Dissertation Proposal Defense
Doctor of Philosophy in Information Science

“Towards Understanding the Challenges and Needs of Practitioners in Developing Privacy-Preserving Machine Learning Systems” by Maryam Aldairi

Date: Aug 3, 2023
Time: 1:00PM – 3:00PM
Place: Room 828, Information Science Building, 135 North Bellefield Ave, Pittsburgh PA 15260

Committee:
- James Joshi, Professor, Department of Informatics and Networked Systems, School of Computing and Information
- Balaji Palanisamy, Associate Professor, Department of Informatics and Networked Systems, School of Computing and Information
- Paul Munro, Associate Professor, Department of Informatics and Networked Systems, School of Computing and Information
- Hanan Hibshi, Assistant Teaching Professor, The Information Networking Institute, Carnegie Mellon University

Abstract:
Rapid advances in Artificial Intelligence (AI) and Machine Learning (ML) in recent years have shown significant benefits for their use in many applications. Simultaneously, privacy has captured the immense attention of the public, industry, and academic community, especially with regards to AI/ML. Growing concerns in privacy have recently led to new privacy laws, extensive research to identify privacy threats, and the introduction of domain-specific privacy requirements and solutions. Incorporating ML models in systems that lack the necessary foundations to protect privacy can exacerbate the privacy risks for ML enabled systems. Deploying ML models in systems with little regard for privacy can jeopardize the privacy of the systems’ data. Emerging large-scale systems like National Digital ID (NDID) have privacy requirements specific to their needs, while providing opportunities to improve the quality of life of citizens and communities by offering the means to verify identities legally and provide services in online environments. Governments are adopting such NDID systems and exploring the role of ML models in such systems to provide support for different tasks, including proactively applying for citizen's benefits and services, employing digital assistants, creating personalized recommendations, and predicting the societal risks related to housing and food security. These ML models are trained using privacy sensitive and unbalanced data from the NDID ecosystem. The models can be biased due to the algorithm or the dataset used to train them. It is essential to be careful when developing privacy-preserving ML systems to avoid the possibility of bias still being present or even amplified. Furthermore, the absence of privacy protection measures in the hosting system and the existing privacy concerns in ML models can lead to additional privacy and fairness issues.

This dissertation examines the challenges and needs of building privacy-preserving ML systems: First, I study privacy threats in NDID ecosystem by analyzing real-world case studies and propose a privacy requirements elicitation framework for NDID systems. Next, I study how
privacy (specifically differential privacy) and fairness mechanisms interact when used in ML models that can be incorporated into systems such as NDID systems. Through an empirical evaluation of these mechanisms in three stages of learning - pre-, in-, and post-processing - I aim to determine the optimal settings for better balance between privacy and fairness, and understand their trade-offs. This dissertation will provide valuable insights into optimizing the performance of these models when considering privacy and fairness issues. Finally, I will investigate ML practitioners' perspectives on privacy in ML. I propose to 1) analyze publicly available data from Community Question Answering websites (CQA), and 2) conduct a user study to understand developers' practices, challenges, and needs when developing private ML models.