Invested Interest in Their Voice Assistant's AI Model: A Sociomaterial Perspective on Owners' Attitudes and Experiences

Mckie, Indra University of Technology Sydney, Australia | indramckie@gmail.com

ABSTRACT

Voice assistants (VAs) are a unique type of semantic media that transects visual, oral, and audio interaction and, at the same time, is heavily personified in its design, and often anthropomorphised by its users (Moriuchi, 2021). VAs are dependent on the use of AI models to provide an accurate response to a user's inquiry. This paper uses semi-structured interviews to examine the attitudes and experiences of long-term users of VAs regarding improving the AI model of their devices to help their user experience of VAs. Findings suggest that long-term users have mixed sentiments toward providing their data for improving the AI model. It also reports on experiences with feedback mechanisms reformatted to voice mediums. The attitudes towards AI transparency and interpretations of how users contributed to improving their VAs are discussed. Implications of the research provide insight into how users perceive their AI VAs and how they are motivated and action continuous improvements to VAs through their everyday interactions.

KEYWORDS

Voice assistants; conversational user interface; sociomateriality; AI transparency; AI data collection

INTRODUCTION

Conversational agents (CAs), depending on their type of interactional user interface, are commonly known as chatbots or voice assistants. CAs are a unique type of semantic media that transects visual, oral and audio interaction and, at the same time, is heavily personified in its design and often anthropomorphised by its users (Moriuchi, 2021) while still operating as a text retrieval system in the backend. Oral and audio-based CAs are referred to as voice assistants (VAs) in the literature and magnify the anthropomorphic effect of CAs' production of an audible human voice.

LITERATURE

Another core perception of CAs and VAs is the system's dependence on artificial intelligence (AI) algorithms. Continuously improving algorithms create better user experiences, for it is an inherent assumption that improved response accuracy leads to better prediction algorithms (Knijnenburg et al., 2012). The more users interact with the VA, the more data the VA collects about them and the VAs' ability to predict what interactions their users prefer, how their users talk (to improve speech models, see Support.Google.com, 2022), or recognise individual users by analysing and storing the sound of their users' voices (i.e., Alexa's VoiceID, see Amazon.com, 2022). All these supposedly lead to a streamlined and pleasant experience in exchange for the collection of data and interference with the privacy of VA users. As well, like any other algorithm, data is collected to improve its AI model. While some brands attempt to be transparent about their methods (see Support.Google.com, 2022), part of those processes remain unexplained as proprietary information or fall into the AI black box category (Hargittai et al., 2020).

Fairness, accountability, transparency and explainability have been shown to determine how users perceive algorithms (Shin et al., 2020; Shin, 2021). However, more qualitative data is needed that illustrates how these factors play out in the context of voice assistant interaction and how it affects user interactions and behaviour. Zhang et al. (2020) discuss the problem of trust calibration, where users actively make sense of how to trust the information provided to them by an AI model. However, the experience of using VAs goes beyond interacting with their algorithm. Companies behind VAs put out a sophisticated voice user interface—packaged with a supplementary app, email updates, marketing, and social media presence. Based on their holistic experience, how are users engaging with their voice assistants and, by extension, the brands that produce them?

RESEARCH DESIGN

For the research design of this study, Orlikowski's sociomateriality theory (2007) was used as a lens to understand the data from ethnographic semi-structured interviews. Ethnography informs the style of research conducted, focusing on the qualitative accounts that provide context-rich explanations of the research subject. Using ethnography, we can draw abductive conclusions from a small but highly contextual research sample. These do not generalise all VA users but uncover novel instances of attitudes and experiences regarding the improvement of the AI model of VAs. Sociomaterialism, written from Orlikowski's perspective, describes the intangible life in the transect of technology and organised life. Particularly how both components evolve in tandem and because of each other. Take this example of Google's search engine algorithm and the people it shapes, and that shape it,

Mid-Year Conference of the Association for Information Science & Technology | April 11-13, 2023, Virtual Event.

‘Google search engine is computer code that was produced and is maintained by software engineers, that executes on computers, and whose operation depends on the millions of people who create and update web pages everyday, and the millions of people who enter particular search terms into it’ (Orlikowski, 2007, p.1440).

Orlikowski’s sociomateriality is used as a lens on the lived experiences of twelve long-term users of voice assistants in their homes who have used them for more than six months. The length of VA ownership ranged from 6 months to 1-4 years. We examine the stories of our participants through this lens to discover intersections of human behaviour changing and transforming the materiality of AI voice assistants. Participants were recruited via callouts at various physical and online venues. After data collection, it was found that the participants were the primary owners of their VAs; this identity may have skewed the data towards a positive sentiment towards VAs (compared to interviewing VA skeptics or new users). Participants ranged in demographics, from age, perceived gender, socio-economic status, amount of voice assistants, and household composition.

Semi-structured interviews were conducted that asked about the understanding, perception and use of voice assistants in their household. As an exploratory case study, the interview protocol was developed to answer the research questions: 1) How do users construct an understanding of voice assistants, and how does this understanding shape their perception and use of voice assistants? And 2) How does our perception and understanding of embodied devices that speak in a human voice change how we seek, use and interact with information? Examples of the interview protocol included, ‘How did you come to own your VA?’ or questions such as, ‘Has your VA made you laugh? Tell us the story or show us the interaction which made you laugh.’ Participants reflected on their own experiences and their observation of interactions with their family members. These interviews were transcribed, then thematically analysed individually and again at an aggregate.

FINDINGS

This paper reports a case study of twelve open-ended semi-structured interviews with long-term voice assistant (VAs) users. During the interviews, we found that all participants were the owners of the devices, introducing VAs to their household members. During these interviews, we asked how these users developed their understanding of VAs, their perception of how VAs work and, based on these constructivist processes—how and why they use VAs in their homes.

Sentiment toward data collection

Findings suggest that as owners of VAs, some participants expressed an invested interest in continuously improving their VAs, as they actively participated in data collection and feedback mechanisms. Participants were incentivised by poor user experience to help train the AI or voice model. Terms like AI, voice model or algorithm that refer to the AI learning components of voice assistants were used interchangeably due to the lack of widespread technical knowledge on account of the newness of voice assistants.

> P3: Yeah, and I do, I have the interest of improving the voice model, I guess, or the response. Yeah, 'cause there'll be a few times where I've given "play this song on Spotify,” and it gives me the other end of the spectrum response; I’m like, that was shocking.

Participants’ responses to the interviews were mixed, and their beliefs and recollections of the experiences were diverse. Their sentiments conveyed themes of acceptance of data collection, invasion of information ownership and the significance of trading information for a service.

> P2: There's nothing you can hide from them. Well, if you use a platform, then basically, your data is there. There are always ways they can find the information from you, so... Why do I come to this topic? I was not worried that I'm putting too much information in Alexa. I don't think that's a concern. We just have to get used to what is the environment we're living in.

P4 offered an interesting perspective to those who might not welcome using VAs based on the implications of privacy invasion and data collection.

> P4: I just don't think I buy into all of that, that paranoia that like, oh, you shouldn't have these devices, 'cause it's always listening in... I think it's just... To me, I think this is progress in terms of what these devices can do; I think it's just those skeptical or maybe people that are resistant to change.

These excerpts are a small insight into how users of VAs perceive the tradeoff of personal data and constant surveillance for the service provided in return. As long-term users of voice assistants, their experience shows that owners of VAs share a similar sentiment towards the collection of their data.

Transparency of sociomaterial change

VA feedback mechanisms varied in transparency; for example, our participants spoke about using traditional rating systems prompted by their voice assistant. Given the novel emphasis on voice interaction, supplemented by visual

and touch interfaces, brands behind VAs must think outside traditional media methods (such as email, reviews or surveys) of obtaining user feedback to improve the user experience of their devices. Like other rating and review systems used online, P3 recalled answering a prompt on their linked smartphone from the Google Home app after interacting with their smart speaker VA.

> P3: Google will send you a follow-up. They'll say, I heard you say this. How happy were you with the response? So, [it's] kind of learning as well, which I think is interesting.

It was clear to P3 that Google was inviting him to improve their product in some way. From a user's perspective, it's unclear how their rating changes or improves their VA's AI based on the messaging from Google's simple notification and rating system.

> P3: So it's like it kind of knows that it's stuffed up, and I'm like, Yeah, no, that was not the mark, like one star for that one... Yeah, it's like one out of five.

P3's remarks suggest some awareness of the VA's low confidence score and how this triggered the 1 out of 5 stars feedback mechanism. Confidence scores are a common indication in AI algorithms to show how the AI performs. This standard metric is used in supervised training to improve the AI's ability to predict the correct answer (Zhang et al., 2020). At this point, confidence scores are not made available to the user during VA interaction, nor are available in the user's data on the app or manual data requests.

Another method described by our participants was the proactive confirmation response by VAs; for instance, Alexa would ask if she had understood you correctly, requiring a second level of confirmation for a potentially low confidence score. VAs conduct disambiguation dialogue based on the confidence score of the original request (Avila et al., 2020). This may be manually embedded or automatically based on the logic of a chatbot's code. Communicating this question directly after the interaction is another way a rating system has been reformatted for the voice medium. While both rating systems presented by Google are transparent and other methods are less so, the effects on the AI model are not.

> P11: This is the first level, I guess, of artificial intelligence that has been something that's accessible to our homes, and it's not quite sharp enough at the moment, but I'm guessing it's learning.

There was a general understanding from participants about the AI being able to learn. As general lay users of VAs, there were multiple interpretations of what this meant regarding their user experience of VAs.

> P3: I think that it's quite intelligent, and for me, it's a lot of the behind the scenes. Like how does it know that I'm saying a particular word? Where does it pull that information from? How does it put those words together and form an instruction, and then how does it execute that properly? To me, that's just crazy.

CONCLUSIONS

This paper presents a diversity of beliefs on how our participant's interactions with their voice assistants improved or altered the conversational AI system and, by proxy, their user experience. There was some sentiment from participants of receiving a mutual benefit from these voice assistants that may inform why VA users choose to exchange their data for the services provided by their AI models.

ACKNOWLEDGMENTS

I would like to acknowledge the Gadigal People of the Eora Nation and the Boorooberongal People of the Dharug Nation upon whose ancestral lands the University of Technology Sydney now stands. We would also like to pay respect to the Elders both past and present, acknowledging them as the traditional custodians of knowledge for these lands.

I also thank my PhD supervisors for their mentorship and support, Dr. Bhuva Narayan, Dr. Baki Kocaballi, and Dr. Gail Kenning. This research is supported by Australian Government Research Training Program Scholarship (RTPS) and the University of Technology Sydney Doctoral Scholarship.

REFERENCES

Amazon.com (2022). *What Is Alexa Voice ID?* Retrieved from https://www.amazon.com/gp/help/customer/display.html?nodeId=GYCXKY2AB2QWZT2X

Avila, C. V. S., Franco, W., Maia, J. G. R., & Vidal, V. M. (2020, June). CONQUEST: A framework for building template-based IQA chatbots for enterprise knowledge graphs. In *International Conference on Applications of Natural Language to Information Systems* (pp. 60-72). Springer, Cham.

Hargittai, E., Gruber, J., Djukaric, T., Fuchs, J., & Brombach, L. (2020). Black box measures? How to study people's algorithm skills. *Information, Communication & Society*, *23*(5), 764-775.

Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., & Newell, C. (2012). Explaining the user experience of recommender systems. *User modeling and user-adapted interaction*, 22(4), 441-504.

Mckie, I., Narayan, B., & Kocaballi, B. (2022). Conversational voice assistants and a case study of long-term users: A human information behaviours perspective. Journal of the Australian Library and Information Association, 71(3), 233-255.

Moriuchi, E. (2021). An empirical study on anthropomorphism and engagement with disembodied AIs and consumers' re-use behavior. *Psychology & Marketing*, *38*(1), 21-42.

Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization studies*, *28*(9), 1435-1448.

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, *146*, 102551.

Shin, D., Zhong, B., & Biocca, F. A. (2020). Beyond user experience: What constitutes algorithmic experiences?. *International Journal of Information Management*, *52*, 102061.

Support.Google.com (2022). *Learn how Google improves speech models*. Retrieved from https://support.google.com/assistant/answer/11140942?hl=en#zippy=%2Cconventional-learning

Zhang, Y., Liao, Q. V., & Bellamy, R. K. (2020, January). Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 295-305).