Efe Tascioglu

Software Engineer & ML Researcher

Experience

HUAWEI | Machine Learning Researcher

September 2024 - April 2025 | Toronto

- Designed and distilled large language models (LLMs) to implement Speculative Decoding, achieving a lossless 3x performance increase.
- Designed and developed an LLM quantization strategy using EfficientQAT, outlier removal, mixed-precision quantization, double-quant, and vector quant to achieve a 4x speed and memory improvement with 8% rise in perplexity.

AMAZON | Software Development Engineer

May 2024 - August 2024 | Toronto

- Designed and developed an improved versioning architecture for Alexa using C++, directly reducing compatibity errors for over a dozen teams.
- Developed a new low-maintenance notification and API test suite for inter-package communications preventing future downtime.
- Contributed to the Alexa Hybrid Engine to reduce device failures, improve unit test coverage, and simplify existing codebase.

NERVEX NEUROTECHNOLOGIES | Artificial Intelligence Developer

May 2023 - August 2023 | Toronto

- Performed feature engineering and unsupervised anomaly detection to detect seizures in epileptic patients for improved treatment.
- Designed algorithms to denoise biosignals and extract vital signals with 50% higher accuracy using noise detection and removal methods.
- Developed and deployed microservices & algorithms at scale using AWS and Docker, and optimized existing algorithms to perform 3x faster.

TUNESCAPE | Co-Founder, Full Stack Developer

May 2022 - December 2022

- Founded a startup helping individuals discover new music on both web and mobile through the analysis of technical characteristics and AI.
- Developed the backend for the application using Flask with Python, and frontend for the supported platforms using Flutter.

Projects & Activities

UOFT MACHINE INTELLIGENCE STUDENT TEAM (UTMIST)

Technical Project Manager, Director, Reinforcement Learning Developer

September 2022 - April 2024

- Evolved and trained creatures with custom morphologies and neural structures using reinforcement learning (RL) in Unity.
- Lead the design, development, and management of a new multiagent collaboration model.

HACK THE NORTH 2023 - IRIDIS | Computer Vision and Full Stack Developer

September 2023

- Developed a physical heads-up display helmet record attendee faces and information to display it when you encountered them.
- Used Computer Vision (CV) to scan and recognise faces, and created a realtime database to store users for further recollection.
- Designed a concurrent backend API and deployed our code to create a fast, scalable system while allowing rapid prototyping.

Skills

PROGRAMMING

C++ • C • Python • C#
Javascript • RISC-V • Flutter

TOOLS/PLATFORMS

PyTorch • HuggingFace • Git
Docker • Linux • AWS • NodeJS

Awards

2024	Dean's List (x6)
2022	Finalist - CIBC AI Competition
2022	Finalist - UTEK Competition
2022	Amateur Rocket Certification :)
2021	Silver - Canada Science Fair
2021	Featured on Make Magazine

Education

UNIVERSITY OF TORONTO

Engineering Science - Machine Intelligence

September 2021 - Present

Cum. GPA: 3.84 / 4.0

Courses

(CSC413) Deep Learning
(ECE353) Operating Systems
(ROB311) Reinforcement Learning
(ECE358) Algorithms and Data Structures