User-Centric Visual Analytics to Enhance Novice User Accessibility and Understandability of AI Applications

Research Abstract

Artificial Intelligence (AI) has emerged as a powerful tool that is so ubiquitous that it is being used in several domains ranging from home security systems to using assistive robots in medical operations. Despite its potential, the recent proliferation of large models and applications has extended the reach of deep learning to users with little to no knowledge of AI. While such systems do aid novice users in their everyday lives and makes decision making easier, it also raises concerns about potential harms that can be caused by autonomous agent due to differences in mental models. Consequently, promoting explainability and accessibility of AI models has become increasingly important. My doctoral research focuses on tackling these crucial AI challenges by adopting a human-centric approach which involves developing innovative interactive visualization tools aimed at both instilling trust in AI models and also leveraging AI to aid in decision making.

Previous Research

Visual explanations to foster trust in AI: I have been working on developing novel visualization systems to foster novice user's trust to help them better understand AI model's decision making. Existing systems explain ML models like CNNs, GNNs, or GANs, but there's a gap in HCI regarding RL explanations, despite RL's broad application. To address this, we developed PolicyExplainer [5], letting users query autonomous RL agents interactively with three key questions: Why a certain action was taken, Why not take another action, and When actions are typically taken. The interface offers visual explanations via surrogate modeling. Usability testing with HIV experts for drug recommendation showed coherent and informative explanations, indicating PolicyExplainer's effectiveness in fostering trust and interpretability among non-expert RL users.

Our ConceptExplainer [1] work further enhances users trust in image based models by capturing high level latent natural language features called concepts to explain image classifications to non-expert users who can interactively probe and explore the concept space to explain model behavior at the instance/class/global level. The system not only helped in identifying concepts that are important to classification but also identified bias and data issues in training dataset.

Vis+AI tools to increase user accessibility Making visualization interfaces accessible to a broad audience for decision making is imperative to make such tools more prominent and mainstream. My first work was in the field of computational journalism wherein I wanted to explore and differentiate how the different news media outlets covered the same news event using a trio of features - keywords, the entities involved and the underlying emotional biases present to convey the information. We employ a design study methodology to develop and evaluate a novel visual analytics platform, called NewsKaleidoscope [2], which combines interactive visualizations and natural language processing (NLP) techniques to analyze coverage diversity. Our results indicate that NewsKaleidoscope effectively supports analysis of coverage diversity both for news experts and news novices, though interestingly it provides different benefits to each user group (e.g., for news experts, it supports validating hypotheses and assumptions about news sites).

Large Language Models (LLMs) have surged in popularity for their capacity to execute ad-hoc Natural Language Processing (NLP) tasks via simple prompts, eliminating the need for AI/ML expertise. However, the effectiveness of these prompts varies greatly based on linguistic structure, context, and other semantic factors. To enhance user prompts, we introduced PromptAid [4], a tool that guides users visually by offering AI backed linguistic (keywords and paraphrasing) and contextual (k-shot) recommendations for refining prompts iteratively. PromptAid was found to be highly useful in generating high performing prompts across various NLP tasks such as topic classification, sentiment analysis and common sense QA by both novice users and prompt engineers.

Ongoing research

My ongoing research is currently focussed on making AI more accessible and usable to end users. One of the works is building an AI assisted authoring tool to assist authors to write game play narratives.

I am also currently working on building a visual interface to make smaller language models as capable as larger LMs to perform downstream reasoning tasks with human in the loop knowledge extraction and distillation. This work is informed studies performed to extract trustworthy rationales for knowledge intensive tasks done during an internship [3].

Future research

Recent tools such as ChatGPT, Claude and Dall-E has become truly ubiquitous and while my research underscores the importance of enhancing the transparency of AI to bolster user trust, my internship experiences have provided a different perspective. During my time, I witnessed firsthand the integration of GenAI tools in the development of chatbots capable of tasks such as summarization, question answering, and evaluations. It's evident that these tools are here to stay, prompting me to advocate for leveraging them in two key areas - **Improving accessibility for GenAI tools through visualizations.** In today's landscape, numerous data scientists in industry are immersed in the creation of GenAI applications. These endeavors typically entail establishing extensive networks of LLM chains to execute downstream tasks, and subsequently employing large-scale LLM agents for evaluations. Despite the availability of tools like Promptflow, many practitioners struggle to effectively monitor and manage these intricate chains and agents. My forthcoming research endeavors will concentrate on developing interactive interfaces designed to assist data scientists and GenAI experts in crafting their agents with greater efficiency and coherence.

Visual evaluation of GenAI and its impact on user trust. A very important aspect which I feel is being overlooked is how to evaluate these generated outputs and the impact of these generations on user trust. Currently LLM evaluation is an abstract topic and is still being researched on using different methods such as embedding based methods, fine tuning and even surprisingly LLMs as a verifiers. However in practice these evaluations have a very low threshold of being put into production which does cause concern especially in safety critical domains. I wish to develop methods and experimental designs to understand how are these generated outputs impacting user's trust. Are LLMs very convincing and if so what attributes makes them so appealing to end users. I also wish to dive deeper into building visual tools to improve on end user's agency and control over these tools.

In summary, I'm a technical HCI researcher driven by a passion for identifying real-world challenges individuals encounter daily. My approach involves devising intelligent systems that blend human and machine intelligence creatively to address these issues. Subsequently, I deploy these solutions to gather a deeper understanding of their real-world usage.

References

[1] Jinbin Huang et al. "Conceptexplainer: Interactive explanation for deep neural networks from a concept perspective". In: *IEEE Transactions on Visualization and Computer Graphics* 29.1 (2022), pp. 831–841.

[2] Aditi Mishra, Shashank Ginjpalli, and Chris Bryan. "News kaleidoscope: Visual investigation of coverage diversity in news event reporting". In: *2022 IEEE 15th Pacific Visualization Symposium (PacificVis)*. IEEE. 2022, pp. 131–140.

[3] Aditi Mishra et al. "Characterizing Large Language Models as Rationalizers of Knowledge-intensive Tasks". In: *Submitted to ACL, 2024* (2023).

[4] Aditi Mishra et al. "Promptaid: Prompt exploration, perturbation, testing and iteration using visual analytics for large language models". In: *Submitted to TVCG, 2024* (2023).

[5] Aditi Mishra et al. "Why? why not? when? visual explanations of agent behaviour in reinforcement learning". In: *2022 IEEE 15th Pacific Visualization Symposium (PacificVis)*. IEEE. 2022, pp. 111–120.