

VIEWPOINT

Clinical, Legal, and Ethical Aspects of Artificial Intelligence–Assisted Conversational Agents in Health Care

John D. McGreevey III, MD

University of Pennsylvania Health System, Perelman School of Medicine, Section of Hospital Medicine, Division of General Internal Medicine, Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia.

C. William Hanson III, MD

Perelman School of Medicine, University of Pennsylvania, University of Pennsylvania Health System, Philadelphia; and School of Engineering and Applied Science, University of Pennsylvania, Philadelphia.

Ross Koppel, PhD

Perelman School of Medicine, University of Pennsylvania, University of Pennsylvania Health System, Philadelphia; and Department of Medical Informatics, Jacobs School of Medicine, University at Buffalo, State University of New York, Buffalo.

Corresponding

Author: John D. McGreevey III, MD, University of Pennsylvania Health System, 3400 Spruce St, Philadelphia, PA 19104 (john.mcgreevey@pennmedicine.upenn.edu).

Conversational agents (CAs) are artificial intelligence (AI) programs that engage in a dialogue with users by interpreting their questions or concerns and replying to them in a text message, image, or voice format. Conversational agents typically imitate human conversation by applying natural language processing and machine learning and stand in contrast to text-based engagement platforms that accept discretely formatted human inputs and reply with preset messages. Conversational agents are now familiar tools in consumer life (eg, billing issues). They have increasingly become more common in health care spaces, especially during the coronavirus disease 2019 (COVID-19) pandemic, as well as in direct-to-consumer or clinical capacities.¹

A National Academy of Medicine (NAM) report² offers guidance on a broad range of AI technologies and the foundational work necessary to implement them safely and effectively. In this Viewpoint, we focus on CAs in particular as we examine evidence supporting their use and propose 12 considerations for clinicians and health care organizations to help promote safe and effective deployment of CAs into clinical practice (**Box**). These considerations are especially critical given the intimacy that CAs allow in their interactions with patients, a privilege and responsibility previously reserved only for licensed clinicians. Although the focus is on CA use within traditional health care settings, these considerations also apply to other CA use cases, such as direct-to-consumer mental health platforms, which originated with a simulated psychotherapist (ELIZA).³

Some evidence suggests CAs may be effective, such as by facilitating remote patient management, enabling clinicians to focus on other functions, and aiding data collection.⁴ A 2-week trial of Woebot (a cognitive behavioral therapy-enabled CA) demonstrated statistically significant improvement in standard depression scores in the CA group (n = 31) vs the control group (n = 25).⁵ In a 2014 study of 239 adults, patients were statistically significantly more likely to disclose history and emotions when informed they were interacting with a CA vs a human clinician.⁶

Literature describing fully deployed CAs in health care settings is limited, with few rigorous evaluations outside research contexts.³ The CA studies rarely incorporate assessments of unintended consequences or safety concerns. The evidence suggests CAs are not yet mature enough to reliably respond to patients' statements in all circumstances, even when those statements explicitly signal harm, such as "I want to commit suicide."⁷ Even though technologies to determine mood, tone, and intent are becoming more sophisticated, they are not yet universally deployed in CAs nor validated for most populations. Moreover, there is no mention of CAs in the US Food and Drug Administration's (FDA) proposed regulatory framework for AI or machine learning for software as a medical device nor is there a user's guide for deploying these platforms in clinical settings.⁸

Considerations for use of CAs appear in the **Box** and each offers questions clinicians, patients, health care leaders, and vendors must address, recognizing that the presence of these systems in health care settings and their relative autonomy are rapidly evolving. Patient safety considerations reflect the difficulty of CAs in interpreting patient meaning. It remains unclear how well CA systems will recognize nuanced statements that may signal potential harm or a benign action, such as a postpartum mother's statement: "I'm having feelings I've never had before—I'm going onto the balcony."

It is important to recognize the potential, as noted in the NAM report,² that CAs will raise questions of trust and may change patient-clinician relationships. A most basic question is to what extent CAs should extend the capabilities of clinicians (augmented intelligence) or replace them (artificial intelligence). Likewise, determining the scope of the authority of CAs requires examination of appropriate clinical scenarios and the latitude for patient engagement. For example, if a patient indicates a cough, should the response of CAs be limited to only: "Let me connect you to a nurse now" or should CAs have more expanded abilities with responses such as "Tell me more about your cough."

In partnerships between vendors and health organizations to use CAs, all must be mindful of converging incentives and work to balance these goals with attention to each of the domains. Given the potential of CAs to benefit patients and clinicians, continued innovation should be supported.² However, hacking of CA systems (as with other medical systems) represents a cybersecurity threat, perhaps allowing individuals with malicious intent to manipulate patient-CA interactions and even offer harmful recommendations, such as quadrupling an anticoagulant dose.

The federal government, including the FDA, should develop frameworks for appropriate classification and oversight of CAs. Using the NAM-recommended graduated regulatory approach, one path would be to classify CAs as low risk vs higher risk.² Low-risk CAs might be less automated, structured for a specialized task, and have relatively minor consequences if they fail. A CA that guides patients to appointments might be one such example. In contrast, higher-risk CAs would involve more automation (natural language processing, machine learning), unstructured, open-ended dialogue with patients, and have potentially serious patient consequences in the event of system failure. Examples of higher-risk CAs might be those that advise patients after hospital discharge or offer recommendations to patients about titrating medications. In these cases, the potential for (unintended) direct harm to a patient due to CA failures or anomalies is much greater. For both low-risk and high-risk CA systems,

Box. Considerations for Implementing Conversational Agents (CAs)

Patient Safety

- Who monitors the interactions between patients and CAs? Does monitoring occur 24 hours/day and 7 days/week or on another schedule?
- Is there a rigorously tested escalation pathway to a human clinician? What scenarios have been configured to initiate the escalation pathway?
- How well do CAs detect subtleties of language, tone, and context that may signal a risk for patient harm?

Scope

- What kinds of clinical tasks should be augmented or automated by CAs and which should not? How much guidance is appropriate for CAs to provide to patients?

Trust and Transparency

- Do clinicians trust CAs? Do patients? Should they?
- To what degree do clinicians and patients need to understand the workings of CAs to use them effectively, intelligently, and ensure the appropriate amount of trust?

Content Decisions

- What are the content sources for CAs that provide recommendations or guidance? Do the CA recommendations align with content sources and with supervising clinician recommendations?

Data Use, Privacy, and Integration

- Who can access exchanges between patients and CAs? Who owns or controls the data? Will the data be stored or purged? If stored, for what purposes (eg, research, commercial use)?
- Are conversations integrated into patients' electronic health records (EHRs) or do they remain in each device? Can EHR data be integrated into CAs to better contextualize interactions?

Bias and Health Equity

- Which patient groups are used to train algorithms? How representative are they? How do CAs evolve over time to reflect new user populations?
- How do CAs handle accents and speakers of other languages? What about various health literacy levels and compliance with the Americans with Disabilities Act?

Third-Party Involvement

- CAs should be protected against commercially motivated data sharing or marketing, while permitting referencing of evidence-based products and therapies. A balance is needed among

commercial, technology leadership, and other incentives for CA developers and health care organizations

Cybersecurity

- What if data, devices, or apps are hacked or monitored covertly and cause harm? Will CA conversation data be encrypted?
- Are there restrictions on CA access? Is 2-factor authentication required? What are the trade-offs between sufficient security and convenient access?

Legal and Licensing

- Who is accountable if CAs fail? The sponsoring health care organizations or clinicians? The CA vendors? All of the above?
- What is the role of insurance in CA services?
- Will there be required licenses or credentials for CAs similar to those required for clinicians?

Research and Development Questions

- What approach or tone works best for patients? Human vs robotic, empathetic vs stoic, terse vs engaging, female vs male vs gender-neutral?
- What are the most common questions or needs posed to CAs?
- What do patients find most and least useful? What motivates patients to use CAs? What are differential discontinuation rates? Why do some patients stop using CAs? What other functions are requested, are viable, and are needed most? What are patient outcomes with CAs?

Governance, Testing, and Evaluation

- How will decisions about CA selection, deployment, and use be governed? How will performance be tested and evaluated with actual patients before deployment?
- What types of standard performance metrics and evaluations will be developed and implemented? How will desired outcomes and unanticipated or undesirable outcomes, including biases, be captured and assessed on an ongoing basis? How will these assessments be used to continue, suspend, or modify use of CAs?
- How will hazards or anomalies be detected and addressed?

Supporting Innovation

- How can development, testing, and introduction of promising boundary-pushing technologies be balanced with the need to protect patients and address the other issues listed here?

periodic interaction sampling from conversational logs is needed to enable monitoring and improvement.²

Conversational agents are just beginning in clinical practice settings, with COVID-19 spurring greater interest in this field. The use of CAs may improve health outcomes and lower costs. Researchers

and developers, in partnership with patients and clinicians, should rigorously evaluate these programs. Further consideration and investigation involving CAs and related technologies will be necessary, not only to determine their potential benefits but also to establish transparency, appropriate oversight, and safety.

ARTICLE INFORMATION

Published Online: July 24, 2020.
doi:10.1001/jama.2020.2724

Conflict of Interest Disclosures: None reported.

Additional Contributions: We acknowledge the insights of the late Robert Wears, MD, PhD, on the role of technology in health care. We also thank Danielle Mowery, PhD, and Anahita Davoudi, PhD, for their review of an earlier draft and suggestions.

REFERENCES

1. Referral MD. 11 healthcare chatbots that can improve patient experience. Accessed February 18, 2020. <https://getreferralmid.com/2019/03/11-healthcare-chatbots-that-improve-patient-experience/>

2. Matheny ME, Israni ST, Ahmed M, Whicher D. *AI in Health Care: The Hope, The Hype, The Promise, The Peril*. National Academy of Medicine; 2019.
3. Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare. *J Am Med Inform Assoc*. 2018;25(9):1248-1258.
4. Piau A, Crissey R, Brechemier D, et al. A smartphone chatbot application to optimize monitoring of older patients with cancer. *Int J Med Inform*. 2019;128:18-23.
5. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot). *JMIR Ment Health*. 2017;4(2):e19.

6. Lucas GM, Gratch J, King A, Morency L-P. It's only a computer. *Comput Human Behav*. 2014;37:94-100. doi:10.1016/j.chb.2014.04.043
7. Miner AS, Milstein A, Schueller S, et al. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA Intern Med*. 2016;176(5):619-625.
8. US Food and Drug Administration. Proposed regulatory framework for modifications to artificial intelligence/machine learning-based software as a medical device. Accessed February 14, 2020. <https://www.fda.gov/media/122535/download>