
EDUCATION

Carnegie Mellon University, Pittsburgh, PA

August 2025 - December 2026

Master of Information Systems Management

Coursework: Object-Oriented Programming in Java, Generative AI, Machine Learning, Unstructured Data Analytics

Anhui University, Anhui, China

September 2021 - June 2025

Bachelor of Engineering in Computer Science and Bachelor of Management

Coursework: Software Engineering, Operating Systems, Artificial Intelligence, Data Structures, Computer Architecture

SKILLS

Programming Languages: Java, Python, SQL, JavaScript, C, CSS, R

Frameworks: Spring Boot, Django, Node.js, React, Next.js, CI/CD

ML/AI Engineering: PyTorch, LangChain, MNE, LDA, Transformer, NLP, MNE

Technologies & Platforms: Kafka, Docker, FastAPI, Git, Airflow, CI/CD

INTERSHIP EXPERIENCE

Beijing Creative & Interactive Digital Technology Co., L, Shanghai, China

July 2024 - October 2024

Software Engineer Intern

- **AI-Enabled Data Platform Architecture:** Designed and implemented resilient batch and near-real-time data pipelines (Python/SQL); built the backend API layer using Django and incorporated a LangChain module to provide semantic data retrieval services, orchestrating workflows with Airflow to cut analyst prep time by 70%
- **Web Refactoring:** Contributed to the Jeep China site redesign, utilizing React components and Next.js architectural principles. Integrated a headless CMS for content decoupling, successfully improving the Lighthouse performance score from 23 to 86

Frost & Sullivan, Beijing, China

January 2024 - March 2024

Software Development Engineer Intern

- **API Service Architecture:** Developed a high-performance REST API layer using Java Spring Boot to serve critical TAM/SAM/SOM metrics. Adhering to a full SDLC (design, unit, and integration testing), this reduced manual reporting preparation time by 50%
- **MLOps and Service Deployment:** Engineered a production-ready Gradient Boosting Model for regional market share forecasting (achieving MAPE of 7.8%). Packaged the model as a modular, containerized Python microservice for reliable deployment and consumption by internal consulting tools

China Mobile Communications Group Co., Ltd., Hangzhou, China

July 2023 - October 2023

Software Engineering Intern

- **Distributed System Architecture and AI-Driven SRE:** Architected a high-reliability telemetry pipeline using C Agents for low-latency logging and Python/Kafka for high-throughput data flow. Deployed an Anomaly Detection Microservice that resolved critical firmware defects, reducing packet loss by 8% and boosting platform throughput by 22%
- **DevOps Excellence and Observability:** Implemented a multi-stage Jenkins CI/CD pipeline, enforcing quality with Sonarqube/Trivy security gating. Deployed the Prometheus/Grafana stack for comprehensive cloud service observability, which successfully reduced deployment-induced rollback incidents by 40%

PROJECT EXPERIENCE

DQN-Enabled Closed-Loop BCI System for Attention Enhancement

September 2023 - June 2025

Tech Stack: Python, PyTorch, P300-BCI, MNE, Reinforcement Learning, LSL

- **Real-Time System Architecture and Performance:** Engineered a closed-loop BCI system with an MNE/LSL pipeline to achieve sub-50ms EEG-to-Stimulus latency. Deployed a DQN microservice that autonomously optimizes visual stimuli, proving system viability for time-sensitive neuro-optimization
- **AI Algorithm Implementation:** Implemented the core Deep Q-Network (DQN) algorithm from scratch in PyTorch, defining the state-space from raw EEG features and the reward function using attention metrics to successfully validate the system's efficacy

High-Throughput ML System for Real-Time TikTok Sentiment Analysis

January 2022 - March 2024

Tech Stack: Python, PyTorch, LDA, Transformer, NLP, Gephi, NetworkX

- **Scalable Data & Analytics Backend:** Architected a backend system to process millions of real-time TikTok comments, featuring a resilient Python ingestion pipeline and a separate NetworkX-based graph analytics API to expose community insights
- **Transformer Serving:** Engineered the production serving layer for a BERT model, deploying it as a containerized REST API (FastAPI/Docker) and implementing batch inference/caching to achieve a 5x increase in QPS and a 60% reduction in P95 latency