EDITOR: Theresa-Marie Rhyne, theresamarierhyne@gmail.com

DEPARTMENT: VISUALIZATION VIEWPOINTS

Grand Challenges in Visual Analytics Applications

Aoyu Wu, *Harvard University, Cambridge, MA, 02138, USA*
Dazhen Deng, *Zhejiang University, Zhejiang, 310027, China*
Min Chen, *University of Oxford, OX1 3QG, Oxford, U.K.*
Shixia Liu, *Tsinghua University, Beijing, 100190, China*
Daniel Keim, *University of Konstanz, 78464, Konstanz, Germany*
Ross Maciejewski, *Arizona State University, Tempe, AZ, 85281, USA*
Silvia Miksch, *TU Wien, A-1040, Vienna, Austria*
Hendrik Strobelt, *MIT-IBM Watson AI Lab, Cambridge, MA, 02142, USA*
Fernanda Viégas and Martin Wattenberg, *Harvard University, Cambridge, MA, 02138, USA*

In the past two decades, research in visual analytics (VA) applications has made tremendous progress, not just in terms of scientific contributions, but also in real-world impact across wide-ranging domains including bioinformatics, urban analytics, and explainable AI. Despite these success stories, questions on the rigor and value of VA application research have emerged as a grand challenge. This article outlines a research and development agenda for making VA application research more rigorous and impactful. We first analyze the characteristics of VA application research and explain how they cause the rigor and value problem. Next, we propose a research ecosystem for improving scientific value, and rigor and outline an agenda with 12 open challenges spanning four areas, including foundation, methodology, application, and community. We encourage discussions, debates, and innovative efforts toward more rigorous and impactful VA research.

Visual analytics (VA) applications leverage software artifacts with VA techniques to solve real-world problems. They combine automated data analysis techniques with interactive visualizations to facilitate effective reasoning and decision-making on large and complex datasets.[1] Since the establishment of the IEEE Conference on Visual Analytics Science and Technology in 2006, VA applications have grown into a significant and influential research field in visualization. For example, research on VA applications occupies about 34% (41/120) of the full papers published in the IEEE Visualization Conference (IEEE VIS) 2022. VA applications have also achieved substantial societal impact by providing success stories on solving real-world problems, especially in wide-ranging high-impact domains, such as bioinformatics, urban analytics, explainable artificial intelligence (AI), and social media.

Nevertheless, decades of research and practices of VA applications have exposed multiple value- and rigor-related questions (see Figure 1). In light of these questions, we organized a panel at IEEE VIS 2022 to discuss grand challenges for making this research field more valuable and rigorous. The panel theme was informed by an informal interview study with 17 active researchers in VA system research.[2]

0272-1716 © 2023 IEEE
Digital Object Identifier 10.1109/MCG.2023.3284620
Date of current version 13 September 2023.

FIGURE 1. Questions on the rigor and value of VA application research have emerged as a grand challenge.

This article compiles interview results and panel discussions to provide the visualization community with a research and development agenda for rigorous and impactful VA application research. We first explain four dilemmas specific to this field and why the rigor and value problems exist. Subsequently, we propose a research ecosystem for increasing research value and rigor and discuss the 12 open challenges surrounding the proposed ecosystem. We hope our article will spark discussions and encourage efforts to help advance the field more vigorously and valuably.

DILEMMAS

We begin by discussing why the VA application research has rigor and value problems. We analyze its characteristics and argue that there exist fundamental dilemmas in the root causes behind the aforementioned provoking questions. We illustrate our idea with a simplified research lifecycle model[3] that represents the four key steps in research that includes: planning the objective, conducting research, publishing results, and enabling reuse (see Figure 2).

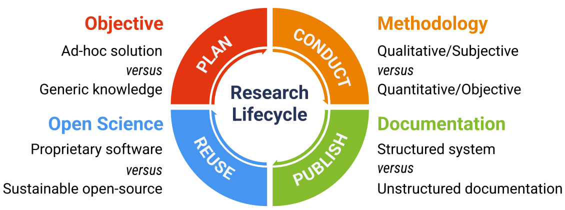

FIGURE 2. Four dilemmas surrounding the research lifecycle of VA application research that could cause the rigor and value problem.

1) *Objective*: The objective of VA application research is usually to create a specific solution to a domain-specific application problem, which creates tension against the drive of academic research to produce general knowledge. This tension leads to the frequently asked question, "What are the scientific contributions of VA applications?"
2) *Methodology*: The methodology is rooted deeply in design study,[4] which is qualitative and subjective in nature. Although it is certainly principled and valid, we need to be aware of the impetus of science and computer science for quantification and objectivity. Particularly, the evaluation of VA applications is sometimes perceived as anecdotal evidence, making it difficult for readers to assess their research value.
3) *Documentation*: The outcome is typically a structured, interactive, multiple-view visualization system. However, there is a lack of shared documentation standards for both academic papers and software artifacts. For example, there exists confusion among terms, such as tasks, goals, and requirements. The field also lacks a principled approach for readers and end users to comprehend the system workflow.
4) *Open science*: Despite the recent open science movement that aims to make research outcomes accessible to all levels of society, amateur or professional, limited progress has been made on VA applications. Only a small number of research studies provide publicly available code, and even fewer establish sustainable open source projects (e.g., continuous community involvement, such as commits, fixing typos or bugs, or commenting on an issue). Therefore, the users of VA applications are usually a limited

number of domain experts and there is huge potential to reach out to more workers in the data science workflow.

The abovementioned discrepancies, which we refer to as dilemmas, reflect fundamental problems in VA application research. By using the term "dilemmas," we emphasize that they are not binary questions with single, obvious answers. Instead, we advocate for an inclusive perspective to derive combined benefits from promoting the research value and rigor of VA applications. Some of our questions include: Can we derive generic knowledge from adhoc applications? Can we augment the current subjective methodology with objective factors? What is the roadmap for establishing, implementing, and unifying standards for VA systems and applications? And, how can we better position VA solutions in the data science ecosystem?

ECOSYSTEM: VALUE AND RIGOR

Faced with these questions, we outline a research and development agenda with 12 open challenges of VA application research (i.e., difficult but important problem sets to encourage solutions). To guide our discussion of those challenges, we systematically organize them by mapping them to the VA application research ecosystem. As shown in Figure 3, VA application research is regarded as a bridging joint between academia and practice, connected via two circles concerning rigor and value, respectively. The term "academia" refers to not only academic institutions but also other organizations that conduct fundamental research.

The rigor cycle bridges academia and VA application research. Components in academia, including foundation, methodology, and academic organization, provide scientific grounds for studying VA applications. In return, we hope that VA research activities can extend the knowledge in academia.

The value cycle connects practice with VA application research. VA application research is driven by real-world application problems, and successful solutions to domain problems generate socio-technical impact and value. Here we discuss two important factors in practice: making real applications and building a vibrant, wider community.

AGENDA: OPEN CHALLENGES

Based on the ecosystem, 12 open challenges were compiled from an informal interview study and a panel discussion. Figure 3 provides an overview of these challenges, organized by four themes: foundation, methodology, application, and community. We acknowledge that the challenges mentioned are not comprehensive, and there remains ample scope for further fundamental research, such as the development of innovative visualization techniques.

Foundation

We start with challenges in the scientific foundations of VA application research.

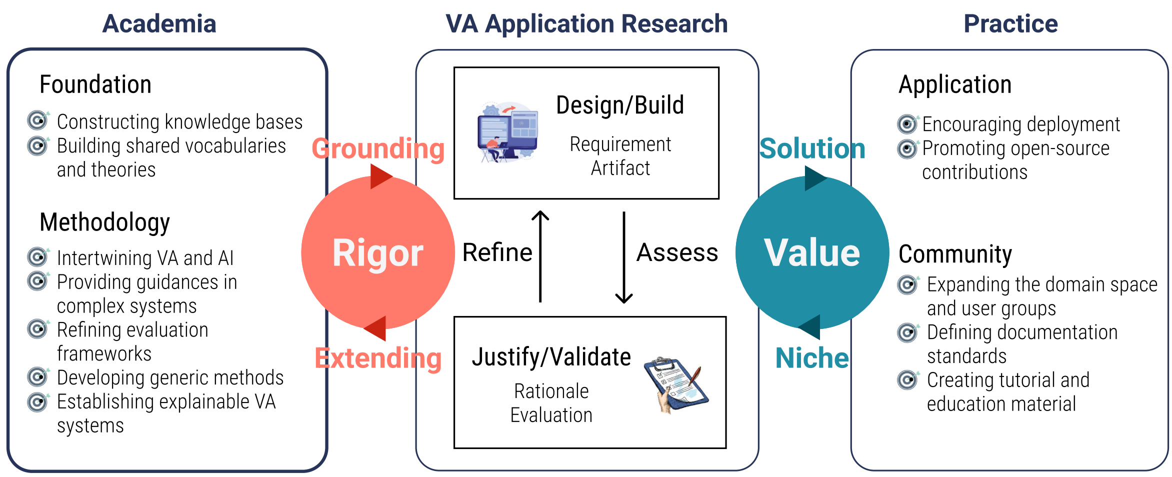

FIGURE 3. We propose a research ecosystem that connects VA application research with academia and practice through the rigor circle and the value circle, respectively. Surrounding this ecosystem, we outline a research and development agenda with 12 open challenges.

Constructing Knowledge Bases

Collecting real-world instances can improve our theoretical understanding of some fundamental questions in VA, such as why and how visualizations work, and how to measure the tradeoff among statistics, algorithms, visualizations, and interactions. Real-world instances will enable us to evidence or falsify existing theoretical postulations and simulate new conceptual and theoretical developments.

For example, Chen and Ebert proposed an ontological framework for designing, evaluating, and improving VA workflows.[5] They adopted the medical terms: symptoms, causes, remedies, and side effects to encourage a systematic methodology for optimizing the utilization and integration of different machine-centric (e.g., statistics and algorithms) and human-centric processes (e.g., visualization and interaction). For instance, given a symptom "a tag map may confuse font size with spatial coverage," researchers have proposed one remedy "use color instead of size to encode the numerical scales of keywords" in a design study. They articulated the need for a large collection of real world instances of symptoms, causes, remedies, and side effects, with which VA researchers and practitioners would be presented with new challenges and opportunities.

Based on this idea, it is a grand challenge to construct a VA encyclopaedia with three A–Z lists for symptoms (side effects), causes, and remedies, respectively, where entities in one list are cross referenced with the relevant entities in other lists. This VA encyclopaedia can be stored as an online knowledge base, providing a digital service to all VA researchers and practitioners, very much in the same way as what the field of medicine has encountered for centuries and millennia.

Building Shared Vocabularies and Theories

Many design requirement sections in VA papers will have generic statements akin to providing access to the raw data, allowing exploration at different scales, supporting brushing and linked views, providing real-time interactivity, etc. A well-designed mapping of design criteria between the problem and solution spaces is missing. Critically, the mapping needs to consider multiple aspects in VA, such as data and model characteristics, goals, and analytical tasks. Are there existing theories and mechanisms that can help us relate design spaces from one problem to another through some transformations? Likely, there needs a paradigm shift toward a science-led protocol with shared vocabularies.

As pointed out by Thomas and Cook[6] in the early days of VA research, a theory of VA is needed to define the building blocks of VA systems and their interfaces. Nice progress has been made,[5] but we are by no means close to having a complete and consistent "Theory of VA." This topic will remain a challenge in the years to come.

Methodology

Methodological challenges concern the system of methods used in designing, implementing, and evaluating VA applications.

Closer Intertwining VA and AI

The intertwining of VA and AI has recently evolved as one of the most exciting areas in our field. As AI continues to revolutionize numerous application areas, challenges of increasing understanding, mitigating bias, and managing human oversight are growing significantly. Often, the methods with the highest accuracy are the least understandable, making integration with VA solutions difficult.

A major challenge is whether these two fields can mutually broaden the theoretical foundation that is crucial for the further development of each field. For example, with the recent call to move from model-centric AI to data-centric AI,[7] more and more machine learning researchers are interested in the theoretical foundations and quantitative measures for data-centric AI. An interesting question here is whether VA research can provide empirical evidence for building this theoretical foundation and form the corresponding quantitative measures, and vice versa.

Another challenge is the smooth integration of machine learning into VA solutions for human-AI collaboration. VA solutions can facilitate sensemaking regarding complex data and scenarios, especially when fully automated models yield insufficient performances or tasks are ill-defined or *a priori* unknown. However, it is still challenging to characterize the role of human and AI in VA solutions and find the "optimal" balance and collaboration between them.

Providing Guidance in Complex VA Systems

VA supports the information-discovery process from a data-dependent, task-specific, and user-oriented perspective. However, for the users, who are usually experts in their application domains but not in VA, it is difficult to determine which VA methods to use for particular data and tasks. Guidance is needed to assist users with the selection of appropriate visual means and interaction techniques, the utilization of analytical methods, as well as the configuration instantiation of these methods with suitable

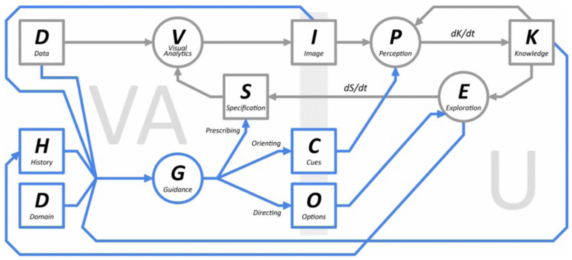

FIGURE 4. Model of guidance in VA.[8]

parameter settings and combinations thereof.[8] As shown in Figure 4, guidance hooks up from the users' knowledge and might consider different inputs, such as data, visualization images, interaction history, and domain knowledge. Guidance can be generated in various forms, such as visual cues and alternative options, which help explore the data, identify interesting data nuggets and findings (e.g., strange behaviors), and collect and group insights to explore high-level hypotheses and gain new knowledge. Guidance, which has its roots in human–computer interaction (HCI), can be envisioned as a mixed-initiative process where both the system and user contribute to the analysis. To this end, it is challenging to provide effective guidance, which is timely, trustworthy, adaptive, controllable, and nondisruptive.

Refining Evaluation Frameworks

We are not only dealing with complex systems to solve complex tasks, but the systems also have varying degrees of automation versus interaction and very few application experts are available for evaluating them. A survey of evaluation methods of VA solutions[9] suggests that 44% of studies performed summative evaluation without any human subjects (e.g., via application scenarios). VA solutions ask for new evaluation strategies because the classic quantitative and qualitative or HCI approaches are not applicable; we need to evaluate the visual information discovery process, which is heavily 1) user, data, and task-specific as well as 2) process-oriented. Research challenges lie in considering which high-level cognitive functions need to be tested for demonstrating the efficacy of changes within a VA framework and which more powerful AI approaches might let us create novel means of evaluation (e.g., automatic unit testing).

Some researchers, especially those from the machine learning field, may question why users prefer a complex visual interface enhanced by an active learning-based approach instead of combining a much simpler user interface with an automatic approach. For example, they may ask the following evaluation-related questions: What are the additional benefits of the carefully designed interactive visualization? Or, would users be willing to spend a lot of time interacting with the complex interface? To tackle these evaluation-related issues, further investigation is necessary.

Developing Generic Methods

Another challenge is to balance generalization and specificity. In the past decades, many domain/task-specific VA systems have been developed. As more and more solutions are needed for different domains/tasks, the required development efforts increase dramatically. With effort concerns in mind, the question of whether these solutions generalize between different analysis tasks and application domains arises and whether similar performance can be maintained when extending a VA system to another application domain/task. For example, as AI for image segmentation becomes increasingly available in the medical domain, practitioners would like to know whether these solutions can generalize between different hospitals and geographies. Researchers and practitioners are, therefore, interested in knowing how to evaluate/judge the benefits and impacts of domain-irrelevant/specific VA solutions. It would be interesting to explore the criteria and guidelines for building and evaluating these two types of solutions.

Constructing Explainable VA

As explainability becomes increasingly important, another larger goal is to build explainable VA—how can we capture one's analytical process and explain it to someone else? The analytical process often relies on implicit and explicit domain knowledge, but it remains challenging to capture and describe such knowledge for communication purposes. The visualization community has discussed this perspective with the idea of provenance.[10] Can we use provenance data to model analytical processes and train ML models for guidance?

Application

Now that we have discussed challenges in academia, the remaining challenges concern VA applications in practice.

Encouraging Deployment

Many application-specific VA systems are developed in the laboratory and not deployed in the field. More research is needed to understand the challenges of field deployment, especially in industry settings, to increase the impact of these systems. As deployment requires enormous resources and years to carry out, it is challenging to build a mechanism that encourages deployment. Specifically, those deployed solutions are unlikely to be published under current reviewing standards since they might lack novel visualizations and current evaluation methods, such as case studies favor more complex solutions.

However, deployment provides opportunities for more comprehensive evaluation methods, such as longitudinal studies. These studies allow researchers to collect more user feedback, perform symptom-cause analysis, and facilitate reflection, which can in turn advance the theorization of our field. Structural changes may also be possible, such as introducing a separate track (e.g., a demo track) in our conference program that favors VA systems that have been deployed in practice.

Promoting Open-Source Contributions

While we are often good at collaborating freely in an open way, we could need more push on being open in publications and software. Open software is key to facilitating comparison and improvement, as it allows us to conduct longitudinal studies if our complex systems are being adopted (e.g., Figure 5). Openly released algorithms are fundamental for reproducibility, while also promoting outreach into other application domains. Releasing a research demo or code comes with high costs if we do not start valuing coding as equally important as writing. How can we motivate Ph.D. students to work on demos? How do we achieve *sustainable open source*?

VA solutions are often released as encapsulated systems that are intrinsically challenging to reuse. If we want to achieve broader adoption (which is not always needed), we need more iterations on design

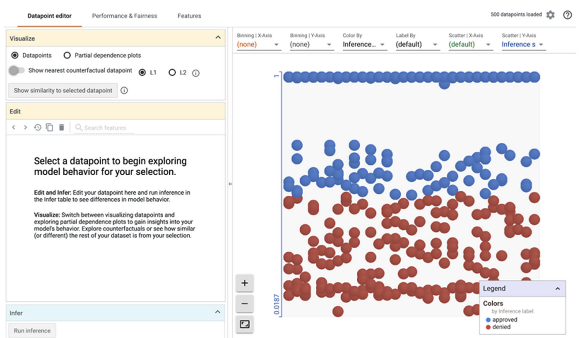

FIGURE 5. The What-if Tool[11] is a successful open-source tool that allows users to visualize and analyze ML model performances. It is packed as an extension in python notebooks and thus fits well into the ML development workflow.

and principled modularization. A metaphor is the German term "eierlegende Wollmilchsau"—the mystical egg-laying wool-milk-pig. This animal has the charm of offering a lot of useful features, but it most likely represents an impossible genetic combination of functions that would not be survivable or practical. Are there principled methods to decompose overly complex VA systems into basic building blocks? How can we modularize complex VA systems into reusable and interchangeable modules?

Community

The last set of challenges focuses on building a vibrant and wider VA community in the wild.

Expanding the Domain Space and User Groups

VA can (continue to) be a critical part of fundamental research in the future if we are careful about how we as VA researchers engage with others. As many VA systems are application oriented, we need continued efforts to scale up our impact through deeper collaboration with other domains and broader contact with end users. For example, VIS 2022 discloses that our current major "clients" are the ML and medical fields (as indicated by sessions, such as "VA and ML," "Interpreting Machine Learning," "Neuro/Brain/Medical Data," "DNA/Genome and Molecular Data/Vis," and "VA of Health Data"). Both fields have visualization programs in their academic conferences (e.g., the ICML Workshop on Visualization for Deep Learning, the workshop on VA in healthcare, and the BioVis event at the International Society for Computational Biology conferences). Similar efforts are needed to enhance interdisciplinary collaboration with other fields and foster the growth of new fields.

Defining Documentation Standards

Communicating application-specific VA systems is often done insufficiently, leaving both readers and users confused about how to use the systems. These challenges lie in all stages of VA application research, from requirement analysis and visualization design to evaluation and discovery. Voluntary demonstration videos are adopted as a remedy in the peer review process. However, VA practitioners still have difficulties explaining VA systems to end-users, visualization and nonvisualization researchers, and programmers. Thus, there is a need to define standards and develop guidelines for documenting VA systems in various forms including academic papers, demonstration videos, interactive demos, software programs, and read-me documents.

Creating Tutorial and Education Material

Developing VA applications requires a rich array of skills such as understanding user requirements, surveying users, programming software, and performing evaluations. Those skills span a range of fields, such as programming, algorithms, visualizations, and HCI. However, there currently exists little educational material or tutorials in training students or helping amateurs conduct VA application research. For example, many undergraduate courses on visualization stop at creating dashboards using software, such as PowerBI and Tableau.

CONCLUSION

To make VA application research more rigorous and valuable, we outline an agenda with 12 open challenges in this viewpoint piece. We situate these challenges within the VA application ecosystem that interconnects research with academia and practice, reflecting the distinct dilemmas faced by VA application research. While most of these challenges are immense, they also present great opportunities, especially given that VA application research is a vibrant yet relatively young research field. Sophisticated real-world issues will continue providing niches for VA applications and create new research and application opportunities, which in turn opens up avenues for promoting research rigor and value. We hope that this article offers a springboard for airing open challenges, sparking discussions, and inspiring continued efforts toward a fruitful future of VA applications in both academia and practice.

REFERENCES

1. D. Keim, G. Andrienko, J. D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon, "Visual analytics: Definition, process, and challenges," *Inf. Vis.*, pp. 154–175, 2008, doi: 10.1007/978-3-540-70956-5_7.
2. A. Wu, D. Deng, F. Cheng, Y. Wu, S. Liu, and H. Qu, "In defence of visual analytics systems: Replies to critics," *IEEE Trans. Vis. Comput. Graphics*, vol. 29, no. 1, pp. 1026–1036, Jan. 2023, doi: 10.1109/TVCG.2022.3209360.
3. Princeton University Library - Research Lifecycle Guide. [Online]. Available: https://researchdata.princeton.edu/research-lifecycle-guide/research-lifecycle-guide
4. M. Sedlmair, M. Miriah, and M. Munzner, "Design study methodology: Reflections from the trenches and the stacks," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 2, pp. 2431–2440, Dec. 2012, doi: 10.1109/TVCG.2012.213.
5. M. Chen and D. S. Ebert, "An ontological framework for supporting the design and evaluation of visual analytics systems," *Comput. Graphics Forum*, vol. 38, no. 3, pp. 131–144, 2019, doi: 10.1111/cgf.13677.
6. J. J. Thomas and K. A. Cook, "A visual analytics agenda," *IEEE Comput. Graphics Appl.*, vol. 26, no. 1, pp. 10–13, Jan./Feb. 2006, doi: 10.1109/MCG.2006.5.
7. N. Sambasivan, S. Kapania, H. Highfill, D. Akrong, P. Paritosh, and L. M. Aroyo, "Everyone wants to do the model work, not the data work: Data cascades in high-stakes AI," *Proc. ACM CHI*, no. 39, pp. 1–15, 2021, doi: 10.1145/3411764.3445518.
8. D. Ceneda et al., "Characterizing guidance in visual analytics," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 1, pp. 111–120, Jan. 2017, doi: 10.1109/TVCG.2016.2598468.
9. M. Khayat, M. Karimzadeh, D. S. Ebert, and A. Ghafoor, "The validity generalizability and feasibility of summative evaluation methods in visual analytics," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 1, pp. 353–363, Jan. 2020, doi: 10.1109/TVCG.2019.2934264.
10. K. Xu, A. Ottley, C. Walchshofer, M. Streit, R. Chang, and J. Wenskovitch, "Survey on the analysis of user interactions and visualization provenance," *Comput. Graphics Forum*, vol. 39, no. 3, pp. 757–783, 2020, doi: 10.1111/cgf.14035.
11. J. Wexler, M. Pushkarna, T. Bolukbasi, M. Wattenberg, F. Viégas, and J. Wilson, "The what-if tool: Interactive probing of machine learning models," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 1, pp. 56–65, Jan. 2020, doi: 10.1109/TVCG.2019.2934619.

AOYU WU is a postdoctoral fellow with Harvard University, Cambridge, MA, 02138, USA. He is the corresponding author of this article. Contact him at aoyuwu@g.harvard.edu.

DAZHEN DENG is a Ph.D. student in the State Key Lab of CAD&CG, Zhejiang University, Zhejiang, 310027, China. Contact him at dengdazhen@zju.edu.cn.

MIN CHEN is a professor of scientific visualization at Oxford University, OX1 3QG, Oxford, U.K., and a fellow of Pembroke College, Oxford, U.K. Contact him at min.chen@oerc.ox.ac.uk.

SHIXIA LIU is a professor in the School of Software, Tsinghua University, Beijing, 100190, China. Contact her at shixia@tsinghua.edu.cn.

DANIEL KEIM is a full professor and head of the Information Visualization and Data Analysis Research Group in the Computer and Information Science Department, University of

Konstanz, Konstanz, 78464, Germany. Contact him at keim@uni-konstanz.de.

ROSS MACIEJEWSKI is a professor in the School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ, 85281, USA, and the director of the Center for Accelerating Operational Efficiency. Contact him at rmacieje@asu.edu.

SILVIA MIKSCH is a university professor and head of the research division "Visual Analytics" (CVAST), Institute of Visual Computing and Human-Centered Technology, TU Wien, A-1040, Vienna, Austria. Contact her at silvia.miksch@tuwien.ac.at.

HENDRIK STROBELT is the explainability lead at the MIT-IBM Watson AI Lab, Cambridge, MA, 02142, USA, and research scientist at IBM Research, Albany, NY, USA. Contact him at hendrik@strobelt.com.

FERNANDA VIÉGAS is a Gordon McKay professor of computer science at the Harvard School of Engineering and Applied Sciences, Boston, MA, 02138, USA, a Chandra Family fellow at Harvard Business School, Boston, MA, USA, and a principal scientist at Google, Inc., Mountain View, CA, USA. Contact her at fernanda@g.harvard.edu.

MARTIN WATTENBERG is a Gordon McKay professor of computer science at the Harvard School of Engineering and Applied Sciences, Boston, MA, 02138, USA, a Chandra Family fellow at Harvard Business School, Boston, MA, USA, and a principal scientist at Google Inc., Mountain View, CA, USA. Contact him at wattenberg@g.harvard.edu.

Contact department editor Theresa-Marie Rhyne at theresamarierhyne@gmail.com.