Evan Zheng

(347)-931-5233 | evan.ty.zheng@gmail.com | linkedin.com/in/evtyz | github.com/evtyz

EDUCATION

University of Waterloo   September 2021 – April 2026 (Expected)
Bachelor of Computer Science (Co-op)   *Waterloo, ON*
- Cumulative Average: 96.4%, Faculty Average: 97.6%, Exchange at National University of Singapore

EXPERIENCE

Jane Street   Starting May 2025
(Incoming) Software Engineering Intern   *New York, NY*

Tesla   January 2025 – April 2024
Machine Learning Intern (Autopilot)   *Palo Alto, CA*

MAGI   May 2024 – August 2024
Machine Learning Engineer   *New York, NY*
- Directed the design and implementation of the main LLM/RAG pipeline for one of the company's core products, an automated relationship therapy app, using LangChain, Firebase, Groq, ChromaDB, and GPT-4
- Reduced user-facing chatbot latency from 8s to 3s by profiling performance with LangSmith, parallelizing the chatbot's chain-of-thought reasoning, optimizing prompts, and experimenting with different models
- Built large-scale unsupervised conversation analysis tools using OpenAI text embeddings and K-means clustering to identify and categorize user conversation patterns and intelligently suggest chatbot improvements
- Led the product development of a new AI video interviewing platform by conducting user research, writing design docs, and scoping engineering tasks, to reduce internal recruiter resume-screening time by 65%

Waabi (formerly Uber ATG)   January 2023 – April 2024
Machine Learning Research Intern   *Toronto, ON*
- Performed fundamental machine learning research for autonomous driving, supervised by Prof. Raquel Urtasun
- Designed a novel transformer-based neural network architecture in PyTorch for lane prediction, achieving a new state-of-the-art by reducing detection and topology errors by 59% compared to competing baselines
- Created a highway graph traversal library for internal use to enable large-scale high-definition automated mapping for 300km+ of highways, reducing manual annotation time by more than 50%
- Parallelized and vectorized Python code in internal model training pipelines to increase training speed by 104% and reduce monthly AWS cloud compute costs by 51%

Faire   May 2022 – August 2022
Software Engineer Intern   *Toronto, ON*
- Proposed and spearheaded an overhaul of advertisement rendering on search pages with 400,000+ daily visits to increase daily ad views on affected pages by 40%+ and speed up estimated development time of future ads by 30%
- Built and tested reusable, responsive React components with TypeScript, Styled-Components, MobX, and Jest to block high-risk financial instruments in payment flows affecting $2+ million USD of daily transaction volume
- Developed recurring automated email tasks using Kotlin, Hibernate, MySQL, and Cron to notify 3,000+ customers of impending charges against their account and reduce associated chargeback dispute rates by 20%+

AWARDS

Top 100, Canadian Mathematical Olympiad (Invitational IMO Qualifier)   March 2021
Distinction, Top 10%, Canadian Computing Competition, Senior Division   February 2021
Gold Division, United States of America Computing Olympiad   March 2020

TECHNICAL SKILLS

Languages: Python, C++, Java, C, Kotlin, Scala, TypeScript, JavaScript, C#, SQL, R
Technologies: PyTorch, NumPy, LangChain, MySQL, Jax, Git, Docker, Google Cloud