Letter from the Special Issue Editors

As AI technologies are increasingly being used to make decisions that impact billions of people around the world, it is important that we take a proactive approach to ensure that these technologies are used responsibly and with the protections necessary to ensure that they are safe, trustworthy and consistent with our deepest ethical commitments.

The current AI systems are typically developed by people who have deep technical knowledge in computer science, mathematics, and optimization. They however may lack the expertise in how AI technologies are deployed and used in various social contexts as well as their potential societal impacts. In contrast, the Human-Computer Interaction (HCI) community has deep knowledge in how humans interact with complex systems and is well positioned to aid the development of responsible AI systems to ensure that they are beneficial to society and they are designed to be transparent, reliable, trustworthy and safe.

In this special issue on **Responsible AI and Human-AI Interaction**, we sought high-quality contributions on human-centered approaches to responsible and trustworthy AI. Leading HCI and AI researchers from both academia and industry worked together to address some pressing issues in developing responsible and trustworthy AI systems such as AI ethics, bias/fairness, explainability, and transparency.

Nora McDonald from University of Cincinnati, and *Aaron Massey* and *Foad Hamidi* from University of Maryland, Baltimore County reflect on why and how Artificial Intelligence (AI)-enhanced Adaptive Assistive Technologies (AATs) need to be designed in collaboration with AAT users belonging to intersecting marginalized groups to ensure that the benefits of AI do not sacrifice privacy for the most vulnerable (e.g., older adults with disabilities).

Alex Okeson from University of Washington and her co-authors from Microsoft Research explore human-centered approaches to Machine Learning (ML) interpretability. They focus on one aspect of interpretability tools, global feature attributions, which are frequently used by ML developers to understand ML model behavior. They conducted an artifact-based interview study intended to investigate whether ML developers would benefit from being able to compare and contrast different global feature attribution methods.

Patrick Gage Kelley and his co-authors from Google and Ipsos present the results of an in-depth survey of public opinion of Artificial Intelligence (AI) conducted with over 17,000 respondents spanning fifteen countries and six continents. Analysis of responses has revealed four emergent themes of sentiment towards AI: exciting, useful, worrying, and futuristic. These sentiments and their relative prevalence may inform how the public influences the development of AI.

John Richards and his colleagues from IBM research explore human-centered methods to address the need for increased transparency in artificial intelligence (AI) for data sets, models, and services. They present a methodology for creating FactSheets, a form of transparent AI documentation. They also share the insights gathered while they creating nearly two dozen FactSheets.

Finally, *Philip Feldman and Aaron Dant* from ASRC Federal and *David Rosenbluth* from Lockheed-Martin present a mechanism to harness the narrative output of large language models and produce "Neural Narrative Maps" (NNMs) that are intended to provide insight into intent and belief and how they evolve in an information space. They demonstrate the utility of their methods in understanding rules of engagement (e.g., if a subordinate is following a commander's intent in a high-risk situation).

Shimei Pan and James Foulds
University of Maryland, Baltimore County