CORLEONE YANG

GitHub LinkedIn Website yahe.yang@gwu.edu

EXPERIENCE

VEA — California, United States
Full-Stack Engineer — Sep 2024 – Present

- Collaborated in an **agile** team to fix **20–25** weekly bugs in data **rendering**, **databases**, **concurrency**, and **APIs**. Used JS (**React**, **Node.js**), Postman, AWS CloudWatch, and ECS logs for full-stack troubleshooting. Wrote Jest tests, maintained 80% SonarQube coverage, and enforced code reviews. Delivered features on schedule in a fast-paced environment.
- Tested Stripe API with Postman, upgrading old endpoints to reduce technical debt. Refactored calls, cutting requests by **30%** and boosting load speed by **25%**. Integrated third-party APIs for broader functionality. Wrote frontend and backend unit tests with Jest, ensuring **80% SonarQube coverage** and passing the quality gate.
- Improved a RAG recommendation system using **LangChain**, **AWS OpenSearch**, and **OpenAI APIs**. Introduced a page caching mechanism and a new storage strategy to fetch the last recommendation result, reducing **OpenAI tokens** consumption without compromising user experience.
- Built a real-time rendering sidebar with React and Node.js, enabling users to view in-progress flow data and improving overall workflow efficiency. Additionally, developed a **GenAI backend module** and integrated its APIs into various frontend features (e.g., allowing users to input symptoms for an AI-revised version or extract key information from the entire flow).

Baynovation — San Jose, United States
Software Engineer Intern — May 2024 – Aug 2024

- Built ML workflows in **AWS SageMaker** to perform **feature engineering** and train a Transformer-based **time series** forecasting model for loan default prediction. Preprocessed data with **AWS Glue** and integrated the trained model into the company's EKS service API, boosting prediction accuracy by **18%**.
- Employed **supervised fine-tuning** and **Lora** to train a preference model to improve chatbot alignment. Additionally, developed two **demo projects**: (1) an AI-powered job application tool using Django for automated matching and one-click applications, deployed to AWS EC2 with Terraform and GitHub Actions, and (2) a Django-based investment simulation web app for ETF return analysis using historical data (e.g., SPY, QQQ, DIA).

Tencent — Shenzhen, China
Machine Learning Engineer Intern — August 2022 - October 2022

- Constructed a machine learning environment using **Conda** and **Ubuntu**. Developed a reinforcement learning framework with **OpenAI Gym** and **PyTorch** to implement the **Deep Q-Network (DQN)** algorithm, achieving an 86% task completion accuracy for smart agents in simulated environments.
- Implemented a **BERT4Rec-inspired ranking system** using TensorFlow, focusing on the **rank** stage of the recommendation pipeline. Employed a sequence-based model to process user click data, leading to a 7% improvement in **click-through rate (CTR)** within the test environment.

XiuNeng Capital — Shenzhen, China
Quant Intern — May 2021 - August 2021

- Designed and implemented an **Event Study Algorithm** using Python, pandas, and matplotlib to analyze the impact of specific events on stock prices. Improved event impact prediction accuracy by **18%** through integrating data from Bloomberg Terminal and internal databases.
- Applied statistical methods and machine learning techniques for **factor mining**, including correlation analysis, PCA, and cross-sectional regression. Incorporated significant factors into a multi-factor model for portfolio construction. Back-tested the strategy using Python, achieving a **12%** increase in risk-adjusted returns compared to the benchmark.

PROJECTS

Smart Inventory Management System | Link

- Engineered a full-stack inventory management system using **Next.js**, **Material UI**, and **Firebase**, integrating **OpenAI Vision** for image classification and **HuggingFace's Llama3.1** for inventory management suggestions, with **CI/CD** implementation via **GitHub Actions** and **Vercel** for streamlined deployment.

Document Chatbot

- Developed a full-stack document chatbot using **Express.js** and **React**, implementing **RAG** with **AWS Bedrock API** and **Pinecone** for efficient vector storage and retrieval. Integrated **LangChain** for enhanced language model interactions. Implemented **CI/CD** pipeline using **GitHub Actions** for automated testing and deployment to **AWS EC2**, ensuring seamless updates and scalability.

TikTok Shop Content Recommendation Engine

- Developed a content recommendation engine for TikTok Shop using **Golang**, **Python**, and **TypeScript**, with **Kafka** for real-time data streaming and **MySQL** for data storage. Implemented personalized product and content suggestions based on user keywords and tags, significantly improving user engagement and driving sales growth.

School Projects

- Implemented a polynomial derivative calculator, A* algorithm for 8-puzzle problem, lightweight database using AVL trees, and matrix operation algorithms, all utilizing **Java.**(Data Structure). Implemented a custom OS kernel with thread creation and signal handling, developed a multi-threaded frog river-crossing simulation, designed a GPU-accelerated virtual memory system using **CUDA** with page replacement algorithms, and created a GPU-based file management system featuring root directory and I/O operations. (OS)

CERTS

Cloud Certificates: Azure AI Engineer Associate | AWS Solution Architect Associate | AWS Data Engineer Associate

Kaggle Medals: Explainable AI for Patent Professionals | Chatbot Arena Human Preference Predictions

SKILLS

Programming Languages:	Python | Java | C | C++ | C# | CUDA | HTML/CSS | JavaScript/TypeScript | MATLAB | R | Go
Backend/Frontend:	Node.js | Express.js | Next.js | React | SpringBoot | Flask | Django | PostgreSQL | MongoDB | Redis
Tools:	AWS/Azure/GCP | Kubernetes | Docker | Git | Jenkins | Terraform | Postman | Huggingface | LangChain
Machine Learning/Data:	PyTorch | TensorFlow | NumPy | Pandas | Matplotlib | Scikit-learn | Databricks | Kafka | Airflow

EDUCATION

The George Washington University — **Washington, DC**
Master of Science in Information Technology — *Aug 2023 - May 2025*

Chinese University of Hong Kong — **Shenzhen**
Bachelor of Science in Mathematics — *Sep 2019 - May 2023*