Trustworthy AI via Advancing Data Integrity and Human-Centered Design

by

Ms. Yunyi LI
University of Texas at Austin

DATE	12 January 2026 (Monday)
TIME	2:30 - 4:00 pm
VENUE	4/F Meeting Room (Room 4047), LSK Business Building

ABSTRACT

In an era where organizations increasingly rely on data-driven systems to inform decisions, shape operations, and interact with customers, the trustworthiness of these socio-technical systems depends not only on the design of the analytical paradigm, but also, critically, on the faithfulness of the data used to build them. In this talk, I will begin by sharing my research that bridges data quality, human–AI interaction, algorithmic fairness, and interpretable machine learning to advance the trustworthiness of socio-technical, data-driven systems, enabling them to serve organizations and society in responsible and lasting ways.

Next, I will delve into research that addresses a notable data integrity challenge: label bias, which refers to systematic errors in a label—a covariate central to a quantitative analysis—such that its quality differs across social groups. This type of bias has been conceptually and empirically explored and is widely recognized as a pressing issue across critical domains. However, effective methodologies for addressing it remain scarce. In this work, we propose Decoupled Confident Learning (DeCoLe), a principled machine learning-based framework specifically designed to detect mislabeled instances in datasets affected by label bias, enabling bias-aware mislabelling detection and facilitating data quality improvement. We theoretically justify DeCoLe's effectiveness and evaluate its performance in the impactful context of hate speech detection, a domain where label bias is a well-documented challenge. Empirical results demonstrate that DeCoLe excels at bias-aware mislabeling detection, consistently outperforming alternative approaches for label error detection. Our work identifies and addresses the challenge of bias-aware mislabeling detection and offers guidance on how DeCoLe can be integrated into organizational data management practices as a powerful tool to enhance data reliability.

BIOGRAPHY

Yunyi Li is a Ph.D. candidate in Information Systems–Computational Data Science at the McCombs School of Business, University of Texas at Austin. Her research advances trustworthy AI to maximize the societal benefits of socio-technical systems while mitigating potential harms. Motivated by the risks of overlooking label bias, her research uncovers its harmful consequences and develops novel methods to mitigate it. More broadly, her work bridges data integrity, algorithmic fairness, human–AI interaction, and interpretable machine learning, with the overarching goal of enhancing the reliability of socio-technical systems to better serve both organizations and society. Her work has been published in premier journals, including INFORMS Journal on Computing, Notices of the American Mathematical Society, Stat, HCOMP, and Genome Research. Her job market paper is under revision at Information Systems Research. She has presented at leading conferences and given invited talks, including CIST, INFORMS, SCECR, ICML, NeurIPS, and the Big XII MIS Research Symposium, and has received the McCombs Dean's Fellowship, Dissertation Excellence Fellowship, the UT Austin Machine Learning Laboratory Research Grant, the Good Systems Award for Ethical AI Research, and an NSF Big Data Grant.