Steric Tsui

280 Dundas St W, Toronto, ON

☎ 437-445-5858 ✉ steric.tsui@mail.utoronto.ca steric-tsui stericishere steric-tsui.com

Education

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

Experience

Flymingos– Website **Toronto, ON**
Software Engineer Intern *May 2025 – Aug 2025*

- **Boosted lead match accuracy by 35%** by architecting a semantic matching pipeline leveraging OpenAI **embeddings** and **Chroma vector search**, enabled intelligent pairing through real-time **vector similarity scoring**.
- Scaled backend infrastructure to **support 500+ concurrent users** by deploying a fully serverless architecture using **Node.js Clou**d Functions and **Firebase**, streamlined authentication and event-driven data workflows.
- Automated high-volume product data collection by building a robust **Python scraping syste**m with **BeautifulSoup** and custom parsing logic, **reducing manual input by 40%** and **improving data accurac**y across seller platforms.

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

- 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.
- Selected for the **Microsoft Startup Club** and funded by **Alterna Savings** to develop a B2B ASL translator.

Projects

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

- Collaborated with a team of 3 to develop CoffeeChat, a global professional networking platform using **React**, **TypeScript**, **RESTful API**, and **PostgreSQL**, resulting in **5,000+ signups across 5+ countries**.
- **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.
- 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.

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

- Built a fault-tolerant, distributed order system with **99% recovery rate during failures**, using microservices (Order, Payment, Inventory, Shipping) and **RabbitMQ** for reliable, async communication.
- Deployed the system using **Docker Compose** on **AWS EC2**, with CloudWatch monitoring and a **CI/CD pipeline** to track queue depth, latency, and uptime metrics.

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

- Re-architected the Agents framework using LangGraph and **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, TypeScript, JavaScript, HTML/CSS, SQL, React, Next.js, Node.js
Frameworks: FastAPI, Django, Express, Spring Boot, LangGraph, LangChain, Streamlit, RESTful API, WebSocket
Databases: PostgreSQL, MySQL, MongoDB, Redis, Firebase
DevOps: Docker, Git, CI/CD, Google Cloud (GCP), AWS (EC2, CloudWatch), RabbitMQ, Kubernetes
Certifications: Oracle Cloud Foundations Associate, Google Cloud Essentials

Leadership / Extracurricular

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

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

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