Letter to the Editor

Advantages and Inconveniences of a Multi-Agent Large Language Model System to Mitigate Cognitive Biases in Diagnostic Challenges

Cedric Bousquet[1,2], PharmD, PhD; Divà Beltramin[3], MD, MSc

[1]Laboratory of Medical Informatics and Knowledge Engineering in e-Health, Inserm, Sorbonne University, Paris, France

[2]Public Health and Medical Information Unit, Saint-Étienne University Hospital Center, Saint-Etienne, France

[3]Medical Information Department, Civil Hospices of Lyon, Lyon, France

Corresponding Author:
Cedric Bousquet, PharmD, PhD
Laboratory of Medical Informatics and Knowledge Engineering in e-Health
Inserm
Sorbonne University
15 rue de l'école de Médecine
Paris, F-75006
France
Phone: 33 0477127974
Email: cedric.bousquet@chu-st-etienne.fr

Related Article:
Comment on: http://www.jmir.org/2024/1/e59439/

(J Med Internet Res 2025;27:e69742) doi: 10.2196/69742

KEYWORDS

large language model; multi-agent system; diagnostic errors; cognition; clinical decision-making; cognitive bias; generative artificial intelligence

We read with great interest a recent article in the *Journal of Medical Internet Research* entitled "Mitigating Cognitive Biases in Clinical Decision-Making Through Multi-Agent Conversations Using Large Language Models: Simulation Study" by Ke et al [1]. Large language models (LLMs) have advanced reasoning skills but most studies have focused on assessing their ability to answer questions [2]. Ke et al [1] evaluated the ability of LLMs to avoid reproducing certain biases that physicians experience when making a diagnosis. The authors simulated decision-making using a multi-agent system in which each agent models the reasoning of a physician using an LLM. The experiment shows a significant increase in performance compared to a human, with an odds ratio of 3.91, which is highly remarkable. This is made possible by a particularly low score for humans, which leaves considerable room for improvement by LLMs.

Previously published studies have often shown disappointing results using LLMs, with performance levels that give cause for concern when used in a clinical context. Some studies have observed performance similar to that of physicians and marginally better in a limited number of cases. We are thus surprised to see that the multi-agent approach has led to such superiority compared to the physicians. This superiority may be explained by using a multi-agent framework, but it could also be limited to the context of addressing cognitive biases.

To investigate this hypothesis, it would be interesting to compare the performance of the multi-agent system with that of a single LLM. In particular, a prompt engineering method known as tree of thought allows several reasoning paths to be explored, enabling the initial choice to be modified according to diverse alternatives [3].

Furthermore, a recent study found no added value in associating an LLM with a physician to improve performance when making a diagnosis [4]. Moreover, the quantity of text and the number of interventions produced by multi-agent systems may be overwhelming, and the physician could ignore the suggestions. This was the case for many studies on computerized decision support systems for drug prescriptions, where too many alerts have led physicians to override relevant alerts [5].

The multi-agent system can lead to significantly higher costs and requires the mobilization of a greater number of resources. This depends on the number of agents involved in the multi-agent system and, above all, on the amount of text generated. The multi-agent system is likely to produce numerous arguments and counterarguments to make a fair diagnosis. It

would be interesting to know how many tokens are produced in the application logs compared to a simpler application.

The authors suggest that similar approaches could revolutionize medical practice by making clinical decisions more reliable and more consistent with the levels of evidence available. We subscribe to the idea that LLMs could help physicians, and the authors' approach seems promising. However, there is still a lack of evidence on how these multi-agent systems would perform if they were generalized to other diagnostic challenges.

Conflicts of Interest

None declared.

Editorial Notice

The corresponding author of "Mitigating Cognitive Biases in Clinical Decision-Making Through Multi-Agent Conversations Using Large Language Models: Simulation Study" declined to respond to this letter.

References

1. Ke Y, Yang R, Lie SA, Lim TXY, Ning Y, Li I, et al. Mitigating cognitive biases in clinical decision-making through multi-agent conversations using large language models: simulation study. J Med Internet Res. Nov 19, 2024;26:e59439. [FREE Full text] [doi: 10.2196/59439] [Medline: 39561363]
2. Bedi S, Liu Y, Orr-Ewing L, Dash D, Koyejo S, Callahan A, et al. Testing and evaluation of health care applications of large language models: a systematic review. JAMA. Oct 15, 2024:e2421700. [doi: 10.1001/jama.2024.21700] [Medline: 39405325]
3. Yao S, Yu D, Zhao J, Shafran I, Griffiths T, Cao Y, et al. Tree of thoughts: deliberate problem solving with large language models. In: Oh A, Naumann T, Globerson A, Saenko K, Hardt M, Levine S, editors. Advances in Neural Information Processing Systems 36 (NeurIPS 2023). San Diego, CA. Neural Information Processing Systems Foundation, Inc; 2023.
4. Goh E, Gallo R, Hom J, Strong E, Weng Y, Kerman H, et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. JAMA Netw Open. Oct 01, 2024;7(10):e2440969. [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.40969] [Medline: 39466245]
5. Slight SP, Seger DL, Franz C, Wong A, Bates DW. The national cost of adverse drug events resulting from inappropriate medication-related alert overrides in the United States. J Am Med Inform Assoc. Sep 01, 2018;25(9):1183-1188. [FREE Full text] [doi: 10.1093/jamia/ocy066] [Medline: 29939271]

Abbreviations

LLM: large language model

Edited by T Leung, L Beri; this is a non–peer-reviewed article. Submitted 06.12.24; accepted 11.12.24; published 20.01.25.

Please cite as:
Bousquet C, Beltramin D
Advantages and Inconveniences of a Multi-Agent Large Language Model System to Mitigate Cognitive Biases in Diagnostic Challenges
J Med Internet Res 2025;27:e69742
URL: https://www.jmir.org/2025/1/e69742
doi: 10.2196/69742
PMID: