tested this possibility by prompting an LLM to suggest a relevant pathway from a list of options, based on the user situation. The answer showed that the model was able to link the situation to a pathway. However, as pointed out by participants in the room, the conclusion drawn by the model potentially failed to consider some caveats.

We also conducted an experiment to see whether an LLM could be useful in the negotiation stage. For example, they could be used to introduce calm language into a conversation, in order to make the parties more likely to settle. We attempted some experiments on this idea, to see whether an LLM could reformulate an aggressive message into a calmer one. This was successful. However, it also raises questions of maintaining the original intent of the message. Perhaps, this type of rewriting is more useful as a suggestion to the parties, by suggesting a way to phrase their message more calmly.

It is possible to use ChatGPT as a mediator or as a support for mediator by giving them specific instructions about how it should moderate the negotiation between two parties (e.g., avoid insults, support open discussion and positive feelings, etc). We gave these kinds of instructions to ChatGPT and added a dialogue between two parties of a dispute. ChatGPT provided the following advice to mediate the discussion:

“Gentlemen, I understand that both of you have concerns about the water leak. Let's take a step back and calmly discuss the issue. Party A, can you share with us when you first noticed the leak and when you informed Party B about it? Party B, can you share your perspective on the issue and the steps you took to address it? Let's work together to find a mutually acceptable resolution to this matter.”

This seems like a useful approach to guide the discussion towards more factual occurrences. However, it starts with the word “Gentlemen”, even though the gender of the parties was not provided in the prompt. This kind of mistake should be monitored for in using such a potential platform.

LLM for adjudication/to support judges

LLMs could potentially be used to help judges in drafting decisions. A possible experiment is to provide materials submitted by the parties to an LLM and see whether it could draft a “facts” section of a decision for a judge to review. Moreover, it would be interesting to train LLMs on previous court decisions and facts to see if it can learn legal reasoning and produce *the ratio decidendi* of a decision.

Another interesting use of ChatGPT would be to give it a case encountered by judges and to ask it to find similar cases within a court decisions database. This task is already well accomplished by some algorithms, but ChatGPT could potentially be more granular when sorting cases into categories. In these ways, it would be interesting to explore to which extent LLMs can support judges by drafting decision sections to be reviewed by judges.

Discussion

A number of important points were raised during the discussion section of the meeting.

An important point that was discussed is the difficulty of using LLMs in domains where there is a need for trustable information. Even when finetuned, the models may make mistakes and hallucinate answers. Potentially, approaches that connect LLMs to explainable models such as decision trees may overcome some of these issues. Or, language models could enhance the efficiency of legal experts that create legal decision support tools. Legal experts have the background knowledge to be able to evaluate and adapt the needed legal information.

To better understand the risks of using GPT models to give legal information directly, it was suggested to perform a test of the JusticeBot and GPT model. This test could show the differences in the capabilities and approaches, including the accuracy and pedagogy of the answers. Such a test will be carried out over the coming weeks.

The discussion also touched on training a new version of a GPT model on legal data. To train a completely new version of ChatGPT does not seem feasible. OpenAI has used an enormous quantity of data, computer performance and human feedback. Replicating this research would be very expensive. OpenAI offers an API that can allow the finetuning of the model on specific data sources. However, depending on the quantity of the data, this can also quickly become quite expensive.

Conclusion

There are many potentially interesting use cases of LLMs in the field of justice. The combination of the structured JusticeBot approach with LLMs is very promising. We are excited to explore these ideas in the coming weeks and months.