Date and Time: Feb 4, 2026, 3:30.PM.

Venue: RM-301(Rajeev Motwani Building, CSE Department).

Title: Understanding Bias in Generative Vision Models: Measurement, Causality, and Mitigation

Abstract:

This talk focuses on understanding, diagnosing, and mitigating biases in text-to-image (TTI) generative models. I will first present a series of methods developed to systematically analyze social and representational biases in these models, including TIBET for dynamic bias evaluation, BiasConnect for studying interactions between multiple bias axes through causal structures, and InterMit for modular, intervention-based bias mitigation. Together, these works aim to move beyond surface-level bias measurements toward more principled and causal analyses of generative behavior. I will conclude by introducing Namya Kavach, an ongoing effort to build scalable, transparent, and community-oriented frameworks for evaluating and governing risks in generative AI systems.

Speaker Bio:

Pushkar Shukla is a researcher in Responsible and Trustworthy AI, currently a Postdoctoral Researcher at the Wharton School of the University of Pennsylvania. His work focuses on evaluating and mitigating bias in generative AI systems, with a particular emphasis on text-to-image and multimodal models. He develops methods for dynamic bias diagnosis, intersectional analysis using causal structures, and modular intervention strategies to improve fairness and safety in AI systems.

Pushkar received his PhD from Toyota Technological Institute at Chicago, where he worked at the intersection of computer vision, fairness, and interpretability under the guidance of Matthew Turk. At Wharton, he collaborates with Kartik Hosanagar on research at the intersection of AI safety, governance, and real-world AI deployment.

His broader research interests include causal reasoning for bias analysis, risk evaluation in large language and multimodal models, and building transparent, community-centered frameworks for governing generative AI.