

Google's AMIE beats primary care doctors in a simulated text-chat diagnosis study

Source copy kept by wheresthe.ai for the story "Google's AMIE beats primary care doctors in a simulated text-chat diagnosis study" (<https://wheresthe.ai/d/google-amie-diagnostic-dialogue-study-2024-01/>).

Source: <https://arxiv.org/abs/2401.05654>

Text saved 18 Oct 2026 (the page as published).

SHA-256 of the saved text: 064579ff9c54a4b9e9bf83d9a61699c6fa1e13869820a9f6c2169b127f7b3c83

-->

Computer Science > Artificial Intelligence

arXiv:2401.05654 (cs)

[Submitted on 11 Jan 2024]

Title: Towards Conversational Diagnostic AI

Authors: Tao Tu , Anil Palepu , Mike Schaeckermann , Khaled Saab , Jan Freyberg , Ryutaro Tanno , Amy Wang , Brenna Li , Mohamed Amin , Nenad Tomasev , Shekoofeh Azizi , Karan Singhal , Yong Cheng , Le Hou , Albert Webson , Kavita Kulkarni , S Sara Mahdavi , Christopher Semturs , Juraj Gottweis , Joelle Barral , Katherine Chou , Greg S Corrado , Yossi Matias , Alan Karthikesalingam , Vivek Natarajan View a PDF of the paper titled Towards Conversational Diagnostic AI, by Tao Tu and 23 other authors

View PDF HTML (experimental)

Abstract: At the heart of medicine lies the physician-patient dialogue, where skillful history-taking paves the way for accurate diagnosis, effective management, and enduring trust. Artificial Intelligence (AI) systems capable of diagnostic dialogue could increase accessibility, consistency, and quality of care. However, approximating clinicians' expertise is an outstanding grand challenge. Here, we introduce AMIE (Articulate Medical Intelligence Explorer), a Large Language Model (LLM) based AI system optimized for diagnostic dialogue.

AMIE uses a novel self-play based simulated environment with automated feedback mechanisms for scaling learning across diverse disease conditions, specialties, and contexts. We designed a framework for evaluating clinically-meaningful axes of performance including history-taking, diagnostic accuracy, management reasoning, communication skills, and empathy. We compared AMIE's performance to that of primary care physicians (PCPs) in a randomized, double-blind crossover study of text-based consultations with validated patient actors in the style of an Objective Structured Clinical Examination (OSCE). The study included 149 case scenarios from clinical providers in Canada, the UK, and India, 20 PCPs for comparison with AMIE, and evaluations by specialist physicians and patient actors. AMIE demonstrated greater diagnostic accuracy and superior performance on 28 of 32 axes according to specialist physicians and 24 of 26 axes according to patient actors. Our research has several limitations and should be interpreted with appropriate caution. Clinicians were limited to unfamiliar synchronous text-chat which permits large-scale LLM-patient interactions but is not representative of usual clinical practice. While further research is required before AMIE could be translated to real-world settings, the results represent a milestone towards conversational diagnostic AI.

Comments: 46 pages, 5 figures in main text, 19 figures in appendix

Subjects:

Artificial Intelligence (cs.AI) ; Computation and Language (cs.CL); Machine Learning (cs.LG)

Cite as: arXiv:2401.05654 [cs.AI]

(or arXiv:2401.05654v1 [cs.AI] for this version)

<https://doi.org/10.48550/arXiv.2401.05654>

Focus to learn more

arXiv-issued DOI via DataCite

Submission history From: Tao Tu [view email] [v1] Thu, 11 Jan 2024 04:25:06 UTC (8,684 KB)

Full-text links: Access Paper:

View a PDF of the paper titled Towards Conversational Diagnostic AI, by Tao Tu and 23 other authors View PDF
HTML (experimental) TeX Source

view license

Current browse context:

cs.AI

< prev

| next >

new | recent | 2024-01

Change to browse by:

cs cs.CL cs.LG

References & Citations

NASA ADS Google Scholar

Semantic Scholar

export BibTeX citation Loading...

BibTeX formatted citation

×

loading...

Data provided by:

Bookmark

Bibliographic Tools

Bibliographic and Citation Tools

Bibliographic Explorer Toggle

Bibliographic Explorer (What is the Explorer?)

Connected Papers Toggle

Connected Papers (What is Connected Papers?)

Litmaps Toggle

Litmaps (What is Litmaps?)

scite.ai Toggle

scite Smart Citations (What are Smart Citations?)

Code, Data, Media

Code, Data and Media Associated with this Article

alphaXiv Toggle

alphaXiv (What is alphaXiv?)

Links to Code Toggle

CatalyzeX Code Finder for Papers (What is CatalyzeX?)

DagsHub Toggle

DagsHub (What is DagsHub?)

GotitPub Toggle

Gotit.pub (What is GotitPub?)

Huggingface Toggle

Hugging Face (What is Huggingface?)

ScienceCast Toggle

ScienceCast (What is ScienceCast?)

Demos

Demos

Replicate Toggle

Replicate (What is Replicate?)

Spaces Toggle

Hugging Face Spaces (What is Spaces?)

Spaces Toggle

XYZ.AI (What is XYZ.AI?)

Related Papers

Recommenders and Search Tools

Link to Influence Flower

Influence Flower (What are Influence Flowers?)

Core recommender toggle

CORE Recommender (What is CORE?)

Author

Venue

Institution

Topic

About arXivLabs

arXivLabs: experimental projects with community collaborators

arXivLabs is a framework that allows collaborators to develop and share new arXiv features directly on our website.

Both individuals and organizations that work with arXivLabs have embraced and accepted our values of openness, community, excellence, and user data privacy. arXiv is committed to these values and only works with partners that adhere to them.

Have an idea for a project that will add value for arXiv's community? Learn more about arXivLabs .

Which authors of this paper are endorsers? | Disable MathJax (What is MathJax?)