Brian Li

b77li@uwaterloo.ca · 647-523-8026 · brianljx.vercel.app/ · Brian Li

Skills

Programming Languages:	Python, C++, Java, HTML/CSS, Tailwind CSS, JavaScript, TypeScript, SQL
Frameworks:	React, Next.js, Node.js, Langchain, PyTorch, TensorFlow, Selenium, Jenkins, PostgreSQL
Tools and Technologies:	Git, Linux, AWS, Docker, Kubernetes, ClearML, STM32, Arduino, Raspberry Pi, OpenCV, MongoDB FPGA, FFmpeg, ElasticSearch, Figma

Relevant Experience

Eon Media
AI/ML Developer Intern April 2024 - August 2024

- Integrated NLP capabilities into **Amazon OpenSearch cluster**, performing deployment of **TorchScript models** and **ingest pipelines**; utilized **BERT** models for conversion of metadata into dense vectors for **KNN-indexing**
- Designed and trained a custom **binary classifier** from scratch to identify logos in images using **CNNs** and **transfer learning** of the **ResNet50 model**. Achieved **93%** accuracy, improving existing classifier's performance by **45%**
- Enhanced text detection model by incorporating dewarping functionality using a **Generalized Additive Model (GAM)** and **Bresenham's algorithm**; achieved 50% increase in recall rate and 80% reduction in runtime of pod
- Facilitated **cloud deployment of pods** through **EKS clusters**, optimized **containerization** of various pods and algorithms, reducing **docker image** sizes by **40%** and **resolving bottlenecks in pipeline**

Wat.AI
Lead Neural Network Developer September 2023 - Present

- Led team in designing and prototyping **sparse** and **denoising autoencoders** using **PyTorch** for compression of IoT cybersecurity data (CICIOT 2023); optimized draft autoencoders with **learning curve analysis**
- Contributed to writing and updating team substack articles to highlight progress of team project; helped with testing of **machine learning models** focused on cyberattack detection for IoT devices

Waterloo Formula Electric
Lead Firmware Developer September 2023 - Present

- Created hardware-in-the-loop (**HIL**) tests using **Python** to determine expected behavior of electric car unit and **RTOS**; utilized **STM32**, **Virtual Box** and **Vagrant** to find translation in code and firmware input values
- Analyzed the Battery Management Unit (**BMU**) through stimulating its state of charge; conducted in-depth examination of **multithreading** firmware code in **C** and schematics to identify source variables and functions

Projects

"A+I" — *Next.js, FastAPI, PostgreSQL, SQLAlchemy, Langchain, docTR, CUDA*

- Developed an **automated** online test grading system using **LLMs** to analyze handwriting and evaluate answers
- Implemented authentication with **NextAuth and OAuth 2.0**, integrating 5+ providers and securing admin routes; optimized **SQLAlchemy ORM queries** for users, classes, and test data in large-scale educational environments
- Trained and developed end-to-end **OCR model** on handwriting dataset, achieving a **120%** improvement in detection accuracy and a **35%** reduction in error rates
- Architectured an inference system leveraging **Langchain memory stores** for continuous question-answering capabilities and **prompt routing** for catered student response evaluation

"Spotify-Roots" — *Next.js, Flask, Langchain, Llama.cpp, Qdrant*

- Developed a web application that uses a **RAG-based approach** to provide origin information of song, album, and artist, integrating Genius API for annotation extraction and **Llama 3 model** for inference and conversion of data
- Built with **OAuth2** to handle Spotify account logins, allowing users to perform origin analysis on personal playlists

"RizzVision" — *Raspberry Pi, React Native, Linux, MongoDB, Groq, OpenCV*

- Engineered glasses that captures **real-time conversational data** via camera and microphone, offering live sentence scoring and suggestions, followed by a post-interaction analysis based on conversation context
- Architected a **LLM agent** using **Groq** inference that performs tailored inference using stored contextual memory
- **Built embedded system** using **Raspberry Pi**, integrating **AssemblyAI** for Speech-To-Text, **OpenAI** for Text-To-Speech, and **OpenCV** for facial landmark detection and semantic analysis

Education

University of Waterloo Waterloo, ON
Candidate for Bachelor of Applied Science in Computer Engineering September 2023 - Present