Yi Lu

(+1) 437-605-1024 | tomlu@cs.toronto.edu | Linkedin | Github | Google Scholar | Home Page

Education

University of Toronto Sept.2024 – Dec.2025 (Expected)
Master of Science in Applied Computing (AI Concentration)
Current GPA: 4.0/4.0

University of Manchester Sept.2020 – Jun.2023
B. Sc. Hons Computer Science (Specialized in Artificial Intelligence)
GPA: 4.0/4.0 (First Class Honours, around Top % 5 among the cohort)

Publications

RSVP: Reasoning Segmentation via Visual Prompting and Multi-Modal Chain-of-Thought ACL 2025 Main

- First-authored and Project Lead, responsible for model implementation, experiment and paper writing.
- Proposed a novel **State-of-the-art** (SOTA) Training-free Reasoning Segmentation Model, unleashed Vision Language Model's localization capability by utilizing region-aware multi-modal visual prompt and hierarchical reasoning architecture.
- Implemented solution surpassed previous SOTA fine-tuned models by up to **9.2 cIoU metric** on ReasonSeg dataset.

VEU-Bench: Towards Comprehensive Understanding of Video Editing CVPR 2025 Highlight

- Accepted as CVPR 2025 Highlight, responsible for dataset curation, benchmarking and paper writing.
- Proposed a comprehensive fine-grained benchmark dataset for high-level video editing comprehension, demonstrated research gap by evaluating **11 SOTA video-language models** on high-level video editing comprehension tasks.
- **Fine-tuned Qwen2-VL 7B** as baseline model using **ModelScope Swift framework**. Resulting model achieved **GPT-4o-level performance** on video editing understanding tasks.

Video Repurposing from User Generated Content: A Large-scale Dataset and Benchmark AAAI 2025 Poster

- Paper accepted as **AAAI 2025 poster**, Contributed to the dataset cleaning, construction and paper writing.
- Proposed cross-modality video long-to-short (Video Repurposing) task, curated a large-scale dataset and a baseline model.

ZeroTrail: Training-Free Trajectory Control Framework for Video Diffusion Models In Submission

- Independent Research Project and single first author, Paper in submission to AAAI 2026.
- Proposed a novel training-free trajectory control framework for video diffusion models utilizing soft cross-attention guidance and test-time latent optimization. The model achieved superior performance across all 3D U-Net-based Video Diffusion Models on Trajectory Control Benchmarks.

VeRL-Tool: Towards Holistic Agentic Reinforcement Learning In Submission

- Co-first authored and project lead for Python Coding, NL2SQL, and Local-Retriever Agent implementation subtasks. Contributed to framework development, experiment, and paper writing. The paper is in submission to ICLR 2026.
- Proposed and implemented the first holistic agentic RL training framework supporting versatile stateful multi-turn tool-calling agent training across multiple modalities.

Research Internships

Research Intern Mar.2025 - Present
TIGER Lab @ University of Waterloo *Topic: Reinforcement Learning, LLM Agent*

- Under the supervision of Prof. Wenhu Chen, currently working on Reinforcement Learning and agentic LLMs.
- Core contributor to **open-source MLLM Reinforcement Learning Framework: VeRL-Tool**, responsible for the development and integration of Python Code Interpreter, Agentic Search and NL2SQL tool, handled Tool-aware LLM Coder, Web Search Agent, and SQL-based Tabular understanding agent's training and benchmarking.

Applied Research Intern May.2025 - Present
ModiFace (L'Oréal's AI Lab) *Topic: Controllable Video Generation.*

- Developing a Portrait-oriented Human Animate video generation model based on Wan-2.2-TI2V and EDTalk. The model is capable of generating motion-guided, emotion-controllable, high-definition animated videos. Work is currently in progress and aiming for ECCV 2026.

Applied Research Intern Jan.2025 - May.2025
Opus AI Research *Topic: Multi-Modal Large Language Models*

- Responsible for developing and benchmarking a training-free reasoning segmentation model. In charge of project leading, model implementation, experiments, and paper writing.
- Responsible for constructing a benchmark dataset for high-level video editing comprehension. Repurposed and categorized mainstream VQA datasets for video editing comprehension tasks. Benchmarked 11 SOTA MLLMs on the curated dataset and contributed to paper writing, fine-tuned Qwen2-VL 7B as the baseline model.

Academic Service

Reviewer of AAAI 2026

Competitions

Multilingual Video Reasoning Evaluation Challenge @CVPR 2025.	Winner
Complex Video Reasoning and Robustness Evaluation Challenge @CVPR2025.	Runner-up
Long-Term VideoQA challenge of the LOVEU Workshop @CVPR2024.	Winner
Hour-long VideoQA challenge of the Second Perception Test challenge @ECCV2024.	Runner-up

Industrial Internships

Machine Learning Engineer Apr.2024 - Jan.2025

OpusClip (a16z Top50 GenAI Startup)

- Developed **Clip-Copilot**, an LLM-driven interactive video editing agent later evolved into core product **Clip Anything**.
- Designed **MM-Screenplay**, a multi-modal **video understanding** framework leveraging **WhisperX**, **GPT-4** and **Gemini** for retrieving visual and contextual information from **hour-long** videos based on user queries.

Machine Learning Engineer Dec.2023 - Mar.2024

FaceMind (AI Startup delivering Customized Local-Deployable Chatbots)

- Implemented a **RAG system** using **Milvus** vector database and **LangChain**, enabling **long-term memory**, **context recall**, and **consistent dialogue** for LLM-based chatbots, expanded retrieval chat history length by **up to 200x**.
- **Fine-tuned LLMs** (LLaMA2-7b, Mixtral-8x7b, Qwen-14b, etc.) into Chatbots using **QLoRA** with **LLaMA-Factory**, characterized models' self-awareness and dialogue tone, injected business-specific domain knowledge to the LLMs.

Data Scientist Aug.2023 - Dec.2023

ByteDance (Parent Company of Tiktok)

- **Fine-tuned an proprietary Visual Language Model** using **LoRA** with **Huggingface-PEFT** for commercial intent understanding, deploying it to identify and analyze business intent in short video content.
- Built a **zero-shot image-matching service** leveraging fine-tuned **CLIP**, **OWL-ViT**, **YOLOv8**, **OpenCV** and **Selenium**, capable of matching products in short videos with images crawled down from related advertising websites.

Technical Skills

Programming Languages: Python, Java, C/C++, SQL
Machine Learning: PyTorch, Huggingface Transformers, PEFT, Llama-Factory, Swift, NumPy, Pandas, Scikit-learn, Tensorflow
Platform, Libraries and Tools: Linux, Git, Conda, Azure, Docker, MySQL, Milvus, Google Cloud Platform