

**Questioning the AI:
Informing Design Practices for Explainable AI User Experiences
Q. Vera Liao, Daniel Gruen, Sarah Miller**

Reviewed by: Rishabh Devgon

Critical Review:

Q. Vera Liao et al. [1] through their research, hope to contribute towards insights into design practices to make Artificial Intelligence (AI) more explainable, identify opportunities to better support this work and reflect upon Explainable AI (XAI) opportunities for the future. The authors delve into what makes for a good XAI system, the agency of which should lie on the lay users without deep technical knowhow and not the algorithm-centric view. The study investigates the problems that industry practitioners face while designing explainable AI systems. There is background provided on various subthemes pertaining to XAI: XAI as a field, supporting AI practitioners by studying work and identifying guidelines for XAI contributions and Question driven explanations. The results section deliberated about how to convey to the user about the functioning of the model and how transparent this particular phenomenon should be without the information getting overbearing or leading to legal or marketing concerns.

The research was not novel in terms of the theme that it explored but rather the novelty came from the methodology adopted by the authors; in the way of investigation for themes related to XAI. The authors employed an alternative taxonomy to map user questions. However, there was no justification provided as to why aspects such as complexity of the model, or applicability to specific models were disregarded. These parameters are predefined, and thus, a shift from it should warrant an explanation for the same. The research echos the title of the paper and provides an in-depth insight into the functioning and dynamics of the user-centred part of XAI. It also provides a means to make AI more explainable and condenses explainability into a question bank for benchmark and parameterisation purposes.

The researchers employ an AI algorithm informed question bank, which acts as a novel probe to their investigation. The research method used by the authors was similar to a card sorting analysis in conjunction with probing. This method seems to be a combination of traditional card sorting analysis and semi-structured interviews. This method selection seemed interesting to me, given that it provided select themes to explore and also had the capability to incorporate any additional themes that may come up. The analysis was done through open coding and axial coding from the grounded research theory, which is an iterative process and thus creates a dynamic feedback loop for further probes. This method is relevant and justified and offers the researchers more flexibility to modify their research questions.

This paper too suffers from the same limitation where all recruits belong to the same institution, ie. IBM, hence this brings about a bias in the research because of institutional values, power dynamics and homogenisation. Additionally, the social, economic, cultural and regional aspects of the informants have not been specified. The authors do talk about recruiting from 7 different

locations, but do these locations represent a global picture or are they also representative of the global north? The research was also not representative of the ground reality because it was second-hand knowledge from UX researchers and designers who essentially acted as user proxies, and thus this could cause appropriation of the user's perspective. Since the paper focuses on explainability for users, it's absurd that it does not investigate the user's themselves. Hence, I would propose the authors also to recruit real users. The selection criterion is not defined and seems rather loose, and there is no particular definition of the roles of people that they have recruited. The sampling method is not mentioned either. The confidentiality and anonymity of participants have not been addressed in this paper. The authors could have also provided their positionality statements for the purposes of this literature. All of these factors added to the work could immensely help the research, in my opinion.

The authors mention the use of MURAL, and hence I would have liked to see the work done there to make this method clearer and more descriptive. This could have been added in an appendix. The section does not talk about subjectivity in terms of explainability at all, i.e. how different users might require various explanations and thus undermines pluralism. The paper offers an insight into the XAI scenario but does not really provide recommendations and guidelines to change the current dynamics. The study could be further extended to advocating for a favourable XAI ecosystem and how that can be achieved using this question bank as a framework. Overall it was an excellent introductory paper to the domain of XAI as it offered in-depth background and knowledge about the field and added a condensation and evaluation mechanism through the means of the question bank.

References:

- [1] Q. V. Liao, D. Gruen, and S. Miller, 'Questioning the AI: Informing Design Practices for Explainable AI User Experiences', in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, Apr. 2020, pp. 1–15, doi: 10.1145/3313831.3376590.