The Ethics of Generative AI & Conversational Agents

Abstract Submission (496 words)

Exploring the role of value-sensitive designs for large-language model-based voice assistants targeting people with dementia

Rasita Vinay[1,2], Stefan Klöppel[3], Mathias Schlögl[4] & Tobias Kowatsch[2,5,6]

[1] Institute of Biomedical Ethics & History of Medicine, University of Zurich, Switzerland
[2] Centre for Digital Health Interventions (CDHI), School of Medicine, University of St. Gallen, Switzerland
[3] University Hospital of Old Age Psychiatry and Psychotherapy, University of Bern, Switzerland
[4] Department of Geriatric Medicine, Clinic Barmelweid, Switzerland
[5] CDHI, Institute for Implementation Science in Health Care, University of Zurich, Switzerland
[6] CDHI, Department of Management, Technology & Economics, ETH Zurich, Switzerland

In recent years, we have seen a rise in digital health interventions (DHIs) to improve public health outcomes and promote healthy aging among older adults, while reducing the economic burden [1–3]. Within dementia research, the impact of DHIs has been explored on caregivers, family members and people with dementia (PWD) [4–10], with results indicating that such interventions can be safely and positively used among this population. With an aging society, more older adults are exposed to voice assistants daily, which are a subset of technologies that are being used as a form of DHIs. Since the introduction of Apple's Siri in 2011 and Amazon's Alexa in 2015, voice assistants have become largely integrated into daily life. Voice assistants are becoming more ubiquitous with older adults due to their low cost, high usability and accessibility, and user-friendly interactions that do not require the dexterity and good eye-sight needed for screen-based technologies [11]. Reported outcome measures utilizing voice assistants with older adults have included positive effects on well-being, such as relieving stress and anxiety [11], and on their quality of life (QoL) [12–14]. Additionally, incorporating large language models (LLMs) into voice assistants can help improve conversational aspects such as different speech characteristics, personalized responses, and stimulating human-like interactions [15]. With the growing adoption of voice assistants in healthcare, there is a burgeoning interest in leveraging these technologies with LLMs to improve lifestyle and cognitive outcomes for PWD. However, it is imperative to carefully consider the ethical implications and balance the values associated with their utilization, such as those associated with autonomy and dependency, dignity, design and deception, privacy and data security, misinformation, and bias and discrimination [15–18].

We aim to first discuss the importance of value-sensitive design for ethical development of such technologies, using LLM-based voice assistants as a case study. We leverage an ethical framework which informs the design of AI systems, such as the Ethics by Design for AI (EbD-AI) [19] to provide discussion on potential prototype designs for use by people with dementia. The AI aspect of LLM-based voice assistants incorporates sophisticated algorithms and machine learning techniques to enable natural language processing and intelligent responses, enhancing user interactions through personalized and contextually relevant assistance [20]. The EbD-AI framework (adopted by the European Commission as part of its ethical review process for AI projects), translates core moral values for AI into design requirements for AI systems [19]. With its established roots in computer science, bioethics, and responsible innovation, it becomes the most suitable framework for assessing value-sensitivity of LLM-based voice assistants as a DHI.

Despite the benefits of using voice assistants in dementia research, there remain significant gaps in our understanding of how to design, optimize and ethically implement these interventions [15,16,21]. In our research, we hope to demonstrate how currently existing design frameworks for DHIs may be utilized for ethical design of voice assistants for PWD, and the roles of using dynamic checklists for development and deployment of trustworthy AI [16,22].

References:

[1] Gulliford M, Alageel S. Digital health intervention at older ages. Lancet Digit Health 2019;1:e382–3. https://doi.org/10.1016/S2589-7500(19)30194-3.

[2] Vinay R, Probst J, Huynh P, Schloegl M, Kowatsch T, Nissen M. Top-funded digital health companies offering lifestyle interventions for dementia prevention: Company overview and evidence analysis [PREPRINT]. MedRxiv 2024. https://doi.org/10.1101/2024.03.29.24305069.

[3] Schneider C, Nißen M, Kowatsch T, Vinay R. Impact of digital assistive technologies on the quality of life for people with dementia: a scoping review. BMJ Open 2024;14:e080545. https://doi.org/10.1136/BMJOPEN-2023-080545.

[4] Liu L, Miguel Cruz A, Ruptash T, Barnard S, Juzwishin D. Acceptance of Global Positioning System (GPS) Technology Among Dementia Clients and Family Caregivers. J Technol Hum Serv 2017;35:99–119. https://doi.org/10.1080/15228835.2016.1266724.

[5] Gibson G, Dickinson C, Brittain K, Robinson L. The everyday use of assistive technology by people with dementia and their family carers: A qualitative study. BMC Geriatr 2015;15:1–10. https://doi.org/10.1186/S12877-015-0091-3.

[6] Sriram V, Jenkinson C, Peters M. Carers' experience of using assistive technology for dementia care at home: a qualitative study. BMJ Open 2020;10:e034460. https://doi.org/10.1136/BMJOPEN-2019-034460.

[7] Lorenz K, Freddolino PP, Comas-Herrera A, Knapp M, Damant J. Technology-based tools and services for people with dementia and carers: Mapping technology onto the dementia care pathway. Dementia 2019;18:725–41. https://doi.org/10.1177/1471301217691617.

[8] Yousaf K, Mehmood Z, Saba T, Rehman A, Munshi AM, Alharbey R, et al. Mobile-Health Applications for the Efficient Delivery of Health Care Facility to People with Dementia (PwD) and Support to Their Carers: A Survey. BioMed Res Int 2019;2019:7151475. https://doi.org/10.1155/2019/7151475.

[9] Davies A, Brini S, Hirani S, Gathercole R, Forsyth K, Henderson C, et al. The impact of assistive technology on burden and psychological well-being in informal caregivers of people with dementia (ATTILA Study). Alzheimer's & Dementia : Translational Research & Clinical Interventions 2020;6:e12064. https://doi.org/10.1002/TRC2.12064.

[10] Stara V, Vera B, Bolliger D, Rossi L, Felici E, Di Rosa M, et al. Usability and acceptance of the embodied conversational agent anne by people with dementia and their caregivers: Exploratory study in home environment settings. JMIR Mhealth Uhealth 2021;9. https://doi.org/10.2196/25891.

[11] Oewel B, Ammari T, Brewer RN. Voice Assistant Use in Long-Term Care. Proceedings of the 5th International Conference on Conversational User Interfaces, CUI 2023 2023:10. https://doi.org/10.1145/3571884.3597135.

[12] Yaghoubzadeh R, Kramer M, Pitsch K, Kopp S. Virtual Agents as Daily Assistants for Elderly or Cognitively Impaired People. Proceedings of the 13th International Conference on Intelligent Virtual Agents (IVA 2013), Edinburgh (UK): 2013, p. 79–91. https://doi.org/10.1007/978-3-642-40415-3_7.

[13] Maia JC, Coutinho JFV, de Sousa CR, Barbosa RGB, Mota FR do N, Marques MB, et al. Assistive technologies for demented elderly: a systematic review. Acta Paul Enferm 2018;31:651–8. https://doi.org/10.1590/1982.

[14] Daly Lynn J, Rondón-Sulbarán J, Quinn E, Ryan A, McCormack B, Martin S. A systematic review of electronic assistive technology within supporting living environments for people with dementia. Dementia 2019;18:2371–435. https://doi.org/10.1177/1471301217733649.

[15] Piñeiro-Martín A, García-Mateo C, Docío-Fernández L, López-Pérez M del C. Ethical Challenges in the Development of Virtual Assistants Powered by Large Language

Models. Electronics (Basel) 2023;12:3170. https://doi.org/10.3390/electronics12143170.
[16] Hastings J. Preventing harm from non-conscious bias in medical generative AI. Lancet Digit Health 2024;6:e2–3. https://doi.org/10.1016/S2589-7500(23)00246-7.
[17] Vinay R, Spitale G, Biller-Andorno N, Germani F. Emotional Manipulation Through Prompt Engineering Amplifies Disinformation Generation in AI Large Language Models [PREPRINT]. ArXiv 2024. https://doi.org/10.48550/arXiv.2403.03550.
[18] Gilbert S, Harvey H, Melvin T, Vollebregt E, Wicks P. Large language model AI chatbots require approval as medical devices. Nature Medicine 2023 29:10 2023;29:2396–8. https://doi.org/10.1038/s41591-023-02412-6.
[19] Brey P, Dainow B. Ethics by design for artificial intelligence. AI and Ethics 2023;1:1–13. https://doi.org/10.1007/S43681-023-00330-4.
[20] Gupta M, Kumar R, Sardalia H. Voice Assistant Technology: The Case of Jarvis AI. 2023 4th International Conference for Emerging Technology (INCET) 2023. https://doi.org/10.1109/INCET57972.2023.10170362.
[21] Howes J, Denier Y, Vandemeulebroucke T, Gastmans C, Howes JaredmichaelHowes J, Howes J. The Ethics of Electronic Tracking Devices in Dementia Care: An Interview Study with Developers. Science and Engineering Ethics 2024 30:3 2024;30:1–29. https://doi.org/10.1007/S11948-024-00478-0.
[22] Goldberg CB, Adams L, Blumenthal D, Brennan PF, Brown N, Butte AJ, et al. To do no harm — and the most good — with AI in health care. Nat Med 2024;30:623–7. https://doi.org/10.1038/s41591-024-02853-7.