YiAn Chen

916-597-3444 | yian.chen261@gmail.com | github.com/yianan261 | Philadelphia, PA | https://yianan261.github.io/Personal_Page/

EDUCATION

University of Pennsylvania *Sep 2024 - May 2026*   **Philadelphia, PA**
Master of Science in Engineering, Artificial Intelligence (Software Engineering concentration)
Northeastern University *Jan 2022 - May 2024*   **San Jose, CA**
Master of Science, Computer Science | GPA: 3.96/4.0

WORK EXPERIENCE

Stevens Institute of Technology *(Schaefer School of Engineering)*   **Hoboken, NJ**
Research Assistant   Sep 2024 – Present

- Implement knowledge graph compression on graph Retrieval Augmented Generation and model compression using quantization techniques with Llama-3-70B-Instruct

Northeastern University *(Khoury College of Computer Science)*   **San Jose, CA**
Research Assistant; paper accepted and peer reviewed, awaiting publication at HCI-COLLAB 2024   Jan 2024 – Aug 2024

- Designed a **concurrent, multi-threaded system** to automate the aggregation and classification of irresponsible artificial intelligence incidents and developed interactive data visualization interface using **reactive programming**
- Leveraged GPT-4 API with **prompt engineering** techniques to fill out 70% missing fields and **improved accuracy by 450%**

Amazon   **San Francisco, CA**
Software Development Engineer Intern (Amazon Music Royalties Calculation Engine Team)   May 2023 – Aug 2023

- Deployed **end-to-end system** that flags discrepancies in worldwide vendor and user data leveraging various **AWS tools**, automated with **Apache Airflow** orchestrator, and impacting over **82 million** end-users globally
- Developed internal and external **GET** and **WRITE RPC APIs** in **Kotlin** using **dependency-injection** with **Google Guice**
- Streamlined big data aggregation with **Apache Spark** in **Scala** to output to **AWS S3** through **EMR serverless**

Starts Foundation   **Remote**
Full-Stack Developer/Project Leader of team of five   Oct 2022 – Present

- Deployed a **serverless full-stack web app** for the Starts Foundation leveraging **React, Gatsby**, **Node.js**, **Express.js**, **MongoDB**, and **GraphQL** to support humanitarian relief in Nepal
- Designed and built a **Content Management System** to streamline user experience and content generation

TECHNICAL SKILLS

Programming Languages: Java, Python, C/C++, JavaScript, Kotlin, Golang, Scala, Solidity, SQL, R
Databases: MongoDB, PostgreSQL, Redis, Liquibase, MySQL, Amazon Aurora
Frameworks/Tools: AWS, Docker, React, Kafka, Spring, Django, Google Guice, Gradle, Tensorflow, PyTorch, Spark, Airflow
Software Development: Distributed and Parallel Computing, API Development, Web Development, Full-Stack Development, CI/CD
AI and Machine Learning: Machine Learning, Linear Optimization, NLP, RAG, Data Mining

SELECTED PROJECTS

eBay 2024 Machine Learning Competition - Develop ML model to identify motor parts to vehicle compatibility   Sep 2024 – Present
Training ML models using a dataset of 5 million listings. Preprocessing and extracting features from unstructured data

Gemini API Developer Competition - Deployed Flutter app that personalizes place recommendations   June 2024 – Aug 2024
Designed recommendation system and AI tour guide with **Gemini APIs** using **Flask** for full-stack customized place-discovery app

LLM Compatibility Test App - Deployed app that leverage LLM to flag Android and iOS application issues   Jan 2024 – April 2024
Deployed **Docker** image of backend **Django** app on **AWS ECS** using continuous integration and deployment with **Github Actions**

Distributed Systems Project - Multi client-server distributed systems   Dec 2023 – Jan 2024
Built a **multi-client-server distributed application** in **Golang** using a **distributed microservices based cluster** over **TCP protocols**

Prognosis of Diabetes Early Readmission - Data mining project training machine learning models for prediction   Nov 2023 – Dec 2023
Leveraged **data mining techniques** and trained six **machine learning models** to stack and **improved model performance by 10%**

Friend Recommendation System - Data mining project that recommends social network friends   Oct 2023 – Oct 2023
Developed friend recommendation system using **MapReduce** and **Apache Spark** for large-scale parallel processing

Cordiance Experiential Project - Natural language processing and advanced algorithm project for data analysis   Nov 2022 – Dec 2022
Implemented advanced **string matching algorithm** and designed **4-level prefix tree** data structure to classify large industry datasets

Tutor Match App - **MERN** stack web app for users to search, match, reserve tutors   Oct 2022 – Nov 2022
Deployed responsive **Node** app using **React** and implemented login, search, tutor-booking, scheduling, and membership endpoints

Career Helper App - **Full-stack web app** that helps users plan, search, manage job applications   Sep 2022 – Oct 2022
Engineered backend with **Express.js,** and **MongoDB** for **full CRUD operations** and dynamic frontend with vanilla **JavaScript**

Steganography Project - **Full-stack web app** that encodes and decodes secret messages in image pixel bits   June 2022 – Aug 2022
Implemented backend **encryption and decryption APIs** with **Java Spring Boot** with multi-factor authentication

Meter Ethereum Hackathon - Scalable decentralized web app for users to create/mint/trade NFTs   Feb 2022 – April 2022
Deployed smart contract to the Meter sidechain with **Solidity** and **Ethers.js** that increased 60x node transactions rate against Ethereum