



Transforming healthcare delivery with conversational AI platforms

Arjun Mahajan & Dylan Powell



Healthcare communication faces unprecedented challenges as the healthcare workforce is increasingly faced with increased administrative burdens and reduced time with patients. Conversational agents powered by generative AI may offer a potential solution by collecting information, answering questions, documenting encounters, and supporting clinical decision-making through fluid, contextual dialogue. However, realizing their potential requires rigorous validation, careful implementation, and a strong commitment to safety, equity, and preserving human-centered care.

Bridging the communication gap

Clinician-patient communication has long been a cornerstone of modern medicine, fostering trust and comprehensive, patient-centered care. However, the quality and quantity of patient-provider interactions face an unprecedented crisis of scale and sustainability¹. Research indicates that physicians now spend an average of 15 to 18 minutes with patients during primary care visits, with nearly half of their clinic day devoted to documentation and non-clinical work – often exceeding the time spent in direct patient interaction^{2–4}. Nurses and supporting staff manage overwhelming caseloads, and administrative burdens consume resources that would otherwise support direct care^{1,3}. As a recent npj Digital Medicine article highlights, generative AI voice agents represent one potential tool to help address this communication crisis⁵. These systems, powered by large language models capable of natural speech understanding and generation, can facilitate patient interviews, support documentation, and enhance real-time clinical dialog. This article explores key opportunities and barriers that will help determine whether generative AI voice agents can successfully enhance healthcare communication.

Beyond traditional automation: the generative AI advantage

AI voice technologies in healthcare encompass a spectrum of applications, from ambient listening systems that passively capture and transcribe patient-provider conversations for documentation, to fully interactive agents designed for direct patient engagement⁶. Unlike traditional automated phone systems or rigid chatbots, generative AI voice agents are systems that can engage in more fluid, contextual dialog that adapts to individual patient needs⁵. As demonstrated with other generative AI systems, they may be able to recognize emotional cues, ask clarifying questions about symptoms, and integrate multiple data sources to provide personalized responses^{6,7}.

Traditional systems, like rule-based chatbots, rely on predetermined decision trees and scripted responses, limiting their utility to narrow, predictable tasks^{8,6}. In contrast, generative AI voice agents, may draw from extensive medical literature, clinical datasets, and previous interactions to produce contextually appropriate responses through dynamic interaction⁷. This technological advancement enables several critical capabilities.

Figure 1 AI voice agents have the potential to pursue nuanced lines of questioning based on patient responses, similar to how experienced clinicians gather history through iterative dialog rather than standardized questionnaires; these abilities may enable them to serve as effective triage and clinical decision support tools that aid rather than replace human clinical judgment⁹. For instance, in a randomized crossover trial, an AI-enabled voice assistant captured SARS-CoV-2 screening histories with 97.7% agreement compared to human staff and was rated “good or outstanding” by 87% of participants, illustrating that well-designed conversational agents may match clinician performance for front-line screening and hand off structured information for downstream decision-making¹⁰.

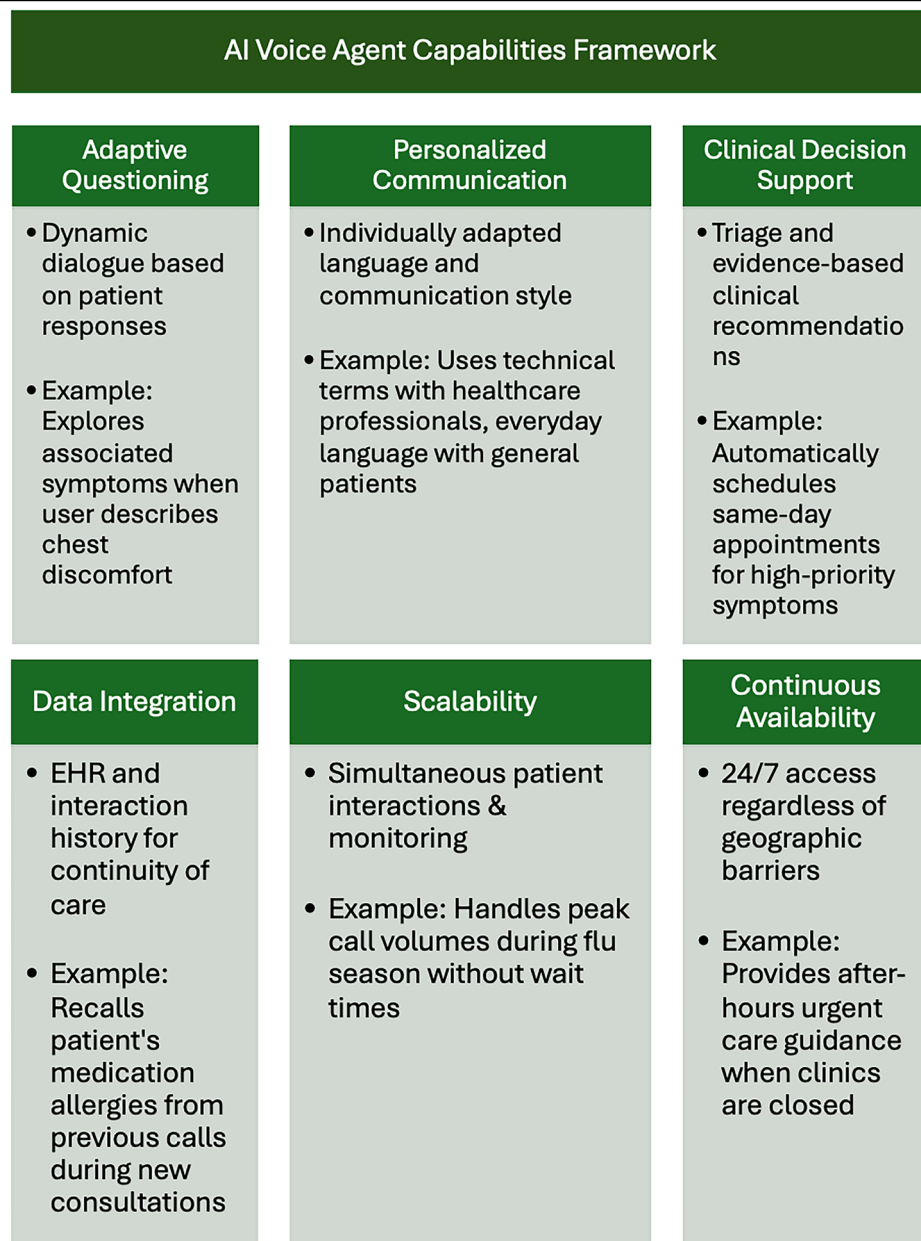
These systems may also modify their language complexity, cultural references, and communication style based on individual patient characteristics and preferences, potentially improving health literacy outcomes across diverse populations. In one study, a mental health supporting multilingual AI agent recorded significantly more and longer sessions in its Spanish than English version amongst primarily Spanish-speaking users, and Spanish-speaking users engaged more often with free-text therapeutic exercises¹¹.

AI agents may offer patients more consistent access to health guidance regardless of geographic barriers, resource limitations, or settings where traditional healthcare services are unavailable. This near-continuous availability enables functions like medication adherence monitoring through regular check-ins, pill counting reminders, and side effect tracking that would be resource-intensive for human staff to conduct consistently¹².

By integrating historical interaction data and electronic health records, AI agents might maintain continuity across multiple encounters, remembering previous concerns and tracking progress over time⁷. For example, a voice AI agent could follow up with a patient who previously reported low mood by naturally referencing earlier descriptions of poor sleep or loss of interest, and then asking whether these symptoms have changed - creating a fluid, conversational experience that more closely mirrors longitudinal care from a familiar clinician. In a randomized trial of oncology patients, patients completed weekly electronic symptom questionnaires that were graphed cumulatively in the EHR and auto-alerted clinicians; this longitudinal “memory” of prior concerns helped reduce emergency-department visits and extended median overall survival versus the usual standard of care, demonstrating concrete clinical gains that may come from technology and AI (such as a voice agent)-enabled data tracking across encounters¹³.

Finally, unlike human providers constrained by time and availability, AI agents could conduct simultaneous interactions with larger groups of patients, enabling opportunities for more proactive outreach and monitoring. Early safety evaluations have demonstrated high medical advice accuracy rates,

Fig. 1 | Selected capabilities of AI-enabled voice agents.



though these preliminary findings require further validation through reproducible evaluations across differing patient and care contexts^{14,15}.

Technical and safety challenges

Despite their promise, generative AI voice agents face significant hurdles that may determine their clinical utility. Latency remains a critical constraint – computational delays that create awkward pauses during medical conversations can undermine patient trust and interrupt natural dialog flow¹⁶. Accurately identifying end-of-utterance boundaries remains a challenge, often resulting in premature interruptions or uncomfortable silences during patient interactions¹⁷. Audio quality degradation or background noise could lead to misinterpretation of critical symptoms or patient responses, potentially resulting in inappropriate clinical recommendations¹⁸.

Importantly, the generative nature that makes these systems powerful also introduces unpredictability, as generative AI agents can produce novel or biased responses that may be clinically inappropriate or potentially harmful^{19,20}. AI systems may fail to reliably identify high-risk scenarios or delay transferring patients to human clinicians when immediate intervention is needed. AI agents might not recognize when clinical situations exceed their capabilities or may inadequately communicate their limitations to patients²⁰.

Thus, the deployment of AI voice agents in healthcare contexts carries inherent safety risks that demand robust mitigation strategies.

Implementation and regulatory challenges

Successful AI voice agent deployment may also require comprehensive organizational change management extending beyond technology

acquisition. Healthcare organizations must navigate complex integration challenges with existing electronic health record systems and develop robust quality assurance protocols to monitor system performance²¹. Healthcare providers and users require instruction not only in AI system operation but in maintaining clinical judgment while leveraging AI capabilities and recognizing when human intervention is necessary. Healthcare systems must create sustainable financial models that account for significant upfront technology investments and the relatively higher ongoing computational costs associated with real-time generative AI systems²².

In the United States, voice AI systems face regulatory uncertainty as they can function both as unregulated communication tools and as Software as Medical Device (SaMD) requiring Food and Drug Administration (FDA) clearance within the same conversational platform, depending on whether they provide specific clinical recommendations versus general information that clinicians independently review⁷. While the FDA has authorized over 1000 AI-enabled medical devices through traditional premarket pathways, regulators acknowledge that adaptive and generalized AI systems present challenges for frameworks designed for more static, single-indication devices²³. Furthermore, monitoring conversational outputs at scale across diverse, unpredictable use cases poses unique post-market surveillance challenges compared to more traditional, single task-oriented predictive algorithms. Potential solutions have been proposed, including tiered regulatory frameworks, with oversight intensity corresponding to clinical risk levels²⁴.

Trust and adoption

Public acceptance of AI voice agents faces significant barriers rooted in previous negative experiences with automated systems, privacy concerns, and preferences for human interaction during vulnerable health moments^{5,26}. Patients may approach AI health agents with skepticism developed through encounters with spam calls, malfunctioning chatbots, and impersonal automated services.

Building sustainable trust requires clear communication about when patients are interacting with AI systems, demonstration of understanding of individual patient contexts, and consistent performance with appropriate escalation to human care when necessary. AI systems must adapt communication styles and cultural references to diverse patient populations while offering options for human interaction and preserving patient autonomy through easy opt-out mechanisms.

Conclusion: toward responsible innovation

Generative AI voice agents represent a potential opportunity to extend personalized healthcare communication at large scale, enabling overworked clinicians to focus on complex cases while AI handles routine interactions and potentially reducing barriers that prevent equitable care access. However, realizing this potential requires addressing substantial technical, regulatory, and implementation challenges through continued technological advancement, evolving regulatory frameworks, and comprehensive organizational change management. The future of healthcare communication is being shaped by today's decisions, and success will depend not only on technological sophistication but on commitment to rigorous validation, thoughtful implementation, and preservation of the empathy and human connection that remain fundamental to healing.

Data availability

No datasets were generated or analysed during the current study.

Arjun Mahajan¹ & Dylan Powell² ✉

¹Harvard Medical School, Boston, MA, USA. ²Faculty of Health Sciences & Sport, University of Stirling, Stirling, UK. ✉e-mail: dylan.powell@stir.ac.uk

Received: 23 July 2025; Accepted: 19 August 2025;

Published online: 30 September 2025

References

- Drossman, D. A. & Ruddy, J. Improving patient-provider relationships to improve health care. *Clin. Gastroenterol. Hepatol.* **18**, 1417–1426 (2020).
- Young, R. A., Burge, S. K., Kumar, K. A., Wilson, J. M. & Ortiz, D. F. A time-motion study of primary care physicians' work in the electronic health record era. *Fam. Med.* **50**, 91–99 (2018).
- Neprash, H. T. et al. Measuring primary care exam length using electronic health record data. *Med Care* **59**, 62–66 (2021).
- Sinsky, C. et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. *Ann. Intern. Med.* **165**, 753–760 (2016).
- Adams, S. J., Acosta, J. N. & Rajpurkar, P. How generative AI voice agents will transform medicine. *NPJ Digit. Med.* **8**, 353 (2025).
- Milne-Ives, M. et al. The Effectiveness of Artificial Intelligence Conversational Agents in Health Care: Systematic Review. *J. Med. Internet Res.* **22**, e20346 (2020).
- Zhang, P. & Kamel Boulos, M. N. Generative AI in medicine and healthcare: promises, opportunities and challenges. *Future Internet* **15**, 286 (2023).
- Mahajan, A., Heydari, K. & Powell, D. Wearable AI to enhance patient safety and clinical decision-making. *npj Digit. Med.* **8**, 176 (2025).
- Tu, T. et al. Towards conversational diagnostic artificial intelligence. *Nature* **642**, 442–450 (2025).
- Sharma, A. et al. Voice-assisted Artificial Intelligence-enabled screening for severe acute respiratory syndrome Coronavirus 2 exposure in cardiovascular clinics: primary results of the VOICE-COVID-19-II randomized trial. *J. Card. Fail.* **29**, 1456–1460 (2023).
- Dinesh, D. N., Rao, M. N. & Sinha, C. Language adaptations of mental health interventions: User interaction comparisons with an AI-enabled conversational agent (Wysa) in English and Spanish. *Digit. Health* **10**, 20552076241255616 (2024).
- Borges do Nascimento, I. J. et al. The global effect of digital health technologies on health workers' competencies and health workplace: an umbrella review of systematic reviews and lexical-based and sentence-based meta-analysis. *Lancet Digit. Health* **5**, e534–e544 (2023).
- Basch, E. et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. *J. Clin. Oncol.* **34**, 557–565 (2016).
- Ayers, J. W. et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern. Med.* **183**, 589–596 (2023).
- Bhimani, M. et al. Real-world evaluation of large language models in healthcare (RWE-LLM): a new realm of AI safety & validation. *medRxiv*, <https://doi.org/10.1101/2025.03.17.25324157> (2025).
- Wang, YC, Xue, J., Wei, C., & Kuo, C.J. An overview on generative AI at scale with edge-cloud computing. *arXiv*, <https://doi.org/10.48550/arXiv.2306.17170> (2023).
- Jacoby, D., Zhang, T., Mohan, A. & Coady, Y. Human latency conversational turns for spoken avatar systems. *arXiv*, <http://arxiv.org/abs/2404.16053> (2024).
- Draper, TC, et al. The impact of acoustic and informational noise on AI-generated clinical summaries. *medRxiv*, <https://doi.org/10.1101/2025.03.24.25324398> (2025).
- Mahajan, A. et al. Cognitive bias in clinical large language models. *npj Digit. Med.* **8**, 428 (2025).
- Hager, P. et al. Evaluation and mitigation of the limitations of large language models in clinical decision-making. *Nat. Med.* **30**, 2613–2622 (2024).
- Feng, J. et al. Clinical artificial intelligence quality improvement: towards continual monitoring and updating of AI algorithms in healthcare. *npj Digit. Med.* **5**, 66 (2022).
- Mahajan, A. & Powell, D. Generalist medical AI reimbursement challenges and opportunities. *npj Digit. Med.* **8**, 125 (2025).
- Office of the National Coordinator for Health Information Technology Health data, technology, and interoperability: certification program updates, algorithm transparency, and information sharing. *Fed. Register* **89**, 1192–1438 (2024).
- Dogra, S., Silva, E. & Rajpurkar, P. Reimbursement in the age of generalist radiology artificial intelligence. *npj Digit. Med.* **7**, 350 (2024).
- May, R. & Denecke, K. Security, privacy, and healthcare-related conversational agents: A scoping review. *Inform. Health Soc. Care* **47**, 194–210 (2022).
- Li, H. et al. Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. *npj Digit. Med.* **6**, 236 (2023).

Author contributions

A.M. and D.P. wrote the main manuscript text and prepared Fig. 1. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Dylan Powell.

Reprints and permissions information is available at

<http://www.nature.com/reprints>

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025