Carnegie Mellon University
Master of Science in Information Technology

privacy ENGINEERING

The Block Center FOR TECHNOLOGY AND SOCIETY

Privacy Seminar

Co-Sponsored with the Block Center for Technology and Society

Speaker: Krishnaram Kenthapadi

Title: Deploying Trustworthy Generative AI

Abstract:

Generative AI models and applications are being rapidly deployed across several industries, but there are several ethical and social considerations that need to be addressed. These concerns include lack of interpretability, bias and discrimination, privacy, lack of model robustness, fake and misleading content, copyright implications, plagiarism, and environmental impact associated with training and inference of generative AI models. In this talk, we first motivate the need for adopting responsible AI principles when developing and deploying large language models (LLMs) and other generative AI models, and provide a roadmap for thinking about responsible AI for generative AI in practice. Focusing on real-world LLM use cases (e.g., evaluating LLMs for robustness, security, bias, etc.), we present practical solution approaches / guidelines for applying responsible AI techniques effectively and discuss lessons learned from deploying responsible AI approaches for generative AI applications in practice. By providing real-world generative AI use cases, lessons learned, best practices, and open research problems, this talk will enable researchers and practitioners to build more reliable and trustworthy generative AI applications. Please take a look at our recent ICML/KDD/FAccT tutorial (https://sites.google.com/view/responsible-gen-ai-tutorial) for an expanded version of this talk.

Bio:

Krishnaram Kenthapadi is the Chief AI Officer & Chief Scientist of Fiddler AI, an enterprise startup building a responsible AI and ML monitoring platform. Previously, he was a Principal Scientist at Amazon AWS AI, where he led the fairness, explainability, privacy, and model understanding initiatives in the Amazon AI platform. Prior to joining Amazon, he led similar efforts at the LinkedIn AI team, and served as LinkedIn's representative in Microsoft's AI and Ethics in Engineering and Research (AETHER) Advisory Board. Previously, he was a Researcher at Microsoft Research Silicon Valley Lab. Krishnaram received his Ph.D. in Computer Science from Stanford University in 2006.

WHEN: October 10th 2023
12:30-1:50pm

WHERE: Hamburg Hall Room 1002

ZOOM LINK
https://cmu.zoom.us/s/97389172852

Password: 429573