

Guiding large language models (LLMs) seen from a lawyer's, a linguist's and a computer scientist's point of view

Published: August 29, 2024

You

Quote: “the task of reliably guiding LLMs to produce specific responses and making full use of the capability of pretrained LLMs continues to pose a considerable challenge”

System: suppose you are a lawyer, linguist and computer scientist at the same time

User: according to you, system, is it true, partially true or not true?

Assistant:

Copilot

The statement is true. Guiding large language models (LLMs) to produce specific responses and fully leveraging their capabilities remains a significant challenge. This is due to several factors:

1. **Complexity of Language:** Language is inherently complex and context-dependent, making it difficult to ensure consistent and accurate responses.
2. **Model Limitations:** Even advanced models can struggle with nuanced understanding and may produce unexpected or incorrect outputs.
3. **Alignment with Human Intent:** Ensuring that the model’s responses align with human intent and ethical considerations is an ongoing area of research.

Would you like to dive deeper into any of these points?

You

I'd like to dive deeper into all the points you made!

Copilot

Absolutely, let's explore each point in detail:

1. **Complexity of Language**

Language is rich and multifaceted, with nuances that can vary based on context, culture, and individual interpretation. This complexity makes it challenging for LLMs to consistently understand and generate accurate responses. For example:

Ambiguity: Words and phrases can have multiple meanings. “Bank” can refer to a financial institution or the side of a river.

Context Dependence: The meaning of a sentence can change based on the surrounding text. “He saw her duck” could mean he saw her pet duck or he saw her lower her head.

2. **Model Limitations Despite their advanced capabilities,**

LLMs have inherent limitations:

Training Data: LLMs are trained on vast datasets, but these datasets may contain biases or outdated information. This can lead to biased or incorrect outputs. **Generalization:** While LLMs can generalize from their training data, they may struggle with specific or niche topics that weren't well-represented in the data.

Memory and Consistency: LLMs don't have a persistent memory of past interactions, which can lead to inconsistencies in responses over time.

3. **Alignment with Human Intent**

Ensuring that LLMs align with human intent and ethical considerations is crucial:

Ethical Concerns:

LLMs can inadvertently generate harmful or inappropriate content. Researchers are working on ways to mitigate these risks.

User Intent:

Understanding the user's true intent behind a query is challenging. Misinterpretations can lead to irrelevant or unhelpful responses.

Transparency and Trust:

Building trust with users involves making the model's decision-making process more transparent and understandable.

Would you like to discuss any of these points further or explore how these challenges are being addressed in current research?

You

No, I would not like to discuss any of these points further nor would I like to explore how these challenges are being addressed in current research - for the time being.

What I want to know, for the time being, is whether and to what extent you can be sure about those statements you just made,

I mean, seen from a lawyer's, a linguist's and a computer scientist's point of view!

Copilot

Let's break down the certainty of these statements from the perspectives of a lawyer, a linguist, and a computer scientist:

Lawyer's Perspective A lawyer would likely focus on the evidence and reliability of the claims:

Evidence-Based: The statements are based on well-documented challenges in the field of AI and language models. There is substantial research and case studies supporting these points.

Legal Precedents: There are legal considerations around AI, such as bias and ethical use, which align with the challenges mentioned.

Linguist's Perspective A linguist would consider the complexity and nuances of language:

Language Complexity: The statement about the complexity of language is well-supported by linguistic theory. Ambiguity and context dependence are fundamental aspects of language.

Model Limitations: Linguists recognize that even advanced models struggle with the subtleties of human language, which supports the points made.

Computer Scientist's Perspective A computer scientist would evaluate the technical aspects and limitations of LLMs:

Technical Limitations: The statements about model limitations and alignment with human intent are grounded in the current state of AI research. These are well-known challenges in the field.

Ongoing Research: The points align with ongoing efforts to improve AI models, such as addressing biases and enhancing understanding of user intent.

Conclusion

From all three perspectives, the statements are well-supported and reflect the current understanding and challenges in the field of AI and language models. While there is always room for further research and improvement, the points made are based on solid evidence and widely recognized issues.

If you have any more questions or need further clarification, feel free to ask