npj | Digital Medicine

www.nature.com/npjdigitalmed

COMMENT OPEN

Check for updates

Cognitive plausibility in voice-based AI health counselors

Thomas Kannampallil[1,2]✉, Joshua M. Smyth[3], Steve Jones[4], Philip R. O. Payne[2] and Jun Ma[5]

Voice-based personal assistants using artificial intelligence (AI) have been widely adopted and used in home-based settings. Their success has created considerable interest for its use in healthcare applications; one area of prolific growth in AI is that of voice-based virtual counselors for mental health and well-being. However, in spite of its promise, building realistic virtual counselors to achieve higher-order maturity levels beyond task-based interactions presents considerable conceptual and pragmatic challenges. We describe one such conceptual challenge—cognitive plausibility, defined as the ability of virtual counselors to emulate the human cognitive system by simulating how a skill or function is accomplished. An important cognitive plausibility consideration for voice-based agents is its ability to engage in meaningful and seamless interactive communication. Drawing on a broad interdisciplinary research literature and based on our experiences with developing two voice-based (voice-only) prototypes that are in the early phases of testing, we articulate two conceptual considerations for their design and use—conceptualizing voice-based virtual counselors as communicative agents and establishing virtual co-presence. We discuss why these conceptual considerations are important and how it can lead to the development of voice-based counselors for real-world use.

npj Digital Medicine (2020)3:72 ; https://doi.org/10.1038/s41746-020-0278-7

INTRODUCTION

Artificial intelligence (AI) affords new opportunities for the practice of medicine ranging from early disease detection to precision medicine tools for cancer treatment[1,2]. The promise of AI is accompanied with corresponding hype regarding its potential for "revolutionizing the healthcare for patients and populations"[3]. One area of prolific growth in AI, especially in home-based settings, has been voice-based (or voice-only) personal assistants (e.g., Amazon's Alexa). Recent reports have suggested that the worldwide sales of Amazon's Alexa and Google Home have crossed 100 million and 50 million, respectively (https://bit.ly/3b5AkKS). However, their use has primarily focused on routine tasks including playing music, initiating conversations (e.g., "Alexa, Good morning"), searching for information (e.g., weather), and performing smart home functions based on voice instructions[4]. Their high adoption notwithstanding, use of voice-based AI technologies may be declining over time[5], attributable to the lack of advanced features for seamless voice interactions[6]. User evaluation studies of voice-based assistants have characterized the conversational interactions similar to "having a bad personal assistant"[7].

Nonetheless, recent initiatives have catalyzed the use of such technologies in the healthcare domain. For example, the United Kingdom's National Health Services wants to enable Alexa devices for patients to find answers to routine questions such as "what are the symptoms of chicken pox?" (https://www.bbc.com/news/health-48925345). Despite considerable skepticism regarding their pragmatic potential for reliable patient interactions[8], there have been some efforts to integrate voice-based applications within the patient care continuum. One domain where voice assistants are expected to grow is that of serving as health coaches or therapists (hereafter referred to as counselors).

Voice-based virtual counselors—as opposed to early forms of automated counseling that involve static interactions, often with text messages or menu-based interfaces—allow for relatively unscripted interactions[9]. These interactions, with seemingly natural communication, can be sensitive to patients' cognitive and emotional states, providing contextually relevant and empirically supported counseling, and thereby, promoting a sense of connection with their "virtual counselor." Such virtual counselors afford opportunities for behavioral health and wellness promotion, given the shortage of clinicians, reimbursement and many barriers to patient access to mental and preventive health services.

These voice-based counselors are different from a variety of prior innovations for virtual counseling. For example, embodied conversational agents (ECA)[10,11] utilized both voice-based and direct manipulation interfaces (e.g., user selection from a list) for clinical and cognitive support in various domains, including marriage counseling[12], palliative care counseling[13], and collecting family health history[14]. A more direct comparisons for the voice-based virtual counselors are the text-based "chat bots" that provided text-based emotional support[9]. A comprehensive review of these existing (predominantly, non-voice based) virtual counseling applications is beyond the scope of this article. Here we focus on the need to develop a fundamental conceptualization that is aligned with the cognitive processes underlying human communicative processes in order to realize the potential of a voice-based virtual counselor for intelligent, "human-like" conversations. Examples of similar cognitive alignment have been shown to affect a number of aspects of healthcare tools, such as usability, workflow, and patient safety[15].

For virtual counselors, given their focus on primarily voice-based interactions, a significant step would be the conceptualization of its *cognitive plausibility*. Cognitive plausibility is the ability of a system to emulate the human cognitive system by simulating how a skill or function is accomplished[16]. In more simplistic terms, cognitive plausibility is characterized as the functional performance of a system where the inputs and outputs to a system are

[1]Department of Anesthesiology, Washington University School of Medicine, St Louis, MO, USA. [2]Institute for Informatics, Washington University School of Medicine, St Louis, MO, USA. [3]Department of Biobehavioral Health, The Pennsylvania State University, University Park, PA, USA. [4]Department of Communication, University of Illinois at Chicago, Chicago, IL, USA. [5]Department of Medicine, University of Illinois at Chicago, Chicago, IL, USA. ✉email: thomas.k@wustl.edu

comparable to that of humans[17] (a more well-known equivalent example is the "Turing test").

For a voice-based virtual counselor, cognitive plausibility refers to its capability of engaging in and replicating conversations that mirror patient conversations with a human counselor. This can include the ability to identify patients' reasons of current concern, their medical or psychological problems, relate to or reference their past counseling sessions, assess their current emotional state (s) and contextual situations, develop a shared understanding regarding problems, and create a plausible action plan to address them with follow-up resolutions or further treatment options for unresolved problems.

Achieving cognitive plausibility in voice-based virtual counselors is challenging due to an array of technical and socio-technical constraints. In this comment, we describe two pragmatic considerations for developing cognitively plausible voice-based agents: creating them as communicative agents and establishing co-presence during communicative interactions. To exemplify the applicability of these conceptual considerations, we discuss them within the context of the design and development of two virtual counselor prototypes for mental health and emotional well-being on Amazon's Alexa platform—a counselor delivering problem-solving therapy for managing depression and anxiety and a counselor for mindfulness-based stress therapy.

VIRTUAL COUNSELORS AS COMMUNICATIVE AGENTS

Developing cognitive plausibility in voice-based virtual counselors requires a change regarding our conceptualization of AI-based conversational agents: as opposed to considering these as interactive objects for technological task-based transactions, these need to be designed as "communicative subjects"[6]. Within this conceptualization, AI voice assistants are interlocutors and play the dual role facilitating conversations and mediating social processes around interactive communication: social communicative processes such as conversational flexibility (e.g., switching topics), context awareness (e.g., knowing previous conversations), or intent recognition (e.g., inferring implicit intentions). Similarly, the conceptualization of voice-based virtual counselors should embody the role of social actors, and interactions should be anchored around how human counselors would normally channel and orient their interactions with patients undergoing counseling for physical or mental health problems[18,19]. In other words, a primary design and functional consideration for cognitive plausibility should be to design these counselors as communicators, as people will perceive and interact with them in the applicable healthcare context.

One of the ways cognitive plausibility can be achieved is through the management of interactivity in conversations. Human interactive communication is a "joint activity in which two or more interlocutors share or synchronize aspects of their private mental states and act together in the world"[20,21]. Achieving interactivity during conversations is largely dependent on utilizing appropriate conceptual frameworks for managing and supporting interactivity. Three cognitive theories are often used to describe interactive communication and its characteristics: the message model, the two-stage model, and the collaborative or grounding model[22]. The message model of communication is an "information transfer" framework based on information theory, where communication involves the transmission of informational content from a sender to a receiver through a channel[23]. In this framework, there is no understanding of nuances or intent of communication, with the focus merely being on transmitting and interpreting the informational content. In contrast, the two-stage model of communication relies on interactive alignment between conversational partners, where the content of communication of one partner primes the other, leading to a shared mental representation regarding the presented content[24,25]. Conversational continuity relies on priming and convergent communicative behaviors, which are typically achieved through mimicry and behavioral adjustments by interpreting and aligning (often, unconsciously) with a partner. The third model, representing in our view a relatively more sophisticated explanation of interactive human communication, postulated by Brennan and Clark[26], describes communication as a joint, collaborative activity. Communicative content and signals are recognized by partners and meanings are coordinated through a process of grounding, where partners provide supporting evidence that they "understand each other" through verbal or situational cues[27].

These models span the spectrum of human-like interactive communication—with the messaging model representing a mechanistic view of communication, and the two-stage and grounding models representing the nuanced, interpretive, and cognitive principles of interactive communication. This spectrum also broadly exemplifies the degree of complexity in simulating realistic voice-based interactions with a virtual counselor. Most current voice-based applications rely on the messaging model, where specific terminology (e.g., keywords) is used as the basis for routine interactions, without accounting for semantic, structural, temporal, or cognitive aspects of the spoken language. For example, a question "Alexa, what is the weather?" or "Alexa, who is the weather?" would elicit similar responses pertaining to the current local weather (relying on the keyword, "weather"). At the other end of the spectrum, the grounding model is a highly collaborative model that relies on visual and verbal cues from a conversational partner. Replicating such a conversational model in a virtual counselor is currently beyond the scope of voice assistants due to the limitations of technology to characterize human intent, such as identifying non-voice (e.g., gestural, facial) or other situational cues and assessing contextual factors in an environment. Such an interaction requires voice-based virtual counselors to be "stateful," by aligning voice-based interactions within the context of past/historical interactions.

However, it is potentially plausible for AI-based tools to more readily align with the two-stage messaging model that relies on behavioral adjustments and mimicry through a process of cognitive priming to a partner's communicative interactions. AI technology, in its current state, is able to detect dynamic changes during voice interactions related to conversational turns (e.g., break and switch between conversations), engage in repair of conversations (e.g., after a conversation breakdown/stoppage)[28–31], create contextual awareness (e.g., through integration with other sensors), and identify mood states (e.g., happy, sad). Such detection is a preliminary precursor for providing cues for adjustments throughout the conversation.

Our prototype design of the two voice-based virtual counselors was aligned with, and informed by, the two-stage model, incorporating features for interactivity and context-sensitive adaptiveness. First, we created conversational continuity through realistic turns for each speaker (virtual counselor, patient) with timed breaks in conversations. This involved the creation of shortened conversational turns for each speaker; for example, avoiding one participant speaking for an extended period. Virtual counselors speaking turn lengths were kept purposefully short; timed interruptions were incorporated to minimize patient conversational turn length. Shorter mean turn lengths are intended to increase interactivity, reducing the potential for distractions/diversions, and engaging patients in a conversation. Second, we developed considerable adaptivity during conversations, including acknowledgement of patient responses as a proxy for virtual counselor engagement, flexibility in the conversations (e.g., repair from breakdowns rather than abrupt stoppage), and awareness of previous sessions to drive current conversations. These design considerations are aligned with the two-stage model. Both prototypes are in the early phase of user testing, and

the findings will inform whether and how the designs help create realistic interactions with voice-based virtual counselors.

ESTABLISHING CO-PRESENCE

A key characteristic of patient–counselor interaction is their synchronous co-presence during a counseling session (e.g., face-to-face, by phone, or via telehealth means). Such synchronous conversational partners, where individuals are "available, accessible, and subject to one another"[32] can potentially help to foster a degree of interactivity and facilitates the building of trust and believability between the conversational partners. Such synchronous interactions are referred to as co-presence or social presence[33]. In the case of voice-based virtual counselors, co-presence of the human is with a digital communication agent, and such interactions are referred to as virtual telecopresence[34].

Aligning interaction designs focusing on establishing virtual telecopresence to enhance the role of virtual counselors as communication agents can potentially help in achieving cognitive plausibility in such interactions. This can be achieved in several ways: first, humans perceive interactive digital devices as social actors[6,35,36]; as such, trust and believability in the virtual counselor can help the patient engage in counseling sessions and adhere to their care plan. Second, creating embodiment, a degree of engagement that is created by the presence of humans in a conversation, can enhance the quality and nature of communicative interactions[13,14]. For example, embodied interactions in the presence of other humans provide cues such as facial expressions, eye contact, and gestures that mediate and drive the conversational interactions. In contrast, voice-only interactions with a virtual counselor have a reduced level of embodiment requiring additional external sources to enhance engagement.

In one of our prototypes, we used a smartwatch for enabling contextually aware responses to create a sense of social presence of the counselor. The purpose of creating such external sensors was as an "add on" for developing more engaging conversations. In on-going early testing, the smartwatch is used to detect stress through ecological momentary assessments; additional activities such as sleep patterns and general physical activities are also captured to create a more situated understanding regarding patient activities. Another promising direction, especially in the case of mental health virtual counselors, are the emotion-sensing features that can determine a patient's mood states via sentiment analysis and analysis of the acoustic features of speech such as pitch, tone, volume, and timbre. Preliminary evaluations suggest that integration of such sensing in a virtual counselor may help to better situate a conversation, providing not only a feeling of co-presence but also helps in developing rapport and trust in the counselor, potentially strengthening such relationships[37–39].

Co-presence can also be achieved through persistence and continuity in the interactions. Design features in virtual counselors such as the ability to continue conversations in different settings and situations through their integration in mobile phone or tablets as opposed to being on a tethered device (such as an Amazon device) can help in developing continuity.

CONCLUSIONS

The widespread adoption of voice-based AI assistants has opened new opportunities for its potential role in the delivery of evidence-based health counseling interventions. Based on our experience with developing voice-based virtual counselors for mental health and emotional well-being, we describe the conceptualization of cognitive plausibility in designing such virtual counselors in order to create seamless and seemingly human-like interactive communication. We specifically focus on voice-based virtual counselors relying on AI, as this is a burgeoning area of technology development with little empirical research. In addition, we are at stage where scholarly discourse about conceptual advances can have a meaningful impact on the emerging science. The concept of cognitive plausibility is relevant to the design of any system (e.g., ECAs) that aims to interact with humans in an intelligent, "human-like" manner. However, it is unclear, based on our review, how cognitive plausibility has been accounted in the design of ECAs. We posit that the considerations of conceptual plausibility are necessary to advance a maturity framework for the design of AI-based voice counselors such that they may achieve higher-order maturity levels beyond technological task-based transactions in order to be pragmatically useful and efficacious. Findings from empirical testing of such virtual counselors will help refine and extend these conceptual considerations, contributing to the maturation of this emerging field.

Received: 16 December 2019; Accepted: 16 April 2020;
Published online: 15 May 2020

REFERENCES

1. Topol, E. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **25**, 44 (2019).
2. Topol, E. *Deep Medicine: How Artificial Intelligence can make Healthcare Human Again* (Basic Books, New York, 2019).
3. Maddox, T. M., Rumsfeld, J. S. & Payne, P. R. Questions for artificial intelligence in health care. *JAMA* **321**, 31–32 (2019).
4. Sciuto, A., Saini, A., Forlizzi, J. & Hong, J. I. A mixed-methods studies of in-home conversational agent usage. In *Proceedings of the 2018 Designing Interactive Systems Conference* 857–868 (ACM, 2018).
5. Cho, M., Lee, S.-s. & Lee, K.-P. Once a kind friend is now a thing: Understanding how conversational agents at home are forgotten. In *Proceedings of the* 2019 *on Designing Interactive Systems Conference*, 1557–1569 (ACM, 2019).
6. Guzman, A. L. & Lewis, S. C. Artificial intelligence and communication: a human-machine communication research agenda. *New Media Soc* **22**, 70–86 (2019).
7. Luger, E. & Sellen, A. Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In *Proceedings of the* 2016 *CHI Conference on Human Factors in Computing Systems*, 5286–5297 (ACM, 2016).
8. Black, S. To benefit the NHS, AI must be guided by intelligence, not hype. https://blogs.bmj.com/bmj/2019/09/20/stephen-black-to-benefit-the-nhs-ai-must-be-guided-by-intelligence-not-hype/ (2019).
9. Laranjo, L. et al. Conversational agents in healthcare: a systematic review. *J. Am. Med. Inform. Assoc.* **25**, 1248–1258 (2018).
10. Provoost, S., Lau, H. M., Ruwaard, J. & Riper, H. Embodied conversational agents in clinical psychology: a scoping review. *J. Med. Internet Res.* **19**, e151 (2017).
11. ter Stal, S., Kramer, L. L., Tabak, M., op den Akker, H. & Hermens, H. Design features of embodied conversational agents in eHealth: a literature review. *Int. J. Hum. Comput. Stud* **138**, 102409 (2020).
12. Utami, D. & Bickmore, T. Collaborative user responses in multiparty interaction with a couples counselor robot. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* 294–303 (IEEE, 2019).
13. Utami, D., Bickmore, T., Nikolopoulou, A. & Paasche-Orlow, M. Talk about death: End of life planning with a virtual agent. In *International Conference on Intelligent Virtual Agents* 441–450 (Springer, 2017).
14. Wang, C. et al. Acceptability and feasibility of a virtual counselor (VICKY) to collect family health histories. *Genet. Med.* **17**, 822–830 (2015).
15. Kannampallil, T. G., Abraham, J. & Patel, V. L. Methodological framework for evaluating clinical processes: a cognitive informatics perspective. *J. Biomed. Inform.* **64**, 342–351 (2016).
16. Wenger, E. *Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge* (Morgan Kaufmann, 2014).
17. Kennedy, W. G. Cognitive plausibility in cognitive modeling, artificial Intelligence, and social simulation. In *Proceedings of the International Conference on Cognitive Modeling (ICCM), Manchester, UK*, 24–26 (ICCM, 2009).
18. Nass, C., Steuer, J. & Tauber, E. R. Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 72–78 (ACM, 1994).
19. Reeves, B. & Nass, C. I. *The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places* (Cambridge University Press, 1996).
20. Brennan, S. E., Galati, A. & Kuhlen, A. K. In *Psychology of Learning and Motivation*, Vol. 53, 301–344 (Elsevier, 2010).

21. Clark, H. H. & Brennan, S. E. In *Perspectives on Socially Shared Cognition* (eds Resnick, L. B., Levine, J. M. & Teasley, S. D.) 127–149 (APA Books, Washington, 1991).
22. Clark, H. H. *Using Language* (Cambridge University Press, 1996).
23. Shannon, C. E. & Weaver, W. *A Mathematical Model of Communication*, Vol. 11 (University of Illinois Press, 1949).
24. Pickering, M. J. & Garrod, S. Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.* **27**, 169–190 (2004).
25. Shockley, K., Richardson, D. C. & Dale, R. Conversation and coordinative structures. *Top. Cogn. Sci.* **1**, 305–319 (2009).
26. Brennan, S. E. & Clark, H. H. Conceptual pacts and lexical choice in conversation. *J. Exp. Psychol. Learn. Mem. Cognition* **22**, 1482 (1996).
27. Brennan, S. E. & Hanna, J. E. Partner-specific adaptation in dialog. *Top. Cogn. Sci.* **1**, 274–291 (2009).
28. Gao, J., Galley, M. & Li, L. Neural approaches to conversational AI. *Found. Trends® Inf. Retr.* **13**, 127–298 (2019).
29. Ram, A. et al. Conversational AI: the science behind the Alexa prize. In *1st Proceedings of Alexa Prize*. Amazon (2018).
30. Venkatesh, A. et al. On evaluating and comparing conversational agents. In *31st Conference on Neural Information Processing Systems* (NIPS, Long Beach, CA, 2018).
31. Yan, R. Chitty-Chitty-Chat Bot: Deep Learning for Conversational AI. In *IJCAI*, 5520–5526 (2018).
32. Goffman, E. *The Presentation of Self in Everyday Life* (Harmondsworth London, 1978).
33. Biocca, F. & Harms, C. Defining and measuring social presence: Contribution to the networked minds theory and measure. In *Proceedings of PRESENCE*, 1–36 (2002).
34. Zhao, S. Toward a taxonomy of copresence. *Presence Teleoperators Virtual Environ.* **12**, 445–455 (2003).
35. Guzman, A. L. *Human-Machine Communication: Rethinking Communication, Technology, and Ourselves* (Peter Lang Publishing, Incorporated, 2018).
36. Mlynář, J., Alavi, H. S., Verma, H. & Cantoni, L. Towards a sociological conception of artificial intelligence. In *International Conference on Artificial General Intelligence*, 130–139 (Springer, 2018).
37. Eyben, F., Weninger, F., Gross, F. & Schuller, B. Recent developments in opensmile, the munich open-source multimedia feature extractor. In *Proceedings of the 21st ACM International Conference on Multimedia*, 835–838 (ACM, 2013).
38. Mostafa, M., Crick, T., Calderon, A. C. & Oatley, G. Incorporating emotion and personality-based analysis in user-centered modelling. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, 383–389 (Springer, 2016).
39. Taguchi, T. et al. Major depressive disorder discrimination using vocal acoustic features. *J. Affect. Disord.* **225**, 214–220 (2018).

ACKNOWLEDGEMENTS

This work was supported, in part, by the Healthcare Innovation Lab and Institute for Informatics at BJC HealthCare and Washington University School of Medicine and by a grant-in-aid from the Division of Clinical and Translational Research of the Department of Anesthesiology, Washington University School of Medicine.

AUTHOR CONTRIBUTIONS

The concept for this paper was developed by T.K. and J.M. and expanded with discussions with J.M.S., S.J. and P.R.O.P. All authors contributed to the writing of this manuscript and approved the final version.

COMPETING INTERESTS

J.M. is a paid scientific consultant for Health Mentor, Inc. (San Jose, CA, USA).

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to T.K.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.