LLMs as Social Cognitive Engines: Simulation, Prediction, and Augmentation Across Sociotechnical Information Systems

Name: Zachary Samson Pangan

Laboratory's name: Social Computing Laboratory

Supervisor's name: Prof. Eiji Aramaki, PhD

Abstract

Large Language Models (LLMs) are increasingly embedded in sociotechnical systems, yet their potential as methodological instruments for computational social science remains underexplored. As major social media platforms continue to restrict researcher access to behavioral data, the need for alternative, replicable, and semantically rich modeling approaches has become more urgent. This dissertation investigates how LLMs can simulate, predict, and augment socio-informational processes across the layers of the modern digital ecosystem: online discourse, media-shaped opinion formation, and algorithmic recommendation.

Study 1 develops an LLM-based multi-agent simulation that generates naturalistic, evolving conversations, enabling the analysis of semantic diffusion and sentiment dynamics without relying on restricted social media APIs. Study 2 introduces a media-diet modeling approach in which LLMs fine-tuned on ideologically distinct news corpora produce stance distributions reflective of media-conditioned biases, offering a scalable method for estimating public opinion. Study 3 proposes a dual semantic–topological augmentation framework for graph-based recommender systems, using LLMs to enrich item semantics and infer relational structure, thereby improving both accuracy and diversity.

Together, these studies position LLMs as social cognitive engines capable of generating discourse, encoding media influence, and reshaping algorithmic behavior. In a research landscape where direct access to platform data is increasingly constrained, this dissertation demonstrates how LLM-driven methods can help sustain empirical inquiry into the dynamics of digital publics and sociotechnical information flows.