Daniel Yu

✉ talldanielyu@gmail.com | 📞 650-278-0824 | in itsyangyu | ⌨ kalig0 | 📍 US Citizen

OBJECTIVE: Master of Computer Science student at WPI with 1+ years of machine learning and software development experience seeking Machine Learning Engineer or Software Engineer opportunities.

EDUCATION

Computer Science — *Masters*   AUGUST 2023 - PRESENT
Worcester Polytechnic Institute   GPA: 4.00

Mathematical Sciences — *Bachelor of Science*   AUGUST 2018 - JUNE 2022
University of California, Santa Barbara

EXPERIENCE

CVTE — *Machine Learning Engineer Intern*   MAY 2024 - JULY 2024
- Developed and implemented logic enhancements for a machine learning model utilized in over 5 million classrooms, improving the generated classroom analysis by 20%
- Used **Python**, **Matplotlib**, and **Pyplot** to process 1000s of classroom data to analyze and identify dips in classroom activity level
- Collaborated in an **Agile/Scrum** team of 8 engineers to deliver new large language model implementations every 2 weeks

ReadMKT — *Software Engineer*   AUGUST 2022 - AUGUST 2023
- Developed a rich text editor in **Javascript** enabling users to edit content, images, links, and other components
- Built an infinite scrolling feature on the main page, dynamically displaying up to 10 posts at a time and loading more groups of 10 as users scroll to the bottom
- Implemented a rate limit management strategy to ensure complete API data fetching from financial data providers
- Used **MongoDB** to store, update, and retrieve user information, trade details, etc
- Deployed product to **AWS Lambda** and **Cloudfront** supporting dual endpoints IPv4 and IPv6

7G BioVentures — *Full Stack Developer Intern*   MAY 2022 - AUGUST 2022
- Implemented an internal system to manage investment project information, boosting employee productivity by 50%
- Built a user-friendly, interactive, and component-based system using **React** and **Express**
- Used **MongoDB** to manage data storage and retrieval

PROJECTS

Research Project — *ASSISTments*   MAY 2024 - PRESENT
- Fine-tuned large language models using **QLoRA** to enhance the accuracy of feedback for responses to math questions
- Trained models using **TensorFlow** on millions of student response data to develop and implement effective improvement strategies for student feedback
- Implemented vector database using **FAISS** to generate feedback and score for studend responses with retrieval augmented generation, improving large language model generated content accuracy by 30%

Personal Project — *MotionWiz*   DECEMBER 2023 - JANUARY 2024
- Developed backend with **Python** which uses OpenAI large language model to generate exercises based on user queries
- Hosted on **AWS** to get the API endpoint
- Leveraged **Next.js** for the interactive interface and stylized with **Tailwind CSS**
- Hosted frontend on Vercel for public access

SKILLS

- **Languages**: Python, JavaScript, Typescript, Java, HTML, CSS, SQL, C++, Rust
- **Environment/Framework**: Pytorch, TensorFlow, Numpy, Pandas, React, Next.js, AWS, Express, FastAPI
- **Databases**: MongoDB, MySQL, SQLite3
- **AI/ML Techniques**: Large Language Models, Fine-tuning, Retrieval Augmented Generation (RAG), Natural Language Processing (NLP), Computer Vision (CV), Prompt Engineering