

Education

Purdue University, West Lafayette, IN

Expected Dec 2026

B.S. in Computer Science | *Dean's List* | Transferred from Northeastern University (China)

- *Data Structures & Algo (TA), Operating Systems, Database Systems, Computer Networks, Problem Solving & OOP*

Experience

TikTok, San Jose | Incoming Software Engineer Intern

Jan. 2026 – May. 2026

- Search Ads team.

Amazon Web Services (AWS Fintech), Seattle | Software Engineer Intern

Aug. 2025 – Dec. 2025

- Building an end-to-end user features on AWS Lambda and React, providing scalable, low-latency AI-powered experiences – boosting API call throughput by 3.1x while maintaining high reliability.

Bank of China, Shanghai | Software Engineer Intern

Jun. 2023 – Sep. 2023

- Built a real-time transaction analytics platform in Python and React, integrating Socket.IO WebSockets for millisecond anomaly alerts – cut response time from 15 min to 30 s, increasing fraud interceptions by 40%.
- Authored Pandas-based ETL scripts to clean and normalize 50K+ daily logs and scheduled PostgreSQL batch loads via Airflow, tripling throughput and ensuring fault-tolerant retries under load spikes.
- Optimized Flask APIs with partitioned tables, indexing, and caching strategies, accelerating historical report queries by 69% (from 20 min to 6 min), and added React Hooks-driven charts/exports for real-time visualization.

Purdue Aerial Robotic Team (PART) | Technical Member

Jan. 2024 – Present

- Built a full-stack web app for sketch-based drone control serving 1K+ concurrent users, and optimized the React sketch pipeline using Web Workers, cutting CPU usage 62% and sustaining 60 FPS for 8K-point sketches.
- Developed a Python/Flask microservice with a gradient-boosted ranking model to generate drone command (+25% precision); designed REST APIs and implemented telemetry logging for debugging in flight tests.
- Containerized services with Docker, orchestrated them using Kubernetes, and provisioned infrastructure via Terraform – reducing cloud operating costs by 28% while improving deployment reliability and rollback speed.

Purdue University AI Lab | AI Research Assistant

Jan. 2024 – Present

- Built a Python/PyTorch video prediction pipeline integrating physical laws for realistic motion forecasts and ran online A/B tests (t-test, confidence intervals), validating a 12% accuracy lift over pure-ML models.
- Accelerated genetic-algorithm evolution with LambdaMART learning-to-rank, cutting convergence time by 65% and boosting accuracy 18%, and built a scalable Pymunk physics engine, reducing collision latency 70% for 10+ objects.

Projects

Moments | Social Media Platform

Mar. 2024 – Sep. 2024

- Architected a Spring Cloud-based microservice social platform, deploying an ML feed-ranking model to keep <100 ms latency, and implemented WebSocket alerts for instant likes/comments in 200ms, lifting CTR by 25%.

SmartGA | LLM-Guided GA Engine

Jan. 2024 – Mar. 2024

- Built a microservice-based GA engine using LLM guidance for adaptive crossover and mutation, with real-time entropy/fitness monitoring – achieving 25% faster convergence and 15% better solution quality.

EchoGen | LLM Optimization Tool

Jan. 2024 – Mar. 2024

- Built an iterative feedback loop using multi-run sampling, similarity scoring, and averaging, and hint injection to self-correct GPT-4 outputs, improving stability and reliability for downstream simulations.

SketchTune | CLIP Fine-Tuning Project

Dec. 2023 – Jan. 2024

- Fine-tuned OpenAI's CLIP model on a custom sketch dataset to improve recognition and understanding of sketch-based inputs, enhancing classification accuracy on visual data outside the model's original domain.

Skills

Programming & Frameworks: Python, JavaScript/TypeScript (React.js, Node.js, Express.js, Vue.js), C++, Java, Go

Data & Databases: PostgreSQL, MySQL, MongoDB, Pandas, NumPy, Apache Airflow

ML & Cloud Infrastructure: PyTorch, TensorFlow, Scikit-Learn, CUDA, AWS, Docker, Kubernetes, Terraform