Variants of knowledge-based chatbots in family caregiving – Human evaluation and comparison

I. Nowak, D. Zühlke, N. Nikzad-Khasmakhi, D. A. Kring, T. Grüb-Okkan, J. Steinborn, I. Zorn

Purpose In Germany, 82% of people in need of care are cared for at home by informal carers. 80% of informal carers have no previous care-related training and 57% are over 60 years old (BARMER 2018). People in need of care as well as their caregivers need reliable information regarding their rights, support structures, or potential ressources for enhancing life with care. Limited accessibility of professional care advice services further exacerbates the need for reliable information resources. Family care is characterized by its complexity and the variety of conditions and specific requirements. Informal carers are ever more in need of easy to access information about coping in life with care for the elderly and are not content with existing care counseling (Bohnet-Joschko/Bidenko 2019). To deliver such information existing knowledge from a variety of sources must be classified in terms of its reliability, credibility and seriousness (Hoffmann et al. 2021). Older people in particular show deficits in digital research skills when it comes to locating, evaluating and using online information (Initiative D21 2021). Against this background, a knowledge-integrated chatbot based on a Large Language Model (LLM) is being developed to provide reliable information on legal entitlements, benefit entitlements and their application procedures. Given the increasing prevalence of LLMs and their application in sensitive areas such as family care, we present results of a comprehensive human evaluation of the effectiveness of different methods of knowledge integration into chatbot technologies. The aim is to increase the accuracy of the information provided and minimize so-called hallucinations, i.e. factual inaccuracies. **Method** Knowledge sources about family care were collected and structured. Two knowledge integration models were programmed. Our evaluation focused on the assessment of 24 generated answers of the chatbot by family care experts in terms of content fidelity, correct language, factual accuracy and completeness of chatbot generated answers to specific questions about family care. The objective was to assess the effectiveness of two different knowledge integration models in terms of accuracy and reliability of answers. **Results and Discussion** The preliminary results show that knowledge-integrated chatbots exhibit a significant increase in performance in the answer generation process and effectively draw on diversified knowledge sources. By providing valid answers, they enable access to reliable information from multiple data sources about care services, regardless of users' digital skills. Particularly noteworthy is the potential of multilingual and speech recognition systems to achieve a broader and more accessible information transfer. Valid answers, evidence-based and without hallucinations shall educate informal caregivers and people in need of care strengthen their decision-making skills, which may not only benefit the quality of care, but may also support self-determination and extend access to potential ressources to support life with care.

References

BARMER; Rothgang H.; Müller, R. (2018). Pflegereport 2018. Schriftenreihe zur Gesundheitsanalyse. BARMER GEK. Berlin. https://www.barmer.de/resource/blob/1028550/d7ae6a0758a91a938be14a0e05433b63/digitale-pressemappe-barmer-pflegereport-2018-data.pdf

Bohnet-Joschko, S.; Bidenko, K. (2019). Was pflegende Angehörige wirklich brauchen. Erste Ergebnisse der deskriptiven Analyse zum ZipA-Projekt. Witten. https://docplayer.org/190019124-Was-pflegende-angehoerige-wirklich-brauchen.html

Hoffmann, M.; Schwarz, C.; Sendlhofer, G. ; 2021. Patienten und Angehörige richtig informieren. Springer Fachmedien. Wiesbaden. https://doi.org/10.1007/978-3-658-35274-5

Initiative D21 (2021). Digital Skills Gap. https://initiatived21.de/uploads/03_Studien-Publikationen/Digital-SKills-Gap/digital-skills-gap_so-unterschiedlich-digital-kompetent-ist-die-deutsche-bevoelkerung.pdf

Keywords: large language models, hallucination, family caregiving, chatbot, response accuracy

Affiliation: Institute for Media Research and Media Pedagogy; TH Cologne; Germany
Email: isa.nowak@th-koeln.de; **ORCID iD:** 0009-0000-4415-734X

Acknowledgement: The authors would like to thank the German Federal Ministry of Labour and Social Affairs for supporting the research in the project "BOTschafft Inklusion: Der inklusive Chatbot zur Eingliederung und Teilhabe von pflegenden Angehörigen in den Arbeitsmarkt" (#CIP.00.00012.23). This research is making an essential contribution to the understanding of external knowledge sources for LLMs and their influence on answer accuracy.

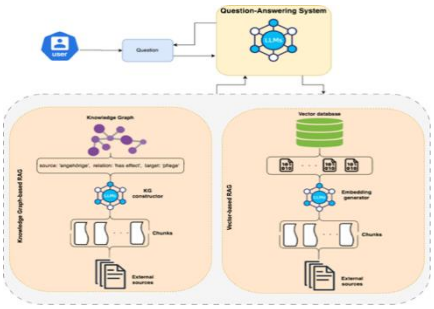

Figure 1. An overview of question answering systems with RAG