Steric Tsui

203 College St, 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) and performing a grid search over key hyperparameters and implementing a Cosine Annealing scheduler, **contributing to 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 2028**
Bachelor of Science in Computer Science and Statistics *Toronto, ON*

Projects

TinyProof – Github | *PyTorch, Google Cloud Platform, LeanDojo, Lean4, Docker, Jira* **Jan 2025**

- Conducted original research under Dr. Mohammad to build an RL-based theorem prover extending AlphaProof using **R'max Tree Search**, **achieving 87% on college-level proofs surpassing expected baselines**.
- Engineered an **ETL pipeline** leveraging LeanDojo to extract formal proof data from Lean4 repositories, transforming and loading it into scalable GCP storage for downstream RL training.
- Containerized workflows with **Docker** and coordinated iterations via **Jira/CI**, ensuring reproducibility across research iterations.

Generative Agents: The Dating Show – Github | *LangGraph, Docker* **May 2025**

- Re-architected the original Generative Agents framework using the **PIANO architecture arXiv:2411.00114)**, enabling scalable and memory-efficient agent behavior.
- Designed and implemented a narrative-driven "dating show" environment, showcasing emergent social behaviors through **advanced prompt engineering, episodic memory, and multi-agent reasoning**.
- Reduced API overhead and token consumption by **consolidating agent context and shared memory into fewer, reusable API calls**, improving system efficiency and lowering cost.

RL agent play Mini-Motorway – Github |*Python, PyTorch, OpenCV, Stable-Baselines3, PyAutoGUI* **Aug 2025**

- Developed a **Deep Q-Network (DQN)** agent with a **CNN architecture** to learn game dynamics directly from **raw visual input**.
- **Engineered a shaped reward function** to overcome sparse feedback by integrating signals such as **trip duration**, **road efficiency**, **congestion level**, **construction usage**, and **point accumulation**.
- **Translated discrete policy outputs** (max-Q actions) into pixel-accurate mouse commands via PyAutoGUI for real-time in-game control, achieved 500+ pts.

ML for Adaptive Education – Github | *Pytorch, NumPy* **Jun 2025**

- Led in a team of four to develop a machine learning model in Python for p**redicting student performance and recommending question difficulty** tailored to learning ability.
- Enhanced traditional Item Response Theory (IRT) to a three-parameter model, **outperformed baseline models by 15%** including neural networks, KNN, SVD, and Bagging.

Technical Skills

Languages: Python, C++, Java, HTML/CSS, JavaScript, PostgreSQL
Frameworks: Tensorflow, Pandas, Numpy, Scikit-learn, LangChain, FastAPI, streamlit
ML: Linear & Logistic Regression, SVM, KNN, Decision Trees, Random Forests, GLM, GAM
Deep Learning: MLP, CNNs, LSTM, TCNs, Attention Mechanisms, Backpropagation, (ReLU, Sigmoid, Tanh)
RL: Monte Carlo, RMax, DQN, Q-Learning, PPO, GRPO
Cloud: GCP (GCS, Cloud Scheduler Functions, BigQuery), AWS (S3, Sagemaker), Jupyter
Certification: AWS Machine Learning Engineer-Associate, Oracle Cloud Foundations Associate, Google Cloud Essentials

Leadership / Extracurricular

UofT AI **Jun 2024**
Conference Associate

The AI Collective **May 2025**
Event Coordinator