

Wenhao He

wenhuohe8@gmail.com | 347-824-4797 | LinkedIn | GitHub | Website

Software Engineer with professional experience building production full-stack and AI-powered systems using Python, TypeScript, React, and cloud platforms.

WORK EXPERIENCE

Software Engineer (AI/Full-Stack)

Clipp

New York, New York

Nov 2024 – Jun 2025

- Collaborated in Agile/Scrum teams using Jira and Git to integrate AI-powered chatbot workflows (Botpress) into production, enhancing user engagement and content relevance.
- Developed scalable full-stack applications (React, TypeScript, GraphQL, REST APIs, Flutter), improving workflows and boosting efficiency by **30%**.
- Built and deployed a real-time chat platform (GetStream, Firebase, TypeScript) supporting **500+ concurrent users** across distributed infrastructure.
- Automated testing and quality checks, achieving **90%+ branch coverage** for critical workflows, ensuring stability and maintainability in production releases.

Software Engineer

CAN International

New York, New York

Apr 2024 – Oct 2024

- Initiated and led development of a membership app with job search and event features, engaging 300+ users.
- Customized Azure Document Intelligence pipelines to extract structured job data, boosting **automation efficiency by 40%**.
- Built Azure Functions triggers for real-time NLP updates, improving responsiveness.
- Reduced avatar load time from **minutes to seconds**, greatly improving UI responsiveness and user experience.

PROJECTS

ResumeMatch (AWS Lambda, DynamoDB, Bedrock, Textract, React)

Serverless AI resume analyzer

Dec 2025 – Present

- Built an event-driven pipeline (S3 → Lambda → Textract → Bedrock Claude Haiku) executing 4-pass LLM analysis: keyword extraction, match scoring, YOE detection, and resume rewriting.
- Served authenticated users via Cognito and API Gateway with a React/TypeScript SPA on CloudFront, persisting history in DynamoDB with GSI-based querying.
- Hardened for production with least-privilege IAM, pre-signed S3 uploads, reserved concurrency, per-user rate limiting, and CloudWatch logging.

Personal Blog Platform (Node.js, Express, MongoDB, React)

Full-stack web app for publishing and managing blogs

Jun 2025 – Present

- Built and deployed a full-stack blogging platform, supporting a public blog and role-based admin CMS.
- Implemented JWT authentication, API rate limiting, and secure Cloudinary-based media uploads.
- Deployed on Vercel with CI/CD pipelines, markdown-to-HTML rendering, SEO-friendly routing, and backend optimizations supporting hundreds of daily visitors.

TECHNICAL SKILLS

Programming: Python, TypeScript, JavaScript, C, C++

AI & Machine Learning: LLM, NLP, Transformer, TensorFlow, PyTorch, Scikit-learn, OpenCV, spaCy

Deployment & Optimization: CI/CD (GitHub Actions, Vercel), API optimization

Cloud & DevOps: AWS (Certified Developer – Associate, DVA-C02), GCP, Firebase, Microsoft Azure, Docker

Backend & Databases: GraphQL, Node.js, Jest, MongoDB, Django, Flask, SQL, MySQL

Frontend & UI: HTML, CSS, React, Next.js, React Native, Flutter

Developer Tools: Git, Jira, Datadog, Figma, VS Code, IntelliJ IDEA, PyCharm

EDUCATION

University at Buffalo, SUNY

Master of Science in Engineering Science (Artificial Intelligence)

Aug 2022 – June 2024

University at Buffalo, SUNY

Bachelor of Science in Computer Science

Aug 2018 – June 2022