Guidelines for Human-AI Interaction
Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz

Reviewed by: Rishabh Devgon

Critical Review:

Saleema Amershi et al. [1] propose a set of 18 concise and generalisable design guidelines pertaining to Human Centered Artificial Intelligence (HCAI). These guidelines are instrumental in investigating knowledge gaps, and highlighting opportunities for future research endeavours. The paper keeps its focus on the primary aim and does not deviate from the central theme. In addition to the aforementioned, the paper also does justice to its claims and highlights the multi levelled implications of AI-infused systems. Due to their characteristically unpredictable and inconsistent nature, it further provides motivation for human centred intervention, explainability and a platform for further elaboration on the subject. The literature review covered a lot of themes and modern paradigms with respect to AI and other analogous fields; however, there was a lack of focus on the AI pipeline and ladder that could aid the research giving a more structural and holistic approach. Each of the guidelines was further facilitated with an example to make the guidelines clearer and more intuitive.

The paper provides a four pronged methodology that gives further insights and is written in a way that captivates the reader. It is robust in nature and each step of the research is put forth in exceptional detail with tangible outcome and conjecture. Methods used for the research in question were critical literature review, modified heuristic evaluation, user study through artefact analysis and later expert evaluation. The entire process is justified and takes an iterative approach. It is thoroughly calculated in its analysis and offers a novel probe into the existing guidelines for Human AI interaction. The selection criteria in every step of the research design were clearly defined and justified. The recruitment was done through snowball sampling and the recruited practitioners were diverse in several ways that added to the participatory nature of the sampling. This included the participants' experience, role, age and region. However there has been no specification of their socio, cultural or economic positionality. Had they named the continents and the countries in their list of specifications, readers could have been provided with further insights into whether or not the Global South was taken into consideration when the research was conducted.

The paper situates itself well with a great balance between researchers and practitioners. This has been achieved by incorporating both in different parts of the study and thus bridging the gap between the research and actual industrial implementation. The phase 4 of the research seemed to be a little ambiguous and not elaborated on thoroughly. The lack of clarity is due to the experience of the experts being almost indistinguishable to the experience of the participants in phase 3 of the research. These participants also belonged to the same organisation and thus fell within a limited purview. So, a case could be made about how the results obtained from the study cannot be generalised and additionally, could be biased as they come from a singular institution with shared beliefs. An additional method that I would have added to this already extensive research, would be to deploy user studies with a diverse research set. More specifically, through surveys and interviews to get a deeper and more holistic perspective of the guidelines and their applicability and generalisability.

I felt that the research lacks focus on the ethics of AI with respect to its interaction with humans. A subjective explanation revolving around the working of AI-infused systems based on the exposure, experience or even expertise of the different types of users involved could have been added to the list of guidelines. An additional theme that could have been explored was the agency and responsibility in case of errors caused by AI systems. The guidelines provided in the paper could have expanded and provided a deeper insight into socio cultural and subjective sub themes. Additionally, it could be made more universal and pluralistic if extensive user studies were taken into consideration on a global scale. To finally conclude my review, although the paper lacks clarity in certain aspects of its guidelines and how the interaction with the users takes place, it does still explore a set of fundamental guidelines that are essential to future practitioners. These guidelines can act as key components to future research by enabling a myriad of work in the domain of HCAI.

References:

[1] S. Amershi *et al.*, ‘Guidelines for Human-AI Interaction’, in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, May 2019, pp. 1–13, doi: 10.1145/3290605.3300233.