Project Title:

Human-Centered Data Science – Trustworthy AI in Healthcare and Medical Research

(Tabular health data concentration)

Supervisor: Prof. Mark Chignell

Application: Please submit CV, unofficial transcript, and a single paragraph describing your interest in the project in one file in a single email to Prof. Mark Chignell at chignel@mie.utoronto.ca and Dr. Lu Wang at lwang71@central.uh.edu. We will decide the acceptance based on your interest and completion in doing a small real data work within our projects instead of your interview performance.

Project Background: Healthcare systems are changing in the era of big data. Advances of artificial intelligence (AI) in healthcare make it possible for healthcare providers to sift through tremendous amounts of information efficiently, which eventually help them take care of their patients better. There are various types of health information ranging from medical literature to pathology reports. Albeit the last few years have witnessed an explosive increase of healthcare data in terms of volume, variety and veracity, it is insufficient to build a robust prediction model in various scenarios due to time, geographical and domain inherent constraints. How to develop and apply Machine Learning (ML) methods that can efficiently utilize Electronic Health/Medical Records (EHRs/EMRs) is significant to facilitate decision making of physicians in their clinical practice.

In addition to ML methods, inspired by the motivation of human-in-the-loop, Human-Centered AI (HCAI) for Data Driven Decision Making (D3M) addressing healthcare/medicine problems attracts more attention to improve the situation awareness and the quality of decisions. More specifically, interactive Machine Learning (iML) improves the ML *prediction* by looping human experts in the learning process and integrating human expertise. From another perspective, eXplainable Artificial Intelligence (XAI) and trustworthy AI in healthcare systems not only improve the uptake of ML model but also increase physician *trust* in ML prediction for clinical decision making.

In this project, you will have opportunities working with the real healthcare and medical data including EHRs/EMRs for multiple cognitive disorders and chronic diseases collaborating with physicians, clinicians and psychiatrists, etc.

Project Description:

1. GEMINI hospital data for delirium detection: Data collected from the GEMINI study supports the Ontario General Medicine Quality Improvement Network (GeMQIN), a program of Ontario Health (Quality) and housed at Unity Health Toronto's St. Michael's Hospital, GeMQIN Ontario Health, Sunnybrook Health Sciences Centre, Sinai Health System, Trillium Health Partners, and the University Health Network. This network brings general medicine physicians and care teams together to learn new ways to improve patient care and reduce variation in care. It also helps to address provincial priorities such as reducing hallway healthcare. (Preferred: Professional programming skills in data preprocessing, preparation and modeling, e.g., Python, R)

2. Centivizer game data for delirium detection with proposing novel machine learning algorithms: Centivizer creates interactive and rewarding technology-based activities

tailored to stimulate physical and cognitive functions of the elderly as well as unleash the potential of technology to promote healthy and active living for the elderly. (Preferred: Strong background in machine learning and professional programming skills, e.g., Python)

3. Depression drug treatment: 10 protocols consisting of phase II, III, and IV clinical trial data on duloxetine and comparators (venlafaxine, paroxetine, and placebo) provided by the Eli Lilly pharmaceutical company was retrospectively compiled and analyzed. The compiled dataset included HAM-D data from 10 randomized clinical trials covering a total of 2,292 participants. (Preferred: Good background in data science including statistical and data mining methods, and professional programming skills, e.g., R, Python)

Start date: Summer/Fall 2025
Research area: Data science, machine learning, artificial intelligence, healthcare, medical research.