Visualization for AI Explainability

L. Miguel Encarnação, *Regions Bank, Birmingham, AL, USA*
Jörn Kohlhammer, *Technical University of Darmstadt, Darmstadt, Germany*
Chad A. Steed, *Oak Ridge National Laboratory, Oak Ridge, TN, USA*

The rapid increase in applications of artificial intelligence (AI) and machine learning (ML) algorithms exacerbates the need for explainable data visualization methods that make algorithmic guidance transparent and understandable. The availability of explainable AI/ML methods is important to a wide range of domains, especially in highly regulated domains such as banking and financial services, healthcare, transportation, and security and defense.

In banking, for instance, AI explainability not only allows data scientists to better validate their models, but it also assists technical regulators, who assess the risk of developed AI systems, and the end-users of such systems, who make decisions based on awareness of system limitations. In healthcare systems, on the other hand, explainable AI gives data scientists the ability to understand the efficacy of algorithms designed to personalize patient treatment and improve healthcare outcomes. At the clinical level, explainable AI methods assist in the evaluation of outcomes, automated interventions, and diagnostic methods, such as the pathology-focused technique described by Plass et al. in this special issue. Yet another growing application field is cyber security, where the rapid increase and growing variability of Internet traffic makes the monitoring of networks and cyber-attack mitigation more difficult and a high priority for companies and governmental agencies. The combination in cyber security of AI methods with explainability approaches allows experts to judge the trustworthiness of detection algorithms.

A significant challenge for explainable AI methods is the transformation of black-box AI and ML technologies into glass-box solutions that humans can understand, ideally trust, and effectively manage in practical applications. Successful deployments of explainable AI techniques, for instance, reveal the most important features for a model, such as model bias, performance measures related to drift and accuracy, and adoption risks. Notwithstanding, building trust in AI methods also requires responsibility, accountability, and fairness of algorithms. There is certainly still a long way to go from explaining AI and ML to data analysts to conveying trust and transparency to domain experts and end users in real-world applications. However, well-designed visualization techniques will continue to play an important role in enabling transparent and understandable computational assistance.

This special issue features articles on the human-centered design and use of user interfaces and data visualizations in support of making systems, which employ AI and ML models, easier to understand and more accurately to interpret, thus supporting their transparency and increasing trust in their application, whether it is during the design and development phase of a model, during its training and execution, or in a *post hoc* phase focusing on the use of models in practical applications.

Mennatallah El-Assady and Caterina Moruzzi are investigating the dynamics between the human users and AI systems toward designing effective, visual and nonvisual, communications between both agents. With the aim of contributing to the design of hybrid intelligence systems, they present potential biases and reasoning pitfalls and propose six desiderata for human-centered explainable AI.

In the context of their *SUBPLEX* system, Jun Yuan and collaborators describe an iterative design process involving expert machine-learning researchers and practitioners, resulting in steerable clustering and projection techniques to identify and select important features for an ML model's interpretation, which in turn are being associated with goals and tasks for explaining a machine-learning model using local explanations.

With *DETOXER*, Mahsan Nourani and her colleagues at the University of Central Florida focus their design goals on multiscope explainability, support for guided exploration, both multilabel and temporal comparisons as well as error detection for iteratively debugging and refining models for video activity recognition.

Finally, Markus Plass and his collaborators at the Medical University of Graz are studying the

0272-1716 © 2022 IEEE
Digital Object Identifier 10.1109/MCG.2022.3208786
Date of current version 13 December 2022.

decision-making of human experts in the context of medical pathology based on implicit knowledge and practical experience. They demonstrate how their findings can inform the design of effective, AI-driven multiscale visualizations of a pathologist's observation and decision path to support the pathologist's examinations of digital gigapixel images of scanned histopathological slides and microscope recordings.

We wish to thank Pak Chung Wong for his support in producing this special issue. We would also like to thank the authors and reviewers for their work and contribution. We hope you enjoy reading this issue and learn more about applying visualization to explainable AI.

L. MIGUEL ENCARNAÇÃO is the head of data visualization (SVP) for the Data and Analytics Organization, Regions Bank, Birmingham, AL, USA. His research interests and current work focus include user-centered visualization design, visual analytics literacy, and data-informed decision making. Encarnação received his Ph.D. degree in computer science from the University of Tübingen, Germany. He is the corresponding author of this article. Contact him at LME@computer.org.

JÖRN KOHLHAMMER is an honorary professor for user-centered visual analytics with the Technical University of Darmstadt, Germany. He is the head of the Competence Center for Information Visualization and Visual Analytics at Fraunhofer IGD, Germany. His research interests include decision-centered visualization and visual analytics. Kohlhammer received his Ph.D. degree in computer science from the Technical University of Darmstadt, Germany. Contact him at Joern.Kohlhammer@igd.fraunhofer.de.

CHAD A. STEED is a distinguished researcher with Oak Ridge National Laboratory, Oak Ridge, TN, USA, where he leads the Augmented Analyst Intelligence Group. His work is focused on interactive data science with an emphasis on multivariate visual analytics and he has applied these techniques in a myriad of domains. Steed received his Ph.D. degree in computer science from Mississippi State University. Contact him at csteed@acm.org.