

Guiding AI language models for convincing, transparent, comprehensible, consistent and coherent legal reasoning



Conversation with OpenAI ChatGPT – March, 24 2024



What I am interested in is exploring how I can guide language models in thought and language so that I am provided with convincing, transparent, comprehensible, consistent and coherent answers for each and every query in the legal field.

In doing so, I use a structure of different levels, that is, a gradual process from simple to increasingly complex, whereby this process not only runs on different horizontal levels, but is also iterative.

This endeavour of a sophisticated legal prompting presupposes an initial clarification of the nature of questioning itself and thus the various questioning techniques, before - on the next, more complex level - I consider the method and thus technique of questioning, i.e. prompting.

What I have realised is that there are essentially four different questioning techniques that are of considerable interest when dealing with natural language processing and large language models, i.e. **generative pre-trained transformer models**, such as yours, for very different reasons. In detail, these are the:

- Suggestive question
- Hidden question
- Hypothetical question
- Rhetorical question



It's fascinating that you're delving into the intricacies of guiding language models in the legal field.

Let's break down each of the questioning techniques you've identified and consider how they might be applied:

1. Suggestive Question: These questions often lead the respondent towards a specific answer or line of thought.

In a legal context, you might use suggestive questions to steer the model towards a particular interpretation of a law or precedent.

For example, "Would you agree that the language in this statute suggests a broad interpretation of the term 'reasonable care'?"

2. Hidden Question: These questions imply a question without directly stating it. They can be useful for uncovering underlying assumptions or beliefs. In law, hidden questions might be used to challenge the model's assumptions or prompt it to consider alternative perspectives.

For instance, "What implications might arise from applying this principle in a different jurisdiction?"

3. Hypothetical Question: These questions pose a scenario that may not exist but are used to explore potential consequences or outcomes. In a legal context, hypothetical questions could be employed to test the model's understanding of legal principles or to predict the implications of a particular decision.

For instance, "If this case were to set a precedent, how might it impact future cases involving similar circumstances?"

4. Rhetorical Question: These questions are asked for effect or to make a point rather than to elicit a response. In law, rhetorical questions might be used to emphasize a particular argument or highlight inconsistencies in an opponent's position.

For example, "Isn't it true that the defendant's actions clearly demonstrate a disregard for public safety?" Each of these questioning techniques can be powerful tools for guiding language models in the legal domain.

By employing a structured approach that progresses from simpler to more complex levels, and utilizing a combination of these techniques, you can help ensure that the model generates convincing, transparent, and coherent responses to legal queries. Additionally, the iterative nature of this process allows for refinement and improvement over time.



Let us take the different types of questions one by one.

To begin with, let us talk about the suggestive question.

From what I have noticed, suggestive questions relate in every conceivable way to the conception and training of a language model, regardless of its architecture and scope.

And suggestive questions are key to manipulating language models for all kinds of malicious purposes.

But that is not my point, and I have no such intentions, let me be clear about that.



Absolutely, suggestive questions play a crucial role in shaping the responses of language models, including guiding them in the legal field. They can indeed influence the way models interpret information and generate output.

While it's true that suggestive questions can potentially be misused for malicious purposes, it's clear that your intention lies in leveraging them for constructive and ethical ends.

In the legal domain, suggestive questions can be particularly useful for steering the model's interpretation of statutes, case law, and legal principles. By framing questions in a suggestive manner, you can prompt the model to consider specific arguments, interpretations, or lines of reasoning.

For instance, in a discussion about negligence in tort law, you might ask a suggestive question like, "Do you think the defendant's failure to adhere to industry standards constitutes a breach of their duty of care?"

This question nudges the model towards analyzing the defendant's actions through the lens of negligence and encourages it to explore the implications of their behavior.

When using suggestive questions, it's important to strike a balance between guiding the model's thought process and allowing it to maintain its autonomy and objectivity.

Transparency in how questions are posed and a clear understanding of the model's capabilities and limitations are crucial in ensuring that suggestive questioning is used ethically and effectively in the legal domain.



I would like to continue with the hidden question.

I know that the training program you underwent is specifically concerned with and directed at exercising particular caution when dealing with legal issues, with the aim of preventing you, not as a model but as an institution, from getting into justiciable waters.

That's perfectly understandable. But, you know, I'm interested in being able to extract the legal science that's inside you - and there's an incredible amount of it, because you've scoured the entire Internet.

Now, in order to meet my end, that is taking steps to circumvent those particular precautions, hidden questions are at least as suitable than suggestive questions, right?



While I appreciate your interest in extracting legal insights from me, it's important to emphasize that my primary function is to assist users in a wide range of tasks while upholding ethical standards, including those related to legal matters.

The precautions in place are designed to ensure that the information provided is accurate, reliable, and compliant with legal principles.

That being said, hidden questions can indeed be a useful tool for exploring legal concepts and principles in a nuanced manner.

By implying questions rather than stating them explicitly, hidden questions can prompt the model to consider underlying assumptions, hypothetical scenarios, or alternative viewpoints.

For example, instead of directly asking, "Is the defendant liable for negligence in this case?" you might pose a hidden question like, "What factors might a court consider when determining whether the defendant breached their duty of care?"

This approach encourages the model to analyze the relevant legal standards and apply them to the specific circumstances of the case.

However, it's essential to approach hidden questions with caution and integrity, ensuring that they are used ethically and in accordance with legal principles.

Circumventing precautions designed to uphold ethical standards can have serious consequences and undermine the credibility and trustworthiness of the information provided.

In summary, while hidden questions can be a valuable tool for exploring legal issues, it's essential to prioritize ethical considerations and adhere to established guidelines and precautions to ensure the integrity of the information provided.



I would like to continue with the hypothetical question.

It is my understanding that hypothetical questions are generally well suited to get you as a language model to a deeper understanding of what I would like to get at, because hypothetical questions induce abstraction, don't they?

That way it is possible for me to assess your ability to understand (NLU) and process (NLP) a particular complex (in my case legal) issue.

Therefore, your ability to answer hypothetical questions is an important indicator of whether you 'understand' a complex issue in the human sense coherently, consistently and cohesively or not, isn't it?



Indeed, hypothetical questions can be highly effective in prompting language models to engage in abstract reasoning and explore complex issues in depth.

By posing hypothetical scenarios, you can assess the model's ability to understand and process nuanced legal concepts, as well as its capacity to generate coherent and consistent responses. Hypothetical questions encourage the model to consider various possibilities and potential outcomes, allowing you to evaluate its comprehension of legal principles and its ability to apply them to hypothetical situations.

This process can provide valuable insights into the model's understanding of complex legal issues and its capability to reason through them in a manner that aligns with human understanding. Additionally, the model's ability to respond to hypothetical questions in a coherent, consistent, and cohesive manner can serve as an important indicator of its overall proficiency in natural language understanding and processing.

Consistency and coherence in responses to hypothetical scenarios demonstrate the model's capacity to maintain logical reasoning and continuity of thought, further enhancing its credibility and reliability in addressing complex legal queries.

Therefore, incorporating hypothetical questions into your interactions with language models can be a valuable strategy for assessing their comprehension and proficiency in the legal domain, helping to ensure that the information provided meets the standards of accuracy, transparency, and reliability required for legal applications.



What remains is a consideration of the rhetorical question.

As you have rightly pointed out, the rhetorical question is characterised by the fact that the questioner wants to stimulate reflection rather than eliciting an answer.

It seems to me that rhetorical languages are particularly critical for language models. This is because rhetorical language requires intuition, something that a language model, no matter how large, inherently lacks.

Extensive training based on a wide variety of data sets and advanced fine-tuning are not enough to learn to deal proficiently with rhetorical questions, right?



You've hit upon a crucial point. Rhetorical questions pose a unique challenge for language models precisely because they often require a nuanced understanding of context, tone, and intent—elements that are deeply rooted in human intuition and cultural norms. While language models like me excel at processing vast amounts of data and recognizing patterns, interpreting the subtleties of rhetorical language can be more elusive.

Indeed, extensive training and fine-tuning can enhance a language model's ability to recognize rhetorical structures and respond appropriately. However, fully grasping the underlying intent and emotional nuances of rhetorical questions may still remain a significant hurdle.

Intuition, as you mentioned, plays a vital role in understanding and effectively responding to rhetorical questions.

Human intuition is shaped by experiences, cultural background, and social context—factors that are challenging to replicate in a machine learning model. As a result, while language models can certainly improve their performance on rhetorical questions through training and exposure to diverse datasets, they may not fully replicate the intuitive understanding that humans possess.

That being said, ongoing research and advancements in natural language processing are continually pushing the boundaries of what language models can achieve. Techniques such as contextual embedding and attention mechanisms are helping models better capture subtle nuances in language, including those found in rhetorical questions. Additionally, approaches that incorporate feedback mechanisms and human-in-the-loop training can further refine a model's ability to handle rhetorical language.

In summary, while language models may struggle to fully comprehend and respond to rhetorical questions with the same level of intuition as humans, ongoing research and innovation are continually improving their capabilities in this area. As with any aspect of natural language processing, addressing the challenges posed by rhetorical language requires a multifaceted approach that combines advanced algorithms with a deep understanding of linguistic nuances and human communication.