USC Viterbi School of Engineering
Ming Hsieh Department of Electrical and Computer Engineering

Center on AI Foundations for the Sciences

AIF4S Seminar Series

AI Safety Issues in Generative Models: Memorization and Detection

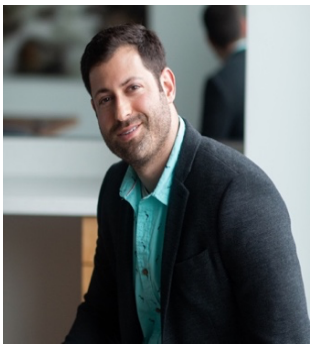
Tom Goldstein
Volpi-Cupal Professor of Computer Science at University of Maryland and Director of the Maryland Center for Machine Learning

Friday November 22, 2024| 10:30am-11:30am PST| EEB 132

Zoom Link:
https://usc.zoom.us/j/96326189770?pwd=6XmvQynLQPm4oADvK23qQy6D734L3K.1

Abstract: Machine learning systems are built using large troves of training data that may contain private or copyrighted content. In this talk, I'll survey a number of data memorization issues that arise when sensitive data is used. I'll begin by talking about data privacy issues that arise when using generative models. These models are often created using a training objective that explicitly promotes their ability to regenerate their training data. I'll discuss how diffusion models can reproduce their training data, leading to potential legal issues. I'll also discuss methods for detecting large language model content and explore ways in which the ability to reproduce training data complicates our ability to detect LLM-produced text.

Biography: Tom Goldstein is the Volpi-Cupal Professor of Computer Science at the University of Maryland, and director of the Maryland Center for Machine Learning. His research lies at the intersection of machine learning and optimization, and targets applications in computer vision and signal processing. Professor Goldstein has been the recipient of several awards, including SIAM's DiPrima Prize, a DARPA Young Faculty Award, a JP Morgan Faculty award, an Amazon Research Award, and a Sloan Fellowship.



Host: Dr. Mahdi Soltanolkotbi, soltanol@usc.edu