

SHI JIE YU

☎ 347-799-5002 ✉ yshijie1999@gmail.com [in LinkedIn](#) [GitHub](#) [Website](#)

Education

New York University

Master's in Computer Science (GPA: 4.0/4)

Sep 2024 – May 2026 (Expected)

New York, NY

National University of Singapore

Bachelor of Science in Business Analytics with Honors (Distinction)

Aug 2020 – May 2024

Singapore

Technical Skills

Languages: Python, SQL, C, C++, JavaScript, Typescript, Solidity, Java

Full Stack Development: React, Next.js, MongoDB, Tailwind, Prisma, Google Cloud Platform, Amazon Web Services

Machine Learning: Pytorch, Transformers, Llama 3, DeepSpeed, Scikit-Learn, Tensorflow, LitGPT, Vertex AI

Certifications: GCP Certified Professional Machine Learning Engineer

Experience

New York University

Jan 2025 – Present

Graduate Teaching Assistant

New York, NY

- Managed a graduate-level operating systems class of 30 students by facilitating weekly consultation sessions, designing grading rubrics and model solutions, and grading and delivering detailed feedback on assignments.

Tensorplex Labs

Apr 2024 – Aug 2024

AI Engineer

Singapore

- Spearheaded continued pre-training of Llama 3 language models on billions of tokens' worth of web3 corpora using state-of-the-art techniques such as domain adaptation, LoRA, and task arithmetic.
- Conducted post-training (SFT and RLHF) on frontend UI datasets using Qwen2.5 Coder model and LLaMA-Factory framework, achieving up to 2x performance gains on HumanEval and other coding benchmarks.
- Led research and implementation of a novel and efficient approach to parameter-efficient model merging of LoRA modules; authored a research paper on the proposed method, assuming the first author role.

WorldQuant BRAIN

Nov 2023 – Apr 2024

Quantitative Research Consultant

Singapore

- Placed 12th out of 500+ participants in WorldQuant's NUS Alphathon 2023; thereafter, onboarded as a quantitative research consultant to develop over 30 alphas (predictive signals) for the US market via market-neutral strategies.

Autodesk

May 2022 – Oct 2022

Data Analyst Intern

Singapore

- Designed interactive Looker dashboards for 3 Autodesk product lines (ReCap Desktop, ReCap Viewer, ReCap Pro), yielding actionable business intelligence capable of informing strategic decisions at the director level.
- Engineered, scheduled, and monitored 20+ daily and weekly Big Data extract, transform, and load workflows on terabyte-scale data with SQL, Jenkins, and Apache Airflow for analytics with Qubole and monitoring with Mixpanel.
- Initiated a machine learning project to predict EC2 peak memory usage using ensemble methods (Random Forests, XGBoosts, LightGBM) and deep neural networks, reducing the cloud cost of AWS EC2 deployments by 50%.

Projects & Extracurriculars

Active Portfolio Management with ML and Time Series Forecasting

Spring 2025 – Present

AI Research Assistant

New York University

- Implemented a Quaternion transformer architecture from scratch for time-series forecasting and conducted pretraining and finetuning; achieved a comparable performance on finance benchmarks to models with 4x more parameters.

Lightning-AI/litgpt

Fall 2024 – Present

Tier 2 Open-source Partner / Top 10 contributor to a project with over 11k GitHub stars

Lightning AI

- Implemented state-of-the-art LLMs, including Phi-4 and Qwen2.5 series models; added general enhancements like introducing new architectures as well as reducing overhead in pretraining and other compute intensive operations.

Rugged

Spring 2025

Full Stack Engineer

Personal Project

- Designed and implemented a REST API for querying live blockchain data from Gecko Terminal, and seamlessly integrated its endpoints into a Next.js frontend; leveraged ElizaOS to build an LLM agent for real-time crypto insights.

NUS Fintech Society Machine Learning Department

Fall 2023 – Spring 2024

Director

National University of Singapore

- Managed a 60-member machine learning interest group, creating learning and networking opportunities for members via workshops and hands-on projects; managed over 10 ML projects in diverse domains, including NLP, LLMs, and XAI.