Closing the Creator-Consumer Gap in XAI: A Call for Participatory XAI Design with End-users

Sunnie S. Y. Kim
Department of Computer Science
Princeton University
suhk@cs.princeton.edu

Elizabeth Anne Watkins*
Intelligent Systems Research
Intel Labs
watkins.elizabethanne@gmail.com

Olga Russakovsky
Department of Computer Science
Princeton University
olgarus@cs.princeton.edu

Ruth Fong
Department of Computer Science
Princeton University
ruthfong@cs.princeton.edu

Andrés Monroy-Hernández
Department of Computer Science
Princeton University
andresmh@cs.princeton.edu

Abstract

Despite the proliferation of explainable AI (XAI) methods, little is understood about end-users' explainability needs and perceptions of XAI explanations. To address this gap, we interviewed 20 end-users of a real-world AI application, the Merlin app for bird identification, and found that participants' AI background and domain interest play a critical role in shaping their XAI needs and perceptions. Further, participants exposed gaps in existing methods and offered valuable solutions. In this position paper, we reflect on our findings and make a call for participatory XAI design with end-users, towards developing methods that serve the needs of diverse end-users and closing the creator-consumer gap in XAI.

1 Introduction

With AI applications supporting more areas of daily life, hundreds of explainable AI (XAI) methods have been developed to make AI models more transparent and understandable to human users. However, arguably these are being developed without embracing the full spectrum of end-user needs. Particularly, for modern neural network-based AI models with millions of model parameters, translating model decisions into understandable insights is so challenging that new XAI methods are frequently limited by what XAI researchers *can do* rather than what end-users *might need*.

To connect XAI development with end-users, we studied end-users' explainability needs and perceptions of existing approaches in a real-world context where XAI methods might be deployed. Our research setting is Merlin, an AI-based mobile phone application that identifies birds in user-uploaded photos and audio recordings. We chose Merlin because it is a widely-used application that allows us to connect with a diverse set of active end-users. Concretely, we interviewed 20 end-users of Merlin who span the range from low-to-high AI background (representing both consumers and creators of

*Most work done while at Princeton Center for Information Technology Policy and HCI Program.

Accepted as a panel presentation to the Human-Centered AI (HCAI) Workshop at the 36th Conference on Neural Information Processing Systems (NeurIPS 2022).

AI systems) and low-to-high domain background (representing both users who know *less* and *more* about birding than the AI system).

In short, we found that end-users' XAI needs and perceptions greatly depend on their particular background and interest in AI and the application domain. Further, end-users surfaced gaps in current XAI research and offered valuable suggestions. In this position paper, we reflect on our findings and make the case for *participatory XAI*, especially involving end-users in the XAI design process, towards developing "explanations (XAI) that serve the needs of diverse end-users" (workshop call).

2 Findings: Participants' XAI needs and perceptions varied based on their AI background and domain interest

XAI needs. In the interviews, we inquired about participants' explainability needs through open-ended questions and a survey we developed based on the XAI Question Bank [3]. Overall, participants were curious about AI system details (e.g., what data the AI was trained on, and how the AI makes decisions). However, there were group differences in levels and types of explainability needs.

Participants with *high-AI background*, who work with AI systems in their day-to-day life, expressed a high need for technical transparency and were willing to learn more about Merlin's AI by emailing the app developers and playing with relevant data. Participants with *high-domain interest*, regardless of AI background, were also very curious about Merlin's AI. They especially wanted to know how the AI identifies birds that are difficult for experienced human birders, e.g., "little brown birds" and mockingbirds, because they sought to learn from the AI and improve their bird identification skills. Conversely, participants with *lower AI background and domain interest* were interested in learning about AI system details, but didn't want to go out of their way to find the information. Some even preferred keeping the AI as a black box, saying *"I don't want to ruin the mystique."*

XAI perceptions. Next, to study how participants perceive existing XAI methods, we mocked up representative approaches that could be embedded into Merlin,[2] and gathered participants' feedback. Here we describe results for two approaches—heatmap and concept-based—where participants' perceptions varied based on their AI background. See Fig. 1 for their mock-ups.

Heatmap-based explanations, while deemed intuitive in the AI research community, received mixed reviews from participants. Participants with *high-AI background*, who often use heatmaps in their work, found them intuitive and helpful for representing information. However, participants with *low-AI background* expressed a strong dislike. One participant remarked, *"I hate those things [...] They are simply not intuitive."* Another participant didn't like them as explanations because *"heatmaps feel like they should be related to weather,"* revealing individual differences in perception.

Similarly, opinions diverged for concept-based explanations. Some participants with *low-AI background* found them confusing. One participant said, *"stuff like this would go right over my head and make no sense for the most part."* Contrarily, participants with *higher-AI background* wanted even more numbers, concepts, and other details about how the AI makes its predictions. Still, they acknowledged that such explanations might overwhelm users who have less AI expertise.

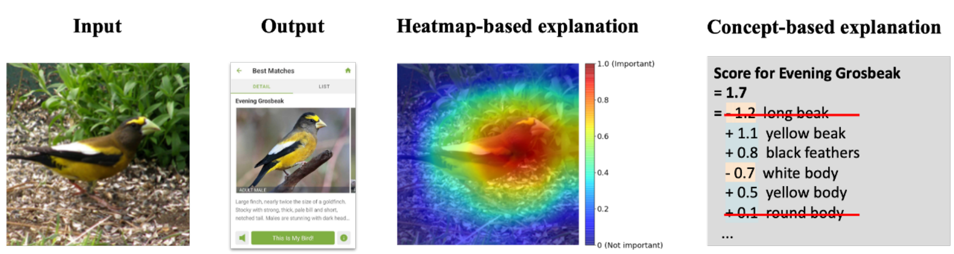

Figure 1: Example of Merlin's identification (real) and corresponding explanations (mock-ups).

[2]Merlin does not have XAI features implemented yet, as with most real-world AI applications.

3 Discussion: The need for participatory design of XAI with end-users

Our study with Merlin end-users revealed a creator-consumer gap in XAI [2]. This gap is unsurprising, since XAI methods have been primarily developed for and used by AI developers to obtain technical transparency and debug AI models [1, 4]. But this gap is critical because end-users may have different explainability needs that XAI methods should but don't yet support.

Indeed, we found that participants with lower-AI background were not as interested in the inner workings of the AI system. Overall, participants wanted XAI explanations for more practical purposes: for determining when to trust the AI, but also for improving their task skills and for changing their behavior to supply better inputs to the AI, which are less discussed in the XAI literature. Further, participants with lower-AI background found some of the popular XAI approaches confusing, making a wake-up call to researchers and developers.

Participants also exposed blind spots in existing XAI methods and proposed solutions. For example, they pointed out that the concepts used in concept-based explanations were disconnected from birders' language: they were too generic compared to birders' field mark terms (e.g., wingbar, supercilium). To solve this disconnect, they suggested developing the bank of concepts with end-users, and offered to contribute their experience and domain expertise.

This example highlights the benefit and need for end-users' participation in the XAI design process. However, very little discussion exists on how to apply participatory approaches [5] to XAI research. Hence, our goal for the workshop is to start a discussion about the opportunities and challenges of participatory XAI and brainstorm together how the human-centered AI and XAI communities can develop "explanations (XAI) that serve the needs of diverse end-users" (workshop call).

References

[1] U. Bhatt, A. Xiang, S. Sharma, A. Weller, A. Taly, Y. Jia, J. Ghosh, R. Puri, J. M. F. Moura, and P. Eckersley. Explainable Machine Learning in Deployment. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, page 648–657, New York, NY, USA, 2020. Association for Computing Machinery.

[2] U. Ehsan, S. Passi, Q. V. Liao, L. Chan, I. Lee, M. J. Muller, and M. O. Riedl. The Who in Explainable AI: How AI Background Shapes Perceptions of AI Explanations. *arXiv preprint arXiv:2107.13509*.

[3] 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*, page 1–15, New York, NY, USA, 2020. Association for Computing Machinery.

[4] T. Miller, P. Howe, and L. Sonenberg. Explainable AI: Beware of Inmates Running the Asylum Or: How I Learnt to Stop Worrying and Love the Social and Behavioural Sciences. *arXiv preprint arXiv:1712.00547*.

[5] M. J. Muller. *Participatory Design: The Third Space in HCI*, page 1051–1068. L. Erlbaum Associates Inc., USA, 2002.