Improving the Utility and Usability of Visualization in AI-driven Scientific Discovery

Qianwen Wang, Sehi L'Yi, Nils Gehlenborg

Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA

Artificial intelligence (AI) is now used in a variety of research fields and has led to exciting scientific breakthroughs, such as predicting protein structures and designing drugs (Jumper et al., 2021). In AI-driven scientific discovery, a key challenge is to build connections between the knowledge learned by AI in the form of parameter weights and the knowledge of humans in the form of laws and theorems. As a result, an urgent need for explainable artificial intelligence (XAI) is emerging.

In XAI, visualization plays a critical role due to its efficiency in organizing and communicating information. For example, many XAI techniques employ visualizations to explain AI results (e.g., heatmaps that highlight important regions in input images). There are also many interactive visualization tools that reveal complicated model structures and working mechanisms of AI models, enabling scientists to understand, diagnose, and improve AI.

However, these visualizations are designed mainly for AI researchers and offer little to no support for domain experts such as biologists or clinicians. Failing to consider the perspective of users and the characteristics of the applied research domains can significantly undermine the utility and usability of visualizations, hindering efficient human-AI communication and decreasing the yields of AI-driven scientific discovery.

Including User Perspectives. One of the main challenges, surprisingly, lies in users' preference for visualizations, which can lead to over-trust of XAI. Prior studies documented that users appreciate visualizations to an extent that they tend to accept explanations without digging into possible issues and even rationalize suspicious observations (Bansal et al., 2020; Kaur et al., 2020). Such over-trust problems can be especially dangerous in scientific discovery, where the ground truth remains unknown and AI predictions require rigid evaluation conducted by domain experts. To promote the application of AI, visualizations should not only help users to understand and trust AI, but also help them to identify when an AI *should not be trusted*.

We should not assume that visualizations can always be interpreted easily and accurately. Many XAI visualizations are designed and developed by AI researchers. Visualizations that are straightforward for AI researchers are not necessarily comprehensible for users in the applied research domain. For example, Kaur et al. (Kaur et al., 2020) showed that few users can accurately interpret the visualizations generated by two popular XAI techniques, SHAP and GAM. Therefore, it is important to incorporate usability guidelines in visualization design and provide the necessary guidance for users to facilitate the interpretation of model predictions.

Considering Domain Characteristics. Effective XAI visualizations require a deep understanding of the applied domain. Over the years, researchers in different domains have developed their own practices of collecting data, formulating hypotheses, and deriving conclusions. The output of AI-driven scientific research is most likely to be maximized when the explanations and their visualizations can seamlessly fit into users' original workflows. Take a recent application of AI in mathematics (Davies et al., 2021) as an example. Instead of directly generating conjectures, which eliminates human intelligence and is technically challenging, AI guides the intuition of mathematicians by explaining the potential relations between mathematical objects with attribution visualizations. The visualization community has developed valuable design guidelines to extract needs in applied domains and transfer them to design

requirements of visualizations. However, adapting and applying these guidelines to XAI visualization in AI-driven scientific discovery require further investigation and exploration.

Apart from working practices, data used by the applied domains also poses challenges to the visualization design. Input data are widely used to explain AI predictions (e.g., explaining image classification using regions of the input image). Visualizing these explanations is non-trivial with complicated input data, which is the case in most research domains. For instance, when applying AI to single-cell biology, explaining AI predictions requires visualizing the large-scale and multi-modal data generated by single-cell analysis, such as RNA sequences, chromatin accessibilities, histone modifications, and spatial positions. Moreover, users can hardly derive meaningful insight merely based on single explanations. Instead, they need to summarize and compare different explanations and connect them with other contextual information. Novel visualization techniques are needed to better support the insight-deriving process, enabling effective exploration, organization, and contextualization of AI explanations.

Facilitating Human-AI Communication. We envision that visualization should also support two-way communication between AI and human users. In other words, users do not just interpret the knowledge learned by AI through reading visualizations, but also provide feedback to improve the AI by interacting with visualizations. Many AI studies recognize the benefits of incorporating human feedback, especially when the learning objectives are hard to define, the exploration space is too large, and the AI is employed in the constantly-changing real world. However, it can be challenging to design visualizations that support efficient human-AI communication. For example, it is overwhelming and frustrating for users to provide feedback on every single AI prediction. Therefore, user feedback should only be asked for when it is most beneficial for AI improvement. What is “*the most beneficial feedback*” can be automatically calculated by the AI, but can also be decided by the human users based on their knowledge and their interpretation of the visualizations. Furthermore, the visualization should incorporate mechanisms to motivate users and improve the feedback quality.

In sum, both challenges and opportunities exist for visualization in AI-driven scientific discovery. Impactful utility and usability of visualization require joint research efforts between visualization, AI, human-computer interaction, and domain researchers.

Reference:

- Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M. T., & Weld, D. S. (2020). Does the Whole Exceed its Parts? The Effect of AI Explanations on Complementary Team Performance. In *arXiv [cs.AI]*. arXiv. <http://arxiv.org/abs/2006.14779>
- Davies, A., Veličković, P., Buesing, L., Blackwell, S., Zheng, D., Tomašev, N., Tanburn, R., Battaglia, P., Blundell, C., Juhász, A., Lackenby, M., Williamson, G., Hassabis, D., & Kohli, P. (2021). Advancing mathematics by guiding human intuition with AI. *Nature*, 600(7887), 70–74.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589.
- Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H., & Wortman Vaughan, J. (2020). Interpreting Interpretability: Understanding Data Scientists’ Use of Interpretability Tools for Machine Learning. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14.