An Explainable AI User Interface for Facilitating Collaboration between Domain Experts and AI Researchers

Meng Shi*1*, Celal Savur*1*, Elizabeth Watkins*1*, Ramesh Manuvinakurike*1*, Gesem Gudino Mejia*1*, Richard Beckwith*1*, Giuseppe Raffa*1*

1 Intel Labs, Intel Corporation, USA

Abstract

The adoption of AI approaches is increasing across domains, yet the study of explainable AI user interfaces for domain experts remains limited. One potential benefit provided by this type of user interface is to facilitate better data collection and improve in-domain model training. With the advancement of explainable AI (XAI) methods for end-users, domain experts can more easily collaborate with AI researchers and contribute to the process of building and deploying domain relevant AI models. We propose an XAI interface for domain experts in a manufacturing setting, that provides transparency into multi-modal AI systems and supports domain expert collaboration with AI researchers to fine-tune models through active feedback. In this paper, we report early findings of a user study with this XAI interface, with participants including both domain experts and AI researchers. These early findings hold promise for supporting improved system understanding by end users as well as cross-functional collaboration between domain experts and AI researchers.

Keywords

Explainable User Interface, XAI for transparency, XAI for human computer collaboration

1. Introduction

The field of artificial intelligence (AI) has progressed remarkably in recent years. It enables new services, such as AI assistance systems which will be adopted not only by households, but also professionals. AI support can streamline operations, support human performance in working environments, and augment human capabilities across tasks.

However, AI assistance systems in these areas encounter very real problems due to a very limited amount of data for training, issues with the incorporation of domain knowledge, and difficulties in fine-tuning the models.

Having domain experts collaborate with AI researchers in development of the AI assistance system could potentially help alleviate the problems mentioned above, since expert feedback is a rich source of data incorporating domain knowledge. However, such collaboration requires the domain expert to also be familiar with the design and back-end knowledge of such AI assistance systems, which is almost impossible without an explainable AI approach.

In this paper, we report a user interface designed to support explainable AI (XAI) for a complex multi-modal AI assistance system with a goal of facilitating collaboration between domain experts in manufacturing (e.g., technicians on the factory floor) and developers working on multimodal ambient sensing systems for human performance and task support.

CEUR Workshop Proceedings (CEUR-WS.org)

2. Related Work

In the first few years of its history, the XAI field focused on developing algorithms to explain or interpret the behavior of "black box" models [4] by collecting and displaying information about, for example, parameters or weights which were especially significant in the production of a prediction or output. Most of this work presumed the consumers of explanations to be AI engineers or researchers looking to fine-tune or improve their own models. More recently, a subset of the field has taken a "human-centered" turn, focusing in part on how non-AI researchers might also benefit from XAI [1,2,3]. This turn included tailoring XAI interfaces and form factors for end-users rather than AI researchers [5,6].

Our contribution builds on this work to facilitate collaboration between two groups: AI researchers and one type of end-user, users who have expert-level knowledge of their domain of work, i.e., domain experts. Facilitating collaboration between these two groups can address key challenges of data availability, data quality, and the resource demands of model fine-tuning. We propose an XAI interface through which domain experts collaborate with AI researchers to give direct input into AI models. A key innovation is the step of actionability: we intend our UI to support end-user action and cross-functional collaboration, in the form of providing feedback on model performance provided by domain experts, to improve system performance.

3. Task Guidance System

We developed a POC (proof of concept) XAI user interface prototype for a multi-modal AI system providing performance support in a manufacturing scenario. As this prototype is a kind of perceptually guided task guidance system, we call it "Task Guidance System" (TGS). TGS, as shown in Fig. 1, is designed to facilitate collaboration between domain experts and AI researchers in fine-tuning the various components of the multi-modal AI system.

The TGS visualizes multiple models used by the multi-modal AI supporting system for Action Recognition (MSTCN[7]), Automatic Speech Recognition (Whisper[8]), Intent Recognition (BERT [9]), etc., which infer the user's actions and predict their next steps. The visualization is represented as a dashboard that uses module design principles for all the AI models being used.

The TGS can help domain experts understand the AI models. With TGS, the domain experts can see how their inputs are taken by the models and compare predictions generated by each model to the phenomenon the predictions represent. This is possible because TGS visualizes how each model analyzes its inputs and presents the output of each model.

In addition, TGS allows domain experts with ground-truth knowledge to critique these predictions through its dialogue system and, thereby, lets AI researchers fine-tune the AI models. To support this collaboration, prompts encourage users to provide feedback at the models' request. These prompts include an uncertainty score for vision, an importance score for noting which words in an utterance were most indicative of a match, highlighted inconsistencies between different models, etc.

The contribution of this work is significant because the obstacles mentioned earlier may hold us back from building systems that can help humans achieve their full potential. The XAI UI we propose is intended to directly address these obstacles.
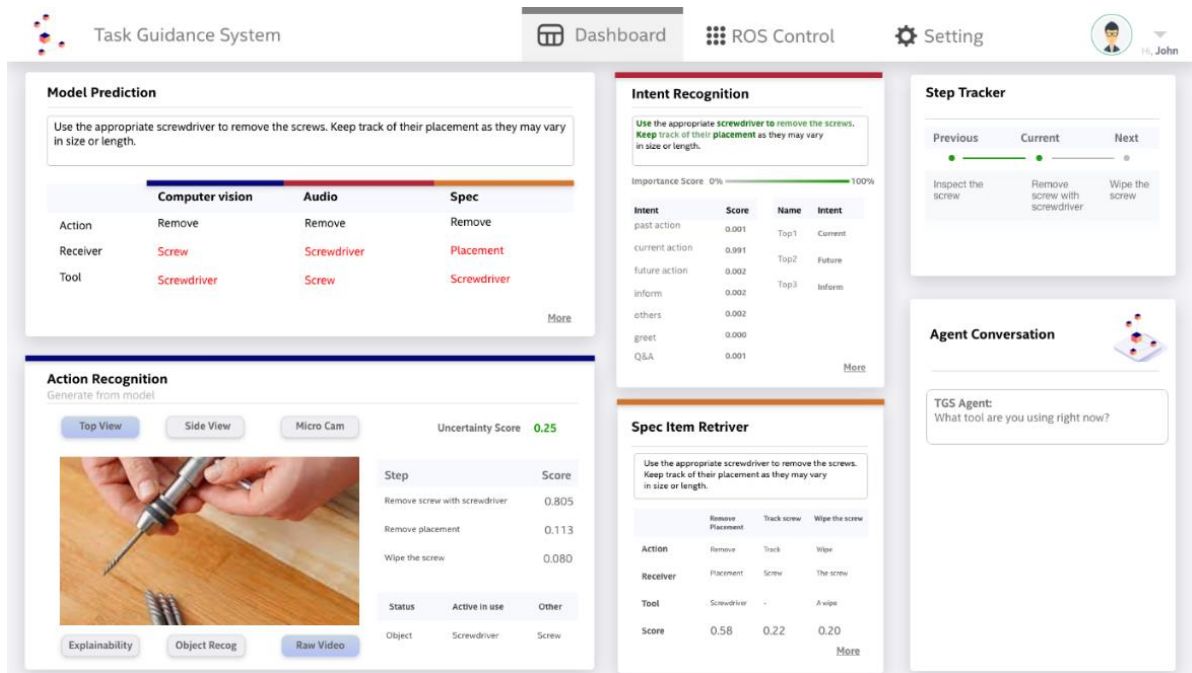
Figure 1: Visualization of each individual model's performance (with, e.g., uncertainty score, importance score and inconsistent outcome) to enable appropriate user feedback (Original images and content replaced due to confidentiality)

4. Methods: Two User Studies

The XAI UI was evaluated remotely with two user studies, one with domain experts and one with AI researchers. Wizard-of-Oz presentations of the UI were combined with qualitative interview protocols for both studies.

The first study was conducted with two domain experts interacting with the AI assistance system including the XAI UI. Both experts had several months of experience working with the AI system (without the XAI UI). A semi-structured interview protocol was designed and carried out to elicit and capture data about the domain experts' cognitive model of system and TGS, whether the information displayed had any impact on their understanding of model predictions, or their willingness to provide feedback to AI researchers.

The second study looked at three AI researchers experienced in multi-modal AI assistance systems. A second semi-structured interview protocol was designed to elicit and capture perceptions of the UI, what domain knowledge might be important for their goals as AI researchers, and how they feel this knowledge can be adequately captured from domain experts through TGS. All interview results were analyzed collaboratively by two of the authors using thematic analysis.

5. Results

All five interviews took place remotely, via a video conferencing platform. Most interviews took between 30 and 90 minutes, with one outlier Wizard/interview process taking roughly two hours.

AI researchers reported that they anticipate deriving important benefits from domain experts' feedback, especially in domains of high dynamism, complexity, or those with diverse users.

AI researchers confirmed that they could derive value from capturing domain experts' knowledge. They expressed that this data could be particularly valuable for fine-tuning models in domains characterized by high degrees of change, or domains or applications that may be made more complex by the need to serve multiple users or multiple different types of users.

Aligned multimodal data are valuable to be streamed together.

AI researchers report it is important to see the entire data flow from domain experts' raw performance data to the final output. For example, with a multi-modal AI system, one models' output may rely on another's output.

It is valuable to have direct access to domain experts' feedback.

Domain experts demonstrated a willingness to help AI researchers using TGS, and identified novel challenges AI researchers might face in negotiating a space of multiple or changing approaches in methods or objects. All domain experts reported that comprehending the information in the UI was challenging. However, even then, domain expert interaction with TGS provided valuable feedback for AI researchers. For example, domain experts pointed out instances where the task model was too specific and pointed to variants on the task that should be supported by the AI assistance system. Without access to the nuanced domain knowledge held by domain experts, AI researchers could face risks when fine-tuning their models of incorrectly labeling actions or objects that they might not know are acceptable alternative methods for accomplishing a task. We found that when domain experts interpret data differently than AI researchers, XAI information can reveal to domain experts that their input is needed to rectify an inaccurate model prediction.

We also found that domain experts' curiosity about model inferences was increased through their exposure to the XAI interface.

When domain experts saw incorrect model predictions, they verbally challenged the discrepancies between the prediction and their understanding of correctness of the process being analyzed: one domain expert asked, unprompted, "Why does [the UI] think I'm holding [item] when I'm clearly holding [a different item]?" when shown the TGS UI. We believe that this promises that further engagement will see an increased willingness to engage with and correct inaccurate predictions. This was confirmed by explicit inquiries about their willingness to provide feedback: both domain experts expressed that they would be both willing and enthusiastic to provide input back to AI researchers via the interface.

6. Future Work

The paper presents our ongoing XAI dashboard and findings of an early user study. A larger-scale user study will be performed to more robustly evaluate the visual content and affordances of the UI and dashboard. We plan to design additional XAI visualizations across multiple form factors, including popular approaches like heatmaps. We also plan visualizations plotting prediction history to identify outlier predictions and generate user-facing alerts. Our plans also include providing more options to edit models directly through TGS, which we hope will make fine-tuning models faster and more flexible. We hope our application can support scaling systems more quickly and easily across domains.

References

[1] U. Ehsan, P. Wintersberger, E. A. Watkins, C. Manger, G. Ramos, J. D. Weisz, H. Daumé Iii, A. Riener, and M. O. Riedl, Human-Centered Explainable AI (HCXAI): Coming of Age. In Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23). Association for Computing Machinery, New York, NY, USA, Article 353, 1–7. https://doi.org/10.1145/3544549.3573832

[2] U. Ehsan, P. Wintersberger, Q. V. Liao, E. A. Watkins, C. Manger, H. Daumé III, A. Riener, and M. O. Riedl, Human-Centered Explainable AI (HCXAI): Beyond Opening the Black-Box of AI. In Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22). Association for Computing Machinery, New York, NY, USA, Article 109, 1–7. https://doi.org/10.1145/3491101.3503727

[3] U. Ehsan, and M. O. Riedl, Human-centered Explainable AI: Towards a Reflective Sociotechnical Approach. arXiv preprint arXiv:2002.01092. https://doi.org/10.48550/arXiv.2002.01092

[4] F. Emmert-Streib, O. Yli-Harja, and M. Dehmer, Explainable Artificial Intelligence and Machine Learning: A reality rooted perspective. J. Cohen (Ed.), Special issue: Digital Libraries, volume 39, 1996. https://doi.org/10.48550/arXiv.2001.09464

[5] S. Kim, S. Choo, D. Park, H. Park, C. S. Nam, J. Y. Jung, and S. Lee, Designing an XAI interface for BCI experts: A contextual design for pragmatic explanation interface based on domain knowledge in a specific context. International Journal of Human Computer Studies, 174, [103009]. https://doi.org/10.1016/j.ijhcs.2023.103009

[6] S. S. Kim, E. A. Watkins, O. Russakovsky, R. Fong, and A. Monroy-Hernández, "Help Me Help the AI": Understanding How Explainability Can Support Human-AI Interaction. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 250, 1–17. https://doi.org/10.1145/3544548.3581001

[7] Y. A. Farha, J. Gall, "Ms-tcn: Multi-stage temporal convolutional network for action segmentation." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR '2019) (pp. 3575-3584). https://doi.org/10.48550/arXiv.1903.01945

[8] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, Robust speech recognition via large-scale weak supervision. In International Conference on Machine Learning (2023) (pp. 28492-28518). PMLR. https://doi.org/10.48550/arXiv.2212.04356

[9] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. (2018) https://doi.org/10.48550/arXiv.1810.04805