YiAn Chen

916-597-3444 | yian.chen261@gmail.com | github.com/yianan261 | New York, NY | https://medium.com/@yian.chen261

EDUCATION

University of Pennsylvania *Jan 2025 - May 2026* **Philadelphia, PA**
Master of Science in Engineering, Artificial Intelligence

Stevens Institute of Technology *Sep 2024 - May 2026* **Hoboken, NJ**
Master of Science, Applied Artificial Intelligence (Software Engineering Concentration) | GPA: 4.0/4.0
Courses: Distributed Systems, Applied Modeling and Optimization, Probability and Stochastic Processes, GPU Programming

Northeastern University *Jan 2022 - May 2024* **San Jose, CA**
Master of Science, Computer Science | GPA: 3.96/4.0

WORK EXPERIENCE

Google **Sunnyvale, CA**
Software Engineer Intern (Core Data) May 2025 – Aug 2025

- Deployed end-to-end media download infrastructure with **C++** that resolves cross-environment incompatibilities to enable downloads in dev, directly supporting clients like Google Gemini accelerate development cycles and mitigate security risks
- Built raw byte blob copying pipeline with **parallelized chunking** to optimize memory and latency**,** achieving **throughput increase by 3531.31%** and **latency reduction by 97.25%**
- **Eliminated** file size limitations previously imposed by network bandwidth and service memory, enabling processing of arbitrarily large media files while reducing memory usage by up to **95%** on average large files

Stevens Institute of Technology **Hoboken, NJ**
Lead Research Assistant (ECE Research Scholarship) Jan 2025 – Present

- Training autoregressive transformers model using Pytorch on notation mappings for optical music recognition system

Northeastern University **San Jose, CA**
Lead Research Assistant; ***published*** *in the Springer Communications in Computer and Information Science 2025* Jan 2024 – May 2024

- Developed a novel classification framework leveraging **LLMs** with strategic **prompt engineering methods** to fill out 70% missing fields and **improved classification accuracy by 450%.** Full paper peer-reviewed and published.

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**,** 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 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

TECHNICAL SKILLS

Programming Languages (* most proficient): C++*, Python*, Java, JavaScript, Kotlin, CUDA C, SQL
Databases: MongoDB, PostgreSQL, Redis, Liquibase, MySQL, Amazon Aurora, Firestore
Frameworks/Tools: AWS, GCP, Azure, Docker, React, Spring, Gradle, Bazel, Next.js, Spark, Airflow, Firebase, CUDA, PyTorch
Engineering: Full-Stack Development, Cloud Infrastructure, Distributed Systems, CI/CD, ML Engineering, GPU Programming, HPC

SELECTED PROJECTS

Distributed Database - In-memory relational database with CRUD operations, data redundancy, and durability April 2025 – Present
Built a degree-3 B+- Tree with **LRU buffer pool**, recovery, point/range queries in O(logn) with single-thread throughput of **300k QPS**

Multi-GPU Gradient Aggregation - Optimizing data-parallel ML training with multi-GPU acceleration April 2025 –May 2025
Designed a multi-GPU training pipeline using **CUDA** and **NCCL** with kernel-level optimizations and scaled data-parallel training across **5 Tesla V100 GPUs**. Using **HPC** fused-ops research to reduce communication bottlenecks and get **40% higher throughput**

IT Quackathon (Team Leader and 1st Place Winner) - Competed against 70+ teams April 2025 –April 2025
Built school digital platforms navigation **AI agent** integrating **RAG** and browser automation. Selected to work with the university

Time Series Forecast Project - Predicting energy consumption training ML and LSTM models. Top class project. Nov 2024 –Dec 2024
Built data processing and feature engineering pipeline and trained ML models resulting in **66% accuracy improvement** from baseline

eBay 2024 Machine Learning Competition - Identify motor parts to vehicle compatibility Sep 2024 – Nov2024
Processed data with **NLP** (N-grams, embeddings); **fine-tuned FLAN-T5 and Longformer** models for year, model, make extraction

Google 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**

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