



RUILIN JIN

216-704-5196 | ruilin.jin@case.edu |  | 

Education

Case Western Reserve University

Master of Science in Computer Science | Artificial Intelligence Track | GPA 3.8/4.0

Sep. 2023 - May 2025

Cleveland, OH

Rensselaer Polytechnic Institute

Bachelor of Science in Information Technology & Web Science | Minor in Philosophy

Sep. 2016 - May 2021

Troy, NY

Relevant Coursework: Data Structures, Computer Networks, Database Systems, Designing High Performance Systems for AI, Machine Learning, Artificial Intelligence, Deep Generative Models, Causal Inference, Computer Vision

Professional Experience

CREC CNIT | Leading construction company in China

Jun. 2021 - Aug. 2023

Software Engineer | Employee of the Year Award in 2022

Beijing, China

• R&D Platform

- * Led the development of a **micro-frontend** scaffold project using **Vue.js**, creating a component library with 50+ unique components, supporting 350+ developers and 500k+ users, accelerating the development cycle by 40%, resulting in 43 additional projects completed in 2022.
- * Implemented CI/CD pipelines using **Nginx**, **Docker**, **Kubernetes**, and **Jenkins**, streamlining large-scale project management and reducing deployment times by 50%.
- * Established coding and documentation standards to improve code quality and knowledge sharing, enforced **ESLint** and **Prettier**, and reduced error rates by 30% through the use of **SonarLint**.

• Marketing Management System

- * Managed an **Agile** development team of 5 engineers, collaborating with 4 cross-department teams to consolidate 40+ pages across 8 modules within a 3-month deadline, improving system efficiency by 20%.
- * Implemented **virtual scrolling** and **lazy loading**, reducing page load times by 90% for 230k+ lines of data, significantly enhancing user experience.
- * Refactored code by **modularizing** and using shared components, optimizing **Webpack** and file compression strategies to reduce the bundle size from 40.25MB to 15.4MB, and cut deployment time from 10 minutes to 4 minutes.

• AI Middle Platform

- * Developed a name comparison function using **BERT**, achieving 95% accuracy and reducing processing times from 3 hours to 2 minutes, solving critical delays in data availability.
- * Finetuned **ChatGLM-6B model** with **LoRA** using a private dataset, improving QA support and increasing response accuracy by 15%, raising customer satisfaction scores from 75.3% to 84.8% over 300k+ users.
- * Engineered a real-time data streaming system with **Kafka**, directing data storage to **ClickHouse** and **MySQL** based on data classification, reducing data processing time and storage costs by 60% for handling 20+ GB of data daily.
- * Trained a face recognition system for construction sites with **PyTorch**, integrating it with private data, achieving 98% recognition accuracy for monitoring over 400+ construction sites, saving \$3,000 in third-party API costs.

IBM

Sep. 2019 - Dec. 2019

Software Engineer Intern

Troy, NY

- * Designed and developed a Use Case Analysis report and a detailed roadmap for IBM Watson's digital twin technology. Created a prototype using **React.js** and **GoLang** with **Gin**, integrating **RESTful** APIs for seamless data communication.
- * Conducted **competitive analysis** on applications, market viability, and technical aspects using **Pandas** and **Tableau**. Provided strategic insights that led to the adoption of new features, generating \$108,363 in total revenue.

Selected Research Project

Generative AI in Education, RAG Agent

Mar. 2024 - Present

Graduate Research Assistant

- * Developed a **RAG agent** using **Next.js** and **FastAPI**, with **MongoDB**, **DynamoDB**, **AWS S3**, and **Redis**, deployed via **Vercel** and GitHub Actions. Saved 32 work hours daily and handled 400,000+ requests per day.
- * Implemented **JWT** authentication and integrated SSO in authorization middleware. Designed an **RBAC** model to manage endpoint permissions, user permissions, and access control, preventing over 300 unauthorized access attempts and API calls.
- * Designed an algorithm to **embed** course-specific and user-specific material into different namespaces in **Pinecone** after embedding with OpenAI endpoints, ensuring data isolation for **multitenancy** and improving data retrieval time from 45 seconds to 10 seconds, even under peak usage.
- * Enhanced the **RAG** process by merging retrievers in **LangChain**. Enabled **multi-source querying** and provide accurate answers, increasing agent response accuracy by 25% for complex multi-source queries and improving user satisfaction by 20%.

Hello Algo Open Source Project |

Dec. 2023 - Present

Contributor

- * Contributed to writing and translating the English version of an open-source project with 88.7k stars, aimed at helping learners quickly get started with data structures and algorithms.

Technical Skills

Languages: JavaScript, TypeScript, Python, GO, Java, C++, SQL

Database: MySQL, PostgreSQL, Redis, DynamoDB, ClickHouse, MongoDB, Pinecone, ChromaDB

Technologies/Frameworks: Vue.js, React.js, Next.js, Redux, Vuex, Vite, Webpack, Tailwind, FastAPI, Flask, PyTorch, TensorFlow, MFlow, Kafka, Docker, LangChain, Pinecone, Git, AWS S3