Research Statement

LIM Ee Peng
School of Computing and Information Systems, Singapore Management University
Tel: (65) 6828-0781; Email: eplim@smu.edu.sg
1 (Day) 12 (Month) 2025 (Year)

Background

My research interests cover data mining, machine learning, and information retrieval. In my research, I develop computational methods that model and analyse user generated structured and unstructured data for the purpose of extracting knowledge embedded in the data and improving social and urban wellbeing through recommendation and prediction. In the following, I summarize a few selected works completed in 2025.

Research Areas

1. Conversational Mental Health Agent Research

I am working on **conversational mental health agent research** as part of a new funded project by NMRC. Below are two research topics studied in 2025.

In the first research topic, we aim to develop conversational counselor agents for addressing the rising demand for scalable and accessible mental health support. This work introduces CAMI, a novel automated counselor agent grounded in Motivational Interviewing (MI)–a client-centered counseling approach designed to address ambivalence and facilitate behavior change. CAMI employs a novel STAR framework, consisting of client's state inference, motivation topic exploration, and response generation modules, leveraging large language models (LLMs). These components work together to evoke change talk, aligning with MI principles and improving counseling outcomes for diverse clients. We evaluate CAMI's performance through both automated and expert evaluations, utilizing simulated clients to assess MI skill competency, client's state inference accuracy, topic exploration proficiency, and overall counseling success. Results show that CAMI not only outperforms several state-of-the-art methods but also shows more realistic counselor-like behavior. Additionally, our ablation study underscores the critical roles of state inference and topic exploration in achieving this performance.

In the second research topic, we focus on simulating human clients in mental health counseling. This is crucial for training and evaluating counselors (both human and simulated) in a scalable manner. Nevertheless, past research on client simulation did not focus on complex conversation tasks such as mental health counseling. In these tasks, the challenge is to ensure that the client's actions (i.e., interactions with the counselor) are consistent with its stipulated profiles and negative behavior settings. In this paper, we propose a novel framework that supports consistent client simulation for mental health counseling. Our framework tracks the mental state of a simulated client, controls its state transitions, and generates for each state behaviors consistent with the client's motivation, beliefs, preferred plan to change, and receptivity. By varying the client profile and receptivity, we demonstrate that consistent simulated clients for different counseling scenarios can be effectively created. Both our

automatic and expert evaluations on the generated counseling sessions also show that our client simulation method achieves higher consistency than previous methods.

2. Career Trajectory Analytics Research

As part of my MOE Tier 2 project, we developed a Multi-Stage Framework with Taxonomy-Guided Reasoning for Occupation Classification Using Large Language Models. Automatically annotating job data with standardized occupations from taxonomies, known as occupation classification, is crucial for labor market analysis. However, this task is often hindered by data scarcity and the challenges of manual annotations. While large language models (LLMs) hold promises due to their extensive world knowledge and in-context learning capabilities, their effectiveness depends on their knowledge of occupational taxonomies, which remains unclear. In this study, we assess the ability of LLMs to generate precise taxonomic entities from taxonomy, highlighting their limitations, especially for smaller models. To address these challenges, we propose a multi-stage framework consisting of inference, retrieval, and reranking stages, which integrates taxonomy-guided reasoning examples to enhance performance by aligning outputs with taxonomic knowledge. Evaluations on a large-scale dataset show that our framework not only enhances occupation and skill classification tasks but also provides a cost-effective alternative to frontier models like GPT-4o, significantly reducing computational costs while maintaining strong performance. This makes it a practical and scalable solution for occupation classification and related tasks across LLMs.

Future Research. I shall continue using my research expertise in AI, data mining, and information retrieval to explore more varied research problems in novel application domains. I will also expand my collaboration with domain experts to identify and work on important social and urban research problems.

Selected Publications and Outputs

Yang, Yizhe, Palakorn Achananuparp, He-Yan Huang, Jing Jiang, Phey Ling Kit, Nicholas Gabriel Lim, Cameron Tan Shi Ern, and Ee-Peng Lim. "Cami: A counselor agent supporting motivational interviewing through state inference and topic exploration." In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 21037-21081. 2025.

Yang, Yizhe, Palakorn Achananuparp, He-Yan Huang, Jing Jiang, Nicholas Gabriel Lim, Cameron Tan Shi Ern, Phey Ling Kit, Jenny Giam Xiuhui, John Pinto, and Ee-peng Lim. "Consistent client simulation for motivational interviewing-based counseling." In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 20959-20998. 2025.

Palakorn Achananuparp, Ee-Peng Lim, Yao Lu. A multi-stage framework with taxonomy-guided reasoning for occupation classification using large language models, to appear at ICWSM 2026.