

## Review

Search



### Understanding misunderstandings: evaluating LLMs on networking questions

Anwar M., Caesar M. ACM SIGCOMM Computer Communication Review 54 (4):14-24,2025.Type:Article

Date Reviewed: Jun 20 2025

Full Text

Large language models (LLMs) have awed the world, emerging as the fastest-growing application of all time--ChatGPT reached 100 million active users in January 2023, just two months after its launch. After an initial cycle, they have gradually been mostly accepted and incorporated into various workflows, and their basic mechanics are no longer beyond the understanding of people with moderate computer literacy. Now, given that the technology is better understood, we face the question of how convenient LLM chatbots are for different occupations. This paper embarks on the question of whether LLMs can be useful for networking applications.

This paper systematizes querying three popular LLMs (GPT-3.5, GPT-4, and Claude 3) with questions taken from several network management online courses and certifications, and presents a taxonomy of six axes along which the incorrect responses were classified:

- Accuracy: the correctness of the answers provided by LLMs;
- Detectability: how easily errors in the LLM output can be identified;
- Cause: for each incorrect answer, the underlying causes behind the error;
- Explainability: the quality of the explanations with which the LLMs support their answers;
- Effects: the impact of wrong answers on users; and
- Stability: whether a minor change, such as a change in the order of the prompts, yields vastly different answers for a single query.

The authors also measure four strategies toward improving answers:

- Self-correction: giving the original question and received answer back to the LLM, as well as the expected correct answer, as part of the prompt;
- One-shot prompting: adding to the prompt "when answering user questions, follow this example" followed by a similar correct answer;
- Majority voting: using the answer that most models agree upon; and
- Fine-tuning: further training on a specific dataset to adapt the LLM to a particular task or domain.

The authors observe that, while some of those strategies were marginally useful, they sometimes resulted in degraded performance.

The authors queried the commercially available instances of Gemini and GPT, which achieved scores over 90 percent for basic subjects but fared notably worse in topics that require understanding and converting between different numeric notations, such as working with Internet protocol (IP) addresses, even if they are trivial (that is, presenting the subnet mask for a given network address expressed as the typical IPv4 dotted-quad representation).

As a last item in the paper, the authors compare performance with three popular open-source models: Llama3.1, Gemma2, and Mistral with their default settings. Although those models are almost 20 times smaller than the GPT-3.5 commercial model used, they reached comparable performance levels. Sadly, the paper does not delve deeper into these models, which can be deployed locally and adapted to specific scenarios.

The paper is easy to read and does not require deep mathematical or AI-related knowledge. It presents a clear comparison along the described axes for the 503 multiple-choice questions presented. This paper can be used as a guide for structuring similar studies over different fields.

#### Recommendations

★ Featured Reviewer

#### Related Topics

Browse

Alerts

Language Models

(1.2.7 ...)

Add

Manage Alerts

More Alerts



Would you recommend this review?

yes

no

Other reviews under "**Language Models**":

**Date**

[A framework for investigating language-mediated interaction with machines](#)

Oct 1 1987

Zoeppritz M. International Journal of Man-Machine Studies 25(3): 295-315, 1986. Type: Article

[Prolog and natural-language analysis](#)

Jun 1 1988

Pereira F., Shieber S., CSLI/Stanford, Stanford, CA, 1987. Type: Book (9789780937073186)

[Competence and performance in the design of natural language systems](#)

Dec 1 1985

Bara B., Guida G., Elsevier North-Holland, Inc., New York, NY, 1984. Type: Book (9789780444875983)

[more...](#)

 [E-Mail This](#)

 [Printer-Friendly](#)

**REVIEWER'S AREA**

**MASTHEAD**

**SUBSCRIBE**

**NEWS**

**TIPS**

**HELP**

**CONTACT US**

Reproduction in whole or in part without permission is prohibited. Copyright 1999-2025 ThinkLoud®

[Terms of Use](#) | [Privacy Policy](#)