Steric Tsui

280 Dundas St W, Toronto, ON
437-445-5858 steric.tsui@mail.utoronto.ca steric-tsui stericishere steric-tsui.com

Experience

Squirl ASL **Toronto, ON**
ML Engineer *Sep 2024 – April 2025*

- Selected for the **Microsoft Startup Club** and funded by **Alterna Savings** to develop a B2B ASL translator.
- Contributed to a Temporal Convolutional Network (TCN) by performing a grid search over key hyperparameters and implementing a Cosine Annealing scheduler **resulting in an award-winning prototype.**
- **Reduced inference computation by 33% using a post-training dynamic frame sampling technique in Azure ML, prioritizing real-time smooth user experience without compromising model accuracy.**

Education

University of Toronto **Expected: May 2027**
Bachelor of Science in Computer Science and Statistics *Toronto, ON*

Projects

CoffeeChat – Website | *React, Typescript, RESTful API, Node.js,, PostgreSQL, Gemini API* **June 2025**

- Collaborated with a team of 3 developers to build CoffeeChat, a global professional networking platform using React, TypeScript, RESTful API, and PostgreSQL, resulting in 5,000+ signups across 3+ countries.
- Implemented semantic user search by ranking candidates with interest embeddings using the Gemini API and cosine similarity, integrated with PostgreSQL filters and served via RESTful APIs for mobile clients.
- Reduced user venue booking time by 80% by developing an AI-driven assistant that streamline the booking requests and automates confirmation and communication via email and AI voice agents.

Fault-Tolerant Order Processing System – Github | *Python, FastAPI, RabbitMQ, PostgreSQL, AWS EC2* **Aug 2025**

- Built microservices (Order, Payment, Inventory, Shipping) connected via RabbitMQ to process orders reliably under service failures.
- Implemented retries and a dead-letter queue, achieving 99% recovery rate in failure simulations.
- Deployed with Docker Compose to AWS EC2, with Cloud Watch monitoring and CI/DI pipeline for queue depth, latency, and service uptime.

Interactive AI multi-Agent Simulation – Github | *LangGraph, Docker, FastAPI, Django, WebSocket* **May 2025**

- Implemented an episodic-based 10-agent simulation using LangGraph, showcasing multi-agent social behaviors and interaction.
- Re-architected the Agents framework using the **PIANO architecture**, enabling scalable and long/short-term memory-efficient agent behavior and decision-making.
- Optimized FastAPI interaction by **consolidating context management and shared memory access**, reducing per-step token usage from 50 to 5 and minimizing latency and cost.

Technical Skills

Languages: Python, Java, HTML/CSS, JavaScript, React, PostgreSQL, My SQL, .Next, Typescript, PostgreSQL, MongoDB, Redis
Frameworks: LangChain, llamaindex, FastAPI, RestfulAPI, Streamlit, SpringBoot
Rest: SAS & SDLC & MATLAB programming, CI/CD pipelines, DevOps
Certification: Oracle Cloud Foundations Associate, Google Cloud Essentials

Leadership / Extracurricular

UTMIST (University of Toronto Machine Intelligence Student Team) **Toronto**
Researcher *Sep 2025 - Present*

- Collaborated with a research team in UTMIST, North America's largest student-led AI/ML organization, to develop a GAN-based computer vision pipeline using PyTorch and Scikit-learn, enhancing video resolution from 480p to 720p.

The AI Collective **Toronto, ON**
Event Coordinator *May 2025 - Present*

- Facilitated a monthly Toronto based AI-focused Coffee Chats, increasing member engagement and fostering networking opportunities within a 70,000+ member community

UofT AI **Toronto, ON**
Conference Team *Jun 2024 - Present*

- Acted as a key communication link between multiple internal teams to coordinate planning & logistics for the conference