



Department of Biomedical Engineering
The Chinese University of Hong Kong



Graduate Seminar – PhD Oral Defence

Student : Mr. WEI Hao
Supervisor : Prof. YUAN Wu Scott
Date : 28 July 2026
Time : 9:30 am
Venue : Room 1122, William M W Mong Engineering Building, CUHK

Title: Collaborative Medical AI Agents via Multimodal Learning and Causal Reasoning

Artificial intelligence (AI) has demonstrated substantial potential in medical image analysis, clinical decision support, and medical education, yet its translation into realistic healthcare workflows remains constrained by several unresolved limitations. Medical AI models may show reduced generalizability when imaging devices, clinical populations, or acquisition protocols change; multimodal clinical evidence is often fragmented across images, text, and structured records; and vision-language systems may produce fluent explanations that are not sufficiently grounded in visual or clinical evidence. In addition, many medical AI systems are still designed and evaluated as isolated predictors, whereas real clinical practice is interactive, multi-role, and dependent on the progressive collection and interpretation of evidence. This thesis defense seminar presents a series of studies that investigate collaborative medical AI agents through a staged framework spanning robust perception, grounded cognition, and agentic collaboration. The first part focuses on multimodal perception in ophthalmology. CauDR studies causality-inspired domain generalization for diabetic retinopathy grading by examining how device-related appearance variations may induce spurious correlations. M3OCTA explores multimodal fusion for ultra-wide OCTA analysis, where multiple retinal and choroidal projections provide complementary diagnostic information. VisionFM further investigates large-scale multimodal and multitask pretraining for generalist ophthalmic image understanding across diverse imaging modalities and clinical tasks. The second part examines the transition from visual recognition to clinically grounded vision-language reasoning. VisionCLIP explores privacy-preserving language-image pretraining using synthetic retinal image-text pairs. OCT-Thinker and IRIS then study structured vision-language reasoning for retinal OCT interpretation and ocular surface disease understanding, respectively, with an emphasis on linking visual findings, anatomical structures, clinical concepts, and language-based explanations. The third part turns to agentic orchestration and causal multi-agent collaboration. Fuxi Cloud is presented as an orchestration platform that connects ophthalmic AI models with large language models through tool-calling workflows. MEDCO investigates a multi-agent framework for medical education by simulating interactions among patients, students, radiologists, and medical experts in case-based learning. AllergyMDT explores active inquiry and causal graph construction for complex allergic disease diagnosis, where missing information and multi-system manifestations are central to diagnostic reasoning. Taken together, these studies examine how robust perceptual models, grounded vision-language systems, and structured agent interactions may contribute to the development of more interpretable and collaborative medical AI systems. The seminar will summarize the main methodological designs, experimental findings, limitations, and future directions, while emphasizing that collaborative medical AI agents remain a research direction requiring careful validation rather than a replacement for clinical expertise.

***** ALL ARE WELCOME *****

For enquiries, please contact Ms. Joyce Chan, Department of Biomedical Engineering at 3943 8278