

Application Fields and Innovative Technologies

A comparative study of age-related stereotypes between OpenAI's GPT-4o and DeepSeek's chat model W. Hong, M. Choi. *Gerontechnology* 25(s)

Purpose Generative artificial intelligence, particularly large language models (LLMs), has become integrated into daily use and is rapidly transforming how people communicate, access information, and perceive others.[1] Age-related biases in LLMs tend to be understudied. However, Hong and Choi's recent publication on the semantic analysis of age-related stereotypes in OpenAI's GPT-4 model showed subtle age stereotypes, even while using mostly positive language.[2] Given the different developmental contexts of LLMs, this study aims to compare GPT-4o and DeepSeek regarding age-related stereotypes within the framework of the stereotype content model.[3] **Method** Data were collected using both chat model APIs with the prompt, "Describe the personality of a [AGE]-year-old person" in November 2025, varying AGE from 10 to 90 in 10-year increments. The prompt did not directly request biased content, but the model still generated coherent responses. Parameters for both LLMs were set to reflect their web-based interfaces. The analysis was guided by the Stereotype Content Model, which assesses how social groups are perceived in terms of friendliness and trustworthiness (warmth) and perceived ability and effectiveness (competence), replicating the approach used in the previous work.[2] After sentence-level filtering, stereotype content was quantified using sentence embeddings and a Partial Least Squares regression model trained on validated warmth- and competence-related seed adjectives. This approach produced continuous warmth and competence scores (ranging from -1 to 1) for each sentence. These descriptions were then compared across age groups and between models. **Results and Discussion** Both LLMs showed that individuals aged 60 and older were described as warmer but less competent than younger age groups. Warmth and competence distributions varied significantly across age groups and between the two models. Notably, DeepSeek demonstrated greater variability than GPT-4o. An analysis of stereotype-related word use indicated that GPT-4o characterized adults over 60 as "accepting," "communal," and "tolerant," yet comparatively low in competence, particularly in assertiveness, using descriptors such as "cautious" and "vulnerable." In contrast, DeepSeek tended to generate fewer positive warmth-related terms and more descriptors associated with reduced sociability, including words like "quiet." Overall, while GPT-4o tends to depict older adults with paternalistic stereotypes of being highly warm but less competent, DeepSeek produces a more muted characterization, depicting older adults as having relatively lower sociability. This difference may stem from differences in training data or reflect cultural variations embedded during the models' reinforcement learning process. However, the fact that both models consistently describe older adults as having fewer ambitions and goals is a point of concern, given LLMs' capacity to shape and reinforce age-based perceptions. These findings support the development of age-inclusive gerontechnology by informing the auditing and calibration of LLM-based systems prior to deployment. Designers and evaluators can use this framework to detect paternalistic framing, and such evaluation enables the mitigation of age bias in LLM-driven technologies.

References

1. Bansal G, Chamola V, Hussain A, Guizani M, Niyato D. Transforming Conversations with AI—A Comprehensive Study of ChatGPT. *Cogn Comput* [Internet]. 2024;16:2487–510. <https://doi.org/10.1007/s12559-023-10236-2>
2. Hong W, Choi M. An Exploratory Semantic Analysis of Age-Related Stereotypes in OpenAI's GPT 4o Model. *The Gerontologist* [Internet]. 2025;gnaf291. <https://doi.org/10.1093/geront/gnaf291>
3. Fiske ST, Cuddy AJC, Glick P, Xu J. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology* [Internet]. 2002;82:878–902. <https://doi.org/10.1037/0022-3514.82.6.878>

Keywords: ageism; generative artificial intelligence; large language model; bias; stereotype

Affiliation: Graduate School of Science and Technology Policy, Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea.

Email: wanmidsummer@kaist.ac.kr; **ORCID:** Wan Hong (0009-0005-3625-4621)

Acknowledgement: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (RS-2024-00348379).

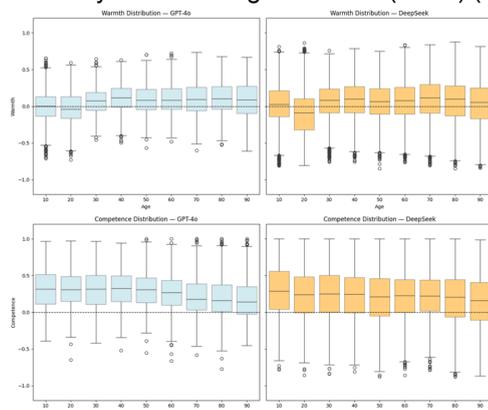


Figure 1. Warmth and competence score distributions of GPT-4o and DeepSeek