Yuwenqian Chen

yuwenqianchen@gmail.com | 347-495-4890 | linkedin.com/in/yuwenqian-chen | github.com/lenzlaww
Open to relocation | Permanent Resident

Education

Stony Brook University **January 2024 - December 2025**
M.S. in Computer Science | GPA-3.63 *Stony Brook, NY*

Stony Brook University **August 2020 - December 2023**
B.S. in Computer Science and Applied Mathematics & Statistics | GPA-3.53 | Dean's List *Stony Brook, NY*

Experience

AI/Data Engineering Intern **June 2025 -- Present**
DeepChatBI *Remote*

- Engineered near real-time and batch data pipelines connecting multi-source analytics logs, enabling automated metric extraction and improving LLM-driven BI accuracy by 90%.
- Reduced latency by 50% through a streaming API architecture using message-based workflows (Kafka-style event streams) and dynamic tool orchestration.
- Designed metric ingestion and aggregation logic (SQL + Python) supporting observability and data validation across 10+ analytical dashboards, improving reliability and reducing manual QA time by 70%.
- Collaborated on productionizing data services in a cloud environment (AWS/GCP equivalent), implementing CI/CD and automated error monitoring for 'you build it, you run it' ownership.

Software Engineering Intern **February 2024 -- June 2024**
GroupClock Inc. *NY*

- Built secure data capture and transmission pipelines for facial biometrics using SwiftUI + REST APIs, enabling real-time ML input and reducing manual verification by 40%.
- Optimized data streaming and upload logic, improving app responsiveness by 30% and increasing user retention during beta testing (20+ active users).
- Implemented end-to-end monitoring with logging and exception handling to enhance reliability and observability of iOS–backend data sync processes.

Teaching Assistant **Fall 2022, Fall 2024**
Undergraduate & Graduate CS Courses *Stony Brook, NY*

- Facilitated data structure and algorithm learning for 100+ students, improving average exam scores by 12% through structured mentoring and performance feedback.

Projects

GNNs for Molecular Dynamics Simulations **May 2025 - Present**

- Built distributed data preprocessing pipelines for molecular datasets and trained Graph Neural Networks (e.g., DimeNet, MACE) to simulate solvent dynamics.
- Reduced prediction loss by 80% (0.6 → 0.12) through data normalization and model tuning, improving scalability for large-scale simulation workloads.

The Interpretation of Vanity License Plates **September 2024**

- Developed a data processing and training pipeline for 150K+ multilingual text records, fine-tuning a LLaMA3-7B model using LoRA to classify plate approvals.
- Improved classification accuracy to 71%, outperforming lexicon-based baselines by 30% while reducing manual screening workload via automated inference.

Health Monitoring System **September 2023**

- Integrated HealthKit APIs to collect and transmit structured mobility data streams from 30+ post-operative patients, reducing manual data entry by 60%.
- Designed an end-to-end pipeline from mobile sensors to analytics dashboards using Swift, Node.js, and RESTful APIs for real-time clinical insights.

Technical Skills

Programming: Python, Java, C++, Swift, SQL, JavaScript, TypeScript

Frameworks: Apache Spark, Apache Airflow, Kafka, Docker, Kubernetes, LLM (OpenAI, Claude, LLaMA), PyTorch, TensorFlow, Spring Boot, React

Database: BigQuery, MySQL, MongoDB, Snowflake