UNIVERSITY of GUELPH

COLLEGE of ENGINEERING AND PHYSICAL SCIENCES

SCHOOL OF COMPUTER SCIENCE

PhD Qualifying Exam

Monday August 21, 2023 at 1pm via Zoom (remote)

Sakib Shahriar

Towards Privacy-Preserving Analytics in Sensitive Social Media Texts

Chair: Dr. Stacey Scott
Advisor: Dr. Rozita Dara
Advisory: Dr. Fei Song
Non-Advisory 1: Dr. Fattane Zarrinkalam [SoE]
Non-Advisory 2: Dr. Charlie Obimbo

Abstract:

Governments and organizations are increasingly harnessing social media data for analytics, leveraging insights to guide decision-making. While this collection of user-generated data provides rich context, the analysis of it raises privacy concerns. Privacy regulations mandate stringent safeguards for personal data, yet identifying and protecting such data within vast, unstructured social media datasets remains a challenge. With emerging technologies like large language models, these challenges are compounded, raising the risk of unintentional sensitive information disclosure. Contemporary research leveraging social media data for sensitive applications often fails to adequately address privacy considerations. In this research, we address the critical privacy challenges associated with sensitive social media texts.

To tackle these issues, we propose a six-component privacy framework targeting direct identifiers, social connections, sensitive attributes, data inaccuracies, authorship attribution, and large language models. Existing solutions, such as differential privacy and anonymization, have limitations, notably ineffective resistance to adversaries and compromised data utility. Our key contributions include a comprehensive privacy-preserving framework identifying and categorizing privacy risks for social media, a novel text transformation mechanism to mitigate the risks of authorship attribution and sensitive information disclosure, an empirical investigation into large language model's implications on text privacy, and a differentially-private deep learning clustering algorithm for high-dimensional social media texts.

Our preliminary results for the innovative transformation mechanism ‘privacy by translation’ are promising in mitigating the risk of authorship attribution in social media texts, demonstrating a significant increase in privacy (24.1%) against a marginal decrease in utility (0.99%). The results highlight the potential of our proposed framework to enhance privacy protection while maintaining utility in the analysis of sensitive social media texts.