# In Real Clinic Study Google's AI Medical Assistant Shows Doctor-Level Diagnostic Reasoning

Researchers conducted a prospective feasibility study to evaluate the real-world performance of AMIE (Articulate Medical Intelligence Explorer) at conducting pre-visit history-taking with 100 adult patients at an urgent care clinic.

While the model was supervised in real time by a physician during patient–AI interactions and evaluated by clinicians at multiple stages of the study, the study found that the LLM-based conversational AI model performed safely with no predefined safety stops triggered during the interactions.

Furthermore, the model was reported to improve patients' attitudes toward medical AI and to generate differential diagnoses of comparable quality to those of human primary care providers when evaluated by blinded physician reviewers.
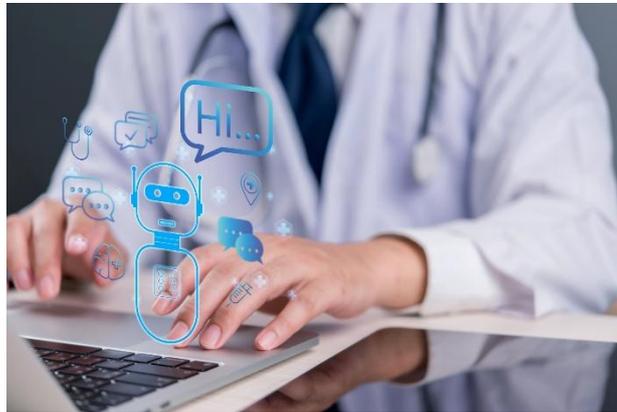
Study

The study was designed as a prospective single-arm feasibility study conducted at Healthcare Associates, an ambulatory primary care practice within Beth Israel Deaconess Medical Center.

The study participants were 100 adult patients already scheduled for non-emergency urgent care visits.

Up to five days before their scheduled appointment, participants engaged in a secure text-based chat with AMIE.

The model's performance was monitored during patient intake, during which AMIE gathered each patient's medical history while dynamically adapting its questions based on suspected conditions and information gaps rather than relying on static questionnaires.

All patient–AI interactions were monitored in real time by a board-certified internal medicine physician via screen sharing.

Following the intake interaction, participants completed surveys assessing their experience.

A summary of the chat transcript, along with an automatically generated clinical summary and participants' survey results, was forwarded to the clinician scheduled to see the patient ahead of the urgent care visit.

Finally, an independent panel of physicians performed a blinded chart review eight weeks later, comparing the accuracy and safety of management plans generated by both AMIE and human clinicians against the patient's finalized clinical assessment documented in the medical record after the visit and follow-up.

Results and Conclusion

In the trial's primary safety outcome, AMIE was judged safe under supervision. Physicians supervising the patient–AI interactions did not trigger a single safety stop across all 100 interactions, although minor clarifications were occasionally provided.

Interacting with the chatbot also significantly improved patient trust.

Survey scores on the General Attitudes toward AI Scale (GAAIS) shifted positively after the chat ($p < 0.001$) and remained elevated even after the patient saw their physician.

When evaluating AMIE's clinical reasoning capabilities, blinded evaluators found no significant difference in the overall quality of differential diagnoses ($p = 0.6$) between AMIE and human clinicians.

Furthermore, the appropriateness ($p = 0.1$) and safety ($p = 1.0$) of the AI's proposed management plans were comparable to those of human clinicians in blinded evaluations of standardized case summaries.

However, human clinicians significantly outperformed AMIE in designing management plans that were both practical ($p = 0.003$) and cost-effective ($p = 0.004$).

These differences likely reflect clinicians' greater access to contextual patient information and real-world healthcare constraints, including access to longitudinal medical records and workflow considerations that were not fully available to the AI during the study.

The study demonstrates that a conversational diagnostic AI system can safely and effectively gather clinical histories from real patients in a busy primary care clinic when used within a supervised research setting.

While AI is not yet ready to practice medicine autonomously, these findings support its emerging role as a collaborative clinical tool and physician assistant. The results highlight the need for larger multi-site studies to confirm safety, effectiveness, and generalizability across diverse patient populations.

Source:
https://www.news-medical.net/news/20260312/Googlee28099s-AI-medical-assistant-shows-doctor-level-diagnostic-reasoning-in-real-clinic-study.aspx