

STERIC TSUI

280 Dundas St W, Toronto, ON

☎ 437-445-5858

✉ steric.tsui@mail.utoronto.ca

🌐 [steric-tsui](https://steric-tsui.com)

🔗 [stericishere](https://stericishere.com)

📁 [Portfolio](#)

Education

University of Toronto

Bachelor of Science in Computer Science and Statistics

Expected: May 2027

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 Cloud Functions** and **Firebase**, streamlined authentication and event-driven data workflows.
- Automated high-volume product data collection by building a robust **Python scraping system** with **BeautifulSoup** and custom parsing logic, **reducing manual input by 40%** and **improving data accuracy** across seller platforms.

Squirrel ASL

Toronto, ON

ML 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

Orbit (Winner @ Hack the Valley 2025) – Devpost | *Next.js, Three.js, FastAPI, LangGraph, Redis, Docker* Sept 2025

- Designed a agentic AI pipeline using **LangGraph** and **Gemini**, orchestrating **11+ interruptible** AI node with **fallback** options and persistent state recovery via **Redis-backed checkpointing**.
- **Reduced generation latency by 66%** through parallelized image and 3D model rendering (**Trellis, Nano Banana**) and streamed client updates with **SSE** at less than 100ms intervals.
- Deployed a production-ready system using **Docker Compose**, with robust error handling, CORS-secured APIs, and modular frontend/backend architecture; **frontend design was specifically recognized by judges**.

CoffeeChat | *React, Typescript, RESTful API, Node.js,, PostgreSQL, Gemini API*

[Website](#)

- Collaborated with a team of 3 to develop a mobile APP 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 automation tools that streamline the booking requests and automates confirmation 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.

Distributed Search Engine (Mini Elasticsearch) | *Java, Spring Boot, Docker, JMH*

[Github](#)

- Developer an internal full-text search engine with inverted index and TF-IDF, sustaining **5K+ queries/sec under 200ms** via JMH/JMeter benchmarking, achieving **1.5× faster retrieval** with consistent hashing and Java concurrency.

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

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

UTMIST (University of Toronto Machine Intelligence Student Team)

Toronto, ON

ML Researcher

Sep 2025 - Present

The AI Collective

Toronto, ON

Event Coordinator

May 2025 - Present