
**Department of Computer Science**
**St. Francis Xavier University**

**M.Sc. Thesis Proposal Presentation**

# Debiasing of Large Language Models for Rural Users

**Presented by**
Patrick Bowen

**Date:** Friday, March 13, 2026
**Time:** 9:30 AM
**Location:** Annex 113

**Abstract:** Large language models, or LLMs, are powerful artificial intelligence systems that generate human-like text. These models are typically trained on internet-derived data, which often reflects urban perspectives disproportionately. This may lead to biased or inaccurate responses when applied to rural contexts, where linguistic patterns, cultural norms, and socioeconomic experiences may differ considerably. Using geotagged Twitter data from the United States in March 2010, this research aims to evaluate the effectiveness of various combinations of debiasing techniques including Counterfactual Data Augmentation (CDA), dropout, Self-Debias, SentenceDebias, and Iterative Nullspace Projection (INLP), that have seen some success with other social biases. The bias mitigation, language modelling, and natural language understanding abilities of the Llama-3, Qwen 2.5, and Gemma 3 LLMs will be evaluated before and after undergoing each of these debiasing techniques. Bias mitigation will be evaluated using the Word Embedding Association Test, Sentence Embedding Association Test, and versions of the StereoSet & CrowS-Pairs datasets modified for geographical bias; language modelling will be evaluated using a heldout test set of tweets; and natural language understanding abilities will be evaluated using the SuperGLUE benchmark.