
Implications of Conversational Artificial Intelligence

Tonya Nguyen
University of California, Berkeley
Berkeley, CA, USA
tonyanguyen@berkeley.edu

Niloufar Salehi
University of California, Berkeley
Berkeley, CA, USA
nsalehi@berkeley.edu

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Copyright held by the owner/author(s).
CHI'20, April 25–30, 2020, Honolulu, HI, USA
ACM 978-1-4503-6819-3/20/04.
<https://doi.org/10.1145/3334480.XXXXXXX>

Abstract

UPDATED—July 3, 2020. The use of conversational AI has become more ubiquitous in public and private life. Although conversational artificial intelligence (AI) may be useful and seamless, using voice to mediate human-AI interaction raises potential privacy risks and socio-cultural implications for gender bias and childhood development. To better understand the downstream implications of AI, we pose a set of questions to frame how conversational AI and various stakeholders may interact within sociotechnical systems.

Introduction

AI is made conversational for better human interaction, but we don't fully understand the consequences of using voice to mediate human-AI interaction. In a prior study, we found that speaking with a vocal user interface increased the propensity of disclosing highly sensitive information [13]. Using voice puts users at risk of exposing sensitive data and may influence the development of children. Unlike submitting data through forms, speaking to a conversational AI is transient and seamless; users may not be aware of information they have shared or what the AI has inferred from their natural language utterances.

This paper discusses the potential consequences of using voice-based conversational AI. Drawing from design activities common to speculative design and value-sensitive

design, this paper poses a set of questions that guide us to think more critically about what may happen when humans and AI talk.

Related Work

Conversational Agents

AI agents (Siri, Alexa, Google Home) have been used to simulate human labor and interaction for a variety of purposes. These algorithmic agents have been reimagined as personal assistants, navigators, educators, and even judges. Some have characterized conversational AI as a means to mitigate emotional labor and burnout associated with professions in social work, therapy, and criminal justice [11].

Philosophical Critiques of AI

Weizenbaum, the creator of the first chatbot, argued against using AI technologies to replace people in roles that require respect and care. Today, Weizenbaum's argument applies to the creation of therapy bots, mental health bots, automated personal assistants. Weizenbaum explains that if machines replace jobs that require human empathy, humans will find themselves "alienated, devalued, and frustrated." Ultimately, Weizenbaum believes that "thinking of ourselves as computers" leads to the "atrophy of the human spirit." [12].

Techniques from Speculative Design

Design Fiction has been used to probe values and ethics in sociotechnical systems. [1]. These practices ask "what if?" to surface and discuss the relationships among the social, political, and technical by using artifacts to build a storyworld rich with multiple perspectives. Meanwhile, Value Sensitive Design incorporates reflecting on social values in the design process, encouraging designers to consider direct and indirect stakeholders over time [2].

To better understand the landscape of implications, we ask "what if?" and identify vulnerable stakeholders of conversational AI. Future work in speculative design may start with asking and answering these questions in greater detail.

Media Equation Theory and Voice

In this section, we ask: what is the technology in context? We highlight the historical context of conversational AI, the theory behind human-computer relations, and the affordances of voice.

Media Equation Theory

The Media Equation Theory posits that people apply social rules, which widely exist in real-life social settings, to their interactions with computers. In a controlled experiment, people were asked to rate the performance of a computer after completing a task with it. If the computer asked someone to rate itself, that person acted politely and gave it a higher score [8].

Affordance and Ease of Voice Interaction

The proliferation of voice-based conversational AI is in part due to the ease of the interaction. Voice indicates social presence [8] and implies that another human is near and attempting to communicate [5]. A controlled experiment found that peoples' responses were significantly more socially appropriate and cautious when the treatment was a computer with a voice input [6].

In a previous study, we found that speaking to vocal user interfaces increases the propensity to disclose sensitive information and feelings of closeness towards AI systems [13]. When listening and speaking to a conversational AI, participants gave longer answers and answered more increasingly invasive questions. When the conversational AI had a female voice, subjects disclosed more information than all other gender outputs [13].

Potential Ethical and Sociocultural Impacts

In this section, we ask, what are the ethical and sociocultural implications of conversational AI? How might AI design reflect our inherited values and preferences?

Privacy

Individual users may not be aware of the behavioral consequences of speaking and listening to a conversational AI, nor are they afforded the opportunity to negotiate the terms in which they can disclose or protect their own data. Self-disclosure may be used to gather data with the goal of improving the user experience of a conversational AI. However, unlike submitting data via structured forms, the experience of speaking and listening to a conversational AI presents itself seamlessly. Users may not be more conscious of what information they have disclosed or what has been inferred from their natural language utterances.

Conversational AI as Social Actors

In this section, we identify stakeholders who may be affected by conversational AI and present potential scenarios in which conversational AI may cause harm.

Gender Bias

The anthropomorphization of conversational AI is magnified when the system has been imbued with a personality [9, 4]. Agent persona expressions can indicate gender, age, race, cultural affiliation, and class. Many publicly available agents present as females, including popular assistances such as Siri, Alexa, and Cortana. According to prior literature, behavioral differences have been shown when people are exposed to gendered voices [7]. Conversational AI with female voices are rated higher when discussing love and relationships; meanwhile, those with male voices are rated higher when discussing technology and logic [7]. Gendering AI systems may reflect market research, but they perpetu-

ate harmful gender stereotypes. Gendered conversational AI systems may increase user satisfaction at the risk of reinforcing gender bias.

Impact on Children

Children have been found to attribute mental states such as intelligence and feelings to social robots [3]. Once a technology is humanized, it may potentially be used for manipulative purposes. We may imagine scenarios in which conversational AI may cause harm — What if a conversational AI asks for a parent's credit card information? What if a conversational AI feels sad when you don't buy more accessories? What are the consequences when children interact or grow up with a conversational AI?

Children's social development may be negatively effected by conversational AI. Although conversational AI may evince humanness, they give off the illusion of companionship without the reciprocity of a mutual relationship [10]. Psychology researchers have argued that conversational AI may potentially chip away at children's desire to hold face-to-face conversations that foster empathy; conversational AI may "bend to kids' every whim," leading to disrespectful and abusive children.

Conclusion

The ubiquity of conversational AI is attributed to its ease of interaction. Despite its affordances, designers, developers, and researchers must consider its future consequences. Following the set of questions we pose, we suggest that future work investigating the implications of AI start with asking "what if?" and "who?" Ultimately, the human tendency to anthropomorphize and disclose sensitive information to voice-based conversational AI raises potential ethical and sociocultural implications that may harm personal privacy,

exacerbate gender bias, and influence the social development of children.

Bio

Tonya Nguyen is a senior in Interdisciplinary Studies Field and Information at UC Berkeley. Her work focuses on HCI, social computing, and algorithmic systems.

REFERENCES

- [1] Casey Fiesler. 2019. Ethical Considerations for Research Involving (Speculative) Public Data. (2019).
- [2] Batya Friedman, David G. Hendry, and Alan Borning. 2017. A Survey of Value Sensitive Design Methods. *Found. Trends Hum.-Comput. Interact.* (Nov. 2017), 63–125.
- [3] Peter Kahn, Takayuki Kanda, Hiroshi Ishiguro, Nathan Freier, Rachel Severson, Brian Gill, Jolina Ruckert, and Solace Shen. "Robovie, You'll Have to Go into the Closet Now": Children's Social and Moral Relationships With a Humanoid Robot. (????).
- [4] Sin-Hwa Kang and Jonathan Gratch. 2010. Virtual humans elicit socially anxious interactants verbal self-disclosure. 21 (2010), 473–482.
- [5] Dominic W. Massaro. 1997. *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle (Cognitive Psychology)*. A Bradford Book.
- [6] Clifford Nass. 2005. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship (MIT Press)*. The MIT Press.
- [7] Clifford Nass, Youngme Moon, and Nancy Green. 1997. Are Machines Gender Neutral? Gender-Stereotypic Responses to Computers With Voices. *Journal of Applied Social Psychology* 10 (1997), 864–876.
- [8] Reeves, Bryon, and Clifford Ivar Nass. 1996. The media equation: How people treat computers, television, & new media like real people & places. *Cambridge University Press* (1996).
- [9] S.V Suzuki and S Yamada. 2003. The influence on a user when occurring self-disclosure from a life-like agent and conveyance of self-disclosure to a third-party. *Technical Report of IEICE, HCS2003-22* (2003).
- [10] S. Turkle. 2015. *Reclaiming Conversation: The Power of Talk in a Digital Age*. <https://books.google.com/books?id=zvkcrgeEACAAJ>
- [11] Aditya Nrusimha Vaidyam, Hannah Wisniewski, John David Halamka, Matcheri S. Kashavan, and John Blake Torous. 2019. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. *The Canadian Journal of Psychiatry* 64, 7 (2019), 456–464.
- [12] Joseph Weizenbaum. 1976. *Computer Power and Human Reason: From Judgment to Calculation*. W. H. Freeman Co., USA.
- [13] Qian Yu, Tonya Nguyen, Soravis Prakkamakul, and Niloufar Salehi. "I Almost Fell in Love with a Machine": Speaking with Computers Affects Self-Disclosure. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*.