Interactive AI Model Debugging and Correction

Research in Artificial Intelligence (AI) has advanced at an incredible pace, to the point where it is making its way into our everyday lives, explicitly and behind the scenes. However, beneath their impressive progress, many AI models hide deficiencies that amplify social biases or even cause fatal accidents. How do we identify, improve, and cope with imperfect models, while still benefiting from their use? I will discuss my work empowering humans to interact with AI models in order to debug and correct them. I will describe both (1) how I help experts run scalable and testable analyses on models in development, and (2) how I help end users collaborate with deployed AI in a transparent and controllable way. In my final remarks, I will discuss my future research perspectives on building human-centered AI through data-centric approaches.

Bio: Sherry Tongshuang Wu is a final year Ph.D. Candidate in Computer Science & Engineering at the University of Washington, advised by Jeffrey Heer and Dan Weld. She received her B.Eng in CSE from the Hong Kong University of Science and Technology. Her research lies at the intersection of Human-Computer Interaction (HCI) and Natural Language Processing (NLP), and aims to empower humans to debug and correct AI models interactively, both when the model is under active development, and after it is deployed for end users. Sherry has authored 19 papers in top-tier NLP, HCI and Visualization conferences and journals such as ACL, CHI, TOCHI, TVCG, etc, including a best paper award (top-1) and an honorable mention (top-3).

Host: Chenhao Tan