

Retrieval-Augmented Action Guidance for Multimodal Robot Control

Faculty advisor: Prof. Chi-Guhn Lee

Autonomous robots that genuinely understand human instructions need to weave together language, visual perception, and their own sensorimotor experience. Recent advances in multimodal transformers and retrieval-augmented generation (RAG) hint at a future where a robot can instantly tap a vast memory of past successes to inform its next move. Harnessing that capability promises faster learning, sharper generalisation, and more natural collaboration between people and machines.

Join a research team exploring how Retrieval-Augmented Generation (RAG) can boost the decision-making power of multimodal transformers that drive language-guided robots.

Preferred Foundation / Experience

- Solid grasp of reinforcement learning and/or imitation learning algorithms
- Hands-on experience with multimodal transformers or vision-language models
- Familiarity with retrieval pipelines (vector DBs, embedding indexes)
- Comfort in robotics simulation frameworks (e.g., IsaacSim, MuJoCo)
- Proficiency in Python + PyTorch; basic MLOps skills for experiment tracking
- Prior coursework or projects in Deep Learning, NLP, or Robotics is a strong plus

Why this project?

You'll work on a fast-growing research frontier where large models meet embodied intelligence, gain hands-on experience bridging retrieval systems with control policies, and leave with portfolio-ready results attractive to both industry R&D labs and PhD programs.

Contact: Prof. Lee, cglee@mie.utoronto.ca