Science | AAAS | SCIENCE JOURNALS | VOL. 384, ISSUE 6697 | PERFORMANCE OF A LARGE LANGUAGE MODEL ON THE REASONING... | 10 MAY 2024 | DOI: 10.1126/science.adq1918

RESEARCH ARTICLE | MACHINE LEARNING

Performance of a large language model on the reasoning tasks of a physician

Yindalon Aphinyanaphongs[1], Lisa B. Mesaros[2], Zoe Szendroi[3], Ruchi Thorpe[4], Ioana Danescu[5], Ilia Markov[6], Ariel Cohen[7], Ronald Gordon[8], Ronald Hirschtick[9], Jonathan Feldman[10], Jonathan Feldman[11], Jonathan Feldman[12], Jonathan Feldman[13], Jonathan Feldman[14], Jonathan Feldman[15], Jonathan Feldman[16], Jonathan Feldman[17], Jonathan Feldman[18], Jonathan Feldman[19], Jonathan Feldman[20], Jonathan Feldman[21], Jonathan Feldman[22], Jonathan Feldman[23], Jonathan Feldman[24], Jonathan Feldman[25], Jonathan Feldman[26], Jonathan Feldman[27], Jonathan Feldman[28], Jonathan Feldman[29], Jonathan

MDAR Reproducibility Checklist

DOWNLOAD | 287.16 KB

References and Notes

1. R. S. Ledley, L. B. Lusted, Reasoning foundations of medical diagnosis; symbolic logic, probability, and value theory aid our understanding of how physicians reason. *Science* **130**, 9–21 (1959).

   + SEE ALL REFERENCES • CROSSREF • PUBMED • GOOGLE SCHOLAR

2. K. Brodman, A. J. Erdmann Jr., I. Lorge, C. P. Gershenson, H. G. Wolff, The Cornell Medical Index-Health Questionnaire. III. The evaluation of emotional disturbances. *J. Clin. Psychol.* **8**, 119–124 (1952).

   GO TO REFERENCE • CROSSREF • PUBMED • WEB OF SCIENCE • GOOGLE SCHOLAR

3. F. T. de Dombal, D. J. Leaper, J. R. Staniland, A. P. McCann, J. C. Horrocks, Computer-aided diagnosis of

SHOW ALL REFERENCES

eLetters (0)

eLetters is a forum for ongoing peer review. eLetters are not edited, proofread, or indexed, but they are screened. eLetters should provide substantive and scholarly commentary on the article. Neither embedded figures nor equations with special characters can be submitted, and we discourage the use of figures and equations within eLetters in general. If a figure or equation is essential, please include within the text of the eLetter a link to the figure, equation, or full text with special characters at a public repository with versioning, such as Zenodo. Please read our Terms of Service before submitting an eLetter.

LOG IN TO SUBMIT A RESPONSE

No eLetters have been published for this article yet.

Recommended articles from TrendMD

How do we know how smart AI systems are?
Melanie Mitchell, Science, 2023

Numerical Cognition Without Words: Evidence from Amazonia
Peter Gordon, Science, 2004

Next-Generation Assessments
Lorrie A. Shepard, Science, 2010

Working Memory
Alan Baddeley, Science, 1992

Editors' Choice
Melissa McCartney, Science, 2013

Development and evaluation of a large language model of ophthalmology in Chinese
Ce Zheng, Br J Ophthalmol, 2024

POS0495 ASSESSING AND OPTIMIZING LARGE LANGUAGE MODELS ON SPONDYLOARTHRITIS MULTI-CHOICE QUESTION ANSWERING (SPAMC-QA)
Ann Rheum Dis, 2024

Cognitive science in the evaluation of medical AI systems
Vimla Lodhia Patel, BMJ Health & Care Informatics, 2023

Assessing the clinical reasoning of ChatGPT for mechanical thrombectomy in patients with stroke
Tse Chiang Chen, J Neurointerv Surg, 2024

Improving diagnostic performance through feedback: the Diagnosis Learning Cycle
Carolina Fernandez Branson, BMJ Quality & Safety, 2021

Powered by TREND MD

View full text | Download PDF

© 2026 American Association for the Advancement of Science. All rights reserved. AAAS is a partner of HINARI, AGORA, OARE, CHORUS, CLOCKSS, CrossRef and COUNTER. *Science* ISSN 0036-8075.