

# Google AMIE: Diagnostic Chatbot Tested with 100 Real Patients Matches Doctors on Accuracy

Kabui, Charles

2026-04-06

---

[Read at ToKnow.ai](#)

---

The infographic features a dark blue background with a white grid pattern. At the top right, there is a faint line graph with three data points. The main title is in large white and teal font. Below it, a subtitle in white text states the study location. Three key statistics are presented in separate boxes with vertical bars on the left: a 90% diagnostic accuracy rate (orange bar), zero safety stops (yellow bar), and a 75% top-3 diagnosis accuracy (blue bar). The date 'April 6, 2026' is at the bottom left, and the 'ToKnow.ai' logo is at the bottom right.

## Google AMIE: Diagnostic Chatbot Matches Doctors on Accuracy

Tested with 100 real patients at a Harvard-affiliated hospital

- 90%** Diagnostic accuracy rate  
Confirmed at 8-week follow-up
- Zero** Safety stops triggered  
Across all 100 patient chats
- 75%** Top-3 diagnosis accuracy  
Correct in top 3 most likely

April 6, 2026

ToKnow.ai

Google tested its conversational diagnostic AI, [AMIE](#), with [100 real patients](#) at Beth Israel Deaconess Medical Center (Harvard-affiliated) in a prospective clinical feasibility study. Previous conversational AI studies used simulated patients or were retrospective; this is the first

prospective study of an LLM-based diagnostic chatbot in a live primary care clinic with real-time physician oversight. Patients chatted with AMIE via text before their primary care appointment while a physician monitored live, ready to intervene on four pre-defined safety criteria. None were triggered. AMIE generated a transcript and summary for the doctor to review before the visit. A blinded panel of evaluators rated AMIE's differential diagnoses (ranked lists of possible conditions) and management plans on par with the primary care physicians (PCPs). AMIE matched the correct final diagnosis, confirmed via chart review eight weeks later, in **90% of cases**, with 75% top-3 accuracy. PCPs only outperformed AMIE on practicality and cost of management plans, expected since AMIE lacked access to medical records or physical exams.

Doctors spend a large chunk of patient visits just gathering basic history and symptoms. PCPs who reviewed AMIE's pre-visit summaries reported their consultations shifted from data collection to data verification and shared decision-making, which is the high-value part of a visit. Patient attitudes toward AI improved significantly after interacting with AMIE ( $p < 0.001$ ), countering fears of technology rejection. For a healthcare system stretched thin on primary care providers, an AI that reliably takes patient history and gets the diagnosis right 90% of the time could meaningfully free up clinician time.

This is a single-center feasibility study, not a controlled efficacy trial, and Google has committed to **larger controlled follow-ups**. But the direction is clear: clinical AI is moving from simulated benchmarks to real patients. The bottleneck is rigorous evidence generation, not technical capability.

Sources:

- [Google Research Blog: AMIE Clinical Study](#)
- [Paper: Prospective Clinical Feasibility Study \(arXiv\)](#)
- [Clinical Trial Registration \(NCT06911398\)](#)
- [AMIE Clinician Study \(Nature\)](#)
- [AMIE Patient Actor Study \(Nature\)](#)

---

*Disclaimer: For information only. Accuracy or completeness not guaranteed. Illegal use prohibited. Not professional advice or solicitation. Read more: [/terms-of-service](#)*