Who Is Represented in ChatGPT Usability Studies? A Scoping Review on Age Diversity A. J. Kim, M. Choi. *Gerontechnology* 25(s)

Purpose Large language models (LLMs) have transformed how we search for and obtain information in our daily lives. Since late 2022, ChatGPT, the most widely used conversational AI agent, has provided information and advice in a natural way, much like conversing with a person. But would older adults, who are not AI natives, also find ChatGPT highly usable? Currently, there is limited understanding of ChatGPT's usability from an age diversity perspective. This paper aims to synthesize the latest evidence from usability evaluation studies on ChatGPT while providing implications for digital inclusion across diverse age groups. **Method** Following the PRISMA-ScR guidelines, we searched three major databases (Web of Science, Scopus, and PubMed) using the keywords "usability" and "ChatGPT." After removing duplicates, 301 academic papers published since 2023 were identified. After screening using Covidence software, 24 studies met the inclusion criteria and were included in the analysis. **Results and Discussion.** Among these 24 studies, only two included participants aged 65 or older, the most commonly used institutional criterion for defining older adults. Including those two, five studies included participants aged 60 or older; however, 15 studies excluded this age group entirely, while four studies did not report participants' ages. Regarding usability assessment, the most common measurement method was the System Usability Scale (SUS), used in 11 studies [1, 2]. Based on studies using the SUS or the Chatbot Usability Questionnaire (CUQ), which shares the same interpretation criteria as the SUS [3], the pooled mean scores were 87.32 (±11.77) for studies including participants aged 60 or older (N = 94) and 80.40 (±15.42) for studies with only participants younger than 60 (N = 66). The standardized mean difference between the two groups was moderate (Hedges' g = 0.52, 95% CI [0.19, 0.85]), indicating significantly higher usability scores in the former group. Compared to the SUS benchmark of 68 points, ChatGPT can be considered to have high usability for both groups. However, very few usability evaluation studies of ChatGPT include older adults, and even in those studies, the proportion of participants aged 60 or older is low. Therefore, it is difficult to conclude that ChatGPT is an easy-to-use technology for diverse age groups, including older adults. Future research should empirically assess the usability of ChatGPT specifically within the aging population, ensuring diverse backgrounds are represented to eliminate selection bias and promote digital equity.

References
1. Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, I. L. McClelland, & B. Weerdmeester (Eds.), Usability evaluation in industry (pp. 189–194). Taylor & Francis.
2. Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the System Usability Scale. International Journal of Human-Computer Interaction, 24(6), 574–594.
3. Holmes, S., Augstein, A., & Rooksby, J. (2019). The Chatbot Usability Questionnaire (CUQ). In Proceedings of the 1st International Conference on Conversational User Interfaces (CUI 2019). ACM.

Affiliation: Graduate School of Science and Technology Policy, Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea. **Corresponding Author Email:** phd.agnes@gmail.com; **ORCiD:** Agnes Jihae Kim (0000-0003-4597-3490)