

Researching AI Legibility through Design

Joseph Lindley, Haider Ali Akmal, Franziska Pilling, Paul Coulton

Reviewed by: Rishabh Devgon

Critical Review:

Joseph Lindley et al. [1] adopt a Research through Design (RtD) approach to explore make AI's role in a system explainable and coherent to the user. The authors focus on three significant aspects in this study: reviewing previous work on the AI legibility problem space, reviewing design proposals aiming to enhance AI legibility and explain the use of RtD in the conjunction of AI and HCI research. The authors provide a brief and elegant summary of historical and present date AI. The paper sheds light on HDI: Human data Interaction and looks at AI's socio-technical implications through the lenses of legibility, agency and negotiability. The researchers highlight the role of signs and symbols in the modern era and provide an overview of the RtD process. The work derives from a lot of previously done research and thus is not novel but deploying an RtD approach in this domain is something that is new and interesting.

The related work delves on the duality of the word Artificial Intelligence and investigates both weak and strong pillars as it explores the epistemic understanding of AI. Legibility and Human Data Interaction further dives into the legibility aspect of HDI and what it constitutes. This subsection, however, does not address why agency and negotiability were not investigated further in this paper, especially because of how these three values are so intertwined together. The next segment focused on Guidelines for Human AI Interaction. These guidelines seem to focus on a totalizing view of explainability in AI and do not address the subjectivity and diversity of all actors. The next subsection talks about Transparency, Interpretation and Understandable AI. A particularly interesting discussion that I find in this area which the paper touches upon, is the motivation to create such explainable systems and how this directly correlates to the very fundamental human emotion of trust. This study extends IBM's research about the FactSheets Question bank and talks about its adoption in various use cases. There is constant stress on interdisciplinary collaboration throughout this paper but no basis or recommendation of how to structure such a collaboration and what kind of break up to aspire towards.

The next section talks about design for legibility and initially explores AI iconography, which is fundamental for improving AI legibility by offering visual representation. Here they have taken an iterative approach and explained their selection and insights along with exploring various tangent themes resulting in tangible outcomes. The authors provide a comprehensive explanation and justification for each step of their design process. However, since there has been no usability testing of their outcomes at any stage of designing, the final set of icons and signs selection could potentially suffer from a belief and confirmation bias due to the empathy gap.

The research method used is a critical literature review and analysis of case studies which is justified given that this is a survey paper. The researchers could have specified their selection

criteria for the literature to provide further insight into the methodology that they adopted. The authors have employed an RtD approach in this paper, which was justified given how the authors evaluated and presented their arguments for the visual components. I felt as if the user's perspective through various ethnographic methods definitely could have been incorporated, especially when evaluating legibility for design and the various visual aids associated with it. The approach that the authors have taken seemed to have created a departure from the human-centred methodology. Putting the user at the centre would have further added a more holistic and representative view of the topic. The authors also do not really talk about their positionality in this paper explicitly, which could have helped us to understand the limitations of the paper.

I felt that the paper lacked a lot of substance when it came to the actual design interventions, which could make AI more legible. The exploration of just visual form factors like signs and icons seemed to be somewhat reductionist to the massive set of requirements and aspirations of modern users and other stakeholders of AI. Hence, a lack of empathy with the user was pretty evident here. The paper could have further explored many more UX concepts. This could range right from the perception of AI to an exploration into the types of interactions. Further, the study could have better addressed the subjectivity of different user groups and used a bottom-up approach. Additionally, there was no focus on HDI in the research design, which seemed to be more of an afterthought with preliminary discussion. To conclude my review, I really appreciated the RtD approach employed by the researchers. However, I felt that this specific approach took away from the user's agency and thus needed to evolve further to uncover and explore various other aspects to make AI more legible.

References:

- [1] J. Lindley, H. A. Akmal, F. Pilling, and P. Coulton, 'Researching AI Legibility through Design', in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, Apr. 2020, pp. 1–13, doi: 10.1145/3313831.3376792.