Zachary Tianyi Tang

+1 (437) 766-8781 | zach.tang@mail.utoronto.ca | e-ndorfin.github.io | Google Scholar

Education

University of Toronto — Toronto, ON
B.S., Computer Science; GPA: 3.89/4.00 — *Expected May 2028*

Work Experience

UofT Dynamic Optimization & Operations Management Lab — Toronto, ON
Researcher — *Sep 2025 - Present*

- Key member driving **robotics and Vision-Language-Action (VLA) industry research** aimed at automating large-scale **car battery assembly using UR5 robotic arm with Prof. Chi-Guhn Lee**
- Engineered simulation environment in **IsaacSim/Lab**, establishing a **teleoperation pipeline** that generated the **critical training data necessary** to successfully transfer robotic policies to **physical hardware**
- Specializing in fine-tuning open-source VLA models ($\pi_{0.5}$) to develop a generalizable wire manipulator

Philer AI — Remote
Software Engineer — *Apr 2025 - June 2025*

- Spearheaded a **10+ member team** to **successfully deploy an end-to-end customer intake legal AI assistant to production**, managing the entire stack including **full-stack deployment** on **Google Cloud/AWS Lambda**
- Executed **experimental RAG research** on advanced retrieval methodologies, including **multi-step retrieval** and **prompt decomposition** to facilitate **complex user legal queries over a large corpus of legal documents**
- Developed a custom, explainable RAG evaluation framework on **Pinecone**, benchmarking performance across metrics like correctness and recall, achieving an **average metric boost of 18%** through hyperparameter tuning
- Built frontend with **Next.js/React**, backend with **Python/Flask** including **LangGraph multi-LLM-agent workflow**

FORMS Syntron Information HK — Hong Kong SAR, China
MLE Intern — *May 2025 - June 2025*

- Developed internal tool for **HSBC's largest branch** (**US$3T+ in assets, 20,000+ employees**) to automate extraction of key data from unstructured police fraud reports for **financial fraud detection**
- Engineered a **distributed fine-tuning pipeline** for 70B LLM models using **LoRa and Transformers** to automate fraud data extraction, deploying and managing **large-scale SLURM jobs across an A100 GPU cluster**
- Led **weekly client meetings** to align evaluation metrics with their criteria, and boosted model accuracy on those metrics from ∼**70% up to 97%** through **few-shot prompt engineering and hard negative mining**
- Built **synthetic data pipeline** constructed from proprietary police report segments, growing dataset from **250 to 10,000+ realistic data samples** to improve domain-specific LLM performance

Research

Soundscape Audio Analysis on Hong Kong Pink Dolphin Population (Abstract) — *Nov 2023*

- Applied Fast Fourier Transforms to marine audio on **endangered pink dolphin conservation research** with **WWF**
- Presented at **IEEE 2024 Young Engineer's Conference** and the **American Geophysical Union Fall Meeting 2022**

Projects & Extracurriculars

Real-Time Video Super-Resolution | *PyTorch, TensorRT, CUDA, C++, OpenCV*

- Spearheading **real-time video super-resolution** research optimizing test-time inference with **TensorRT** and **CUDA**
- Profiled model bottlenecks using **NVIDIA Nsight Compute** and **PyTorch Profiler** to find high-latency operations
- Developing low-latency custom **C++/C** video capture pipeline with **OpenCV** to minimize **GPU/CPU overhead**

Multi-Agent RL Communication in Overcooked | *JAX, Python*

- Co-leading research in collaboration with **Prof. Matthew Taylor at the University of Alberta** investigating **multi-agent RL communication**
- Built Overcooked environment from scratch in **JAX**, achieving over **20x increase in training steps per second**

Technical Skills

Programming Languages: Python, JavaScript, Java, SQL, TypeScript, C#, C++
AI/ML Frameworks: PyTorch, Unsloth, Hugging Face, LangGraph, OpenCV, NumPy, SciPy, IsaacLab, IsaacSim, USDA
Full-stack Development: React, Next.js, Flask, Django, HTML/CSS, Tailwind CSS, shadcn/ui
Databases & Tools: SQLite, Git, Docker, SLURM, Pandas, jsPDF, Linux, Bash, Ubuntu, vim

Feb 2026