Charlie Cheng

6462698719 - charliecheng112@gmail.com - https://www.linkedin.com/in/charlie-tianle-cheng-6147a4325 - github.com/Oak112

SUMMARY

A Software Developer with expertise in **Full-Stack development**, **Agentic AI** and **cloud-native architectures**. Proficient in **Python, C#, Java, TypeScript, AWS**. Interact with my **https://chengai.me** to explore my projects and talk with my AI.

EDUCATION

New York University   **New York, USA**

Master of Science in Urban Data Science (GPA: 3.9/4.0)   *Sep 2024 - May 2026*
• *Core Courses: Big Data, Machine Learning for Cities, Urban Computing and AI*

University of Science and Technology Beijing   **Beijing, China**

Bachelor of Engineering in Artificial Intelligence   *Aug 2019 - Jun 2023*

TECHNICAL SKILLS

Programming Languages: Java, Python, Go, C#, TypeScript, JavaScript, SQL, Bash, C++

Frameworks & DevOps: FastAPI, Next.js, gRPC, Docker, Kubernetes, Helm, Terraform, CI/CD, Git, REST API

AI & ML Tooling: AI Agents, RAG, Semantic Kernel, LangGraph, LangChain, FAISS, BM25, PyTorch, MLflow

Data & Cloud: PostgreSQL, MySQL, Redis, Kafka, Apache Iceberg, AWS (EKS, S3, MSK, EMR), Prometheus

PROFESSIONAL EXPERIENCE

Riot Platforms, Inc.   *C#, .NET, Semantic Kernel, AWS, AI Agent*

Software Engineer Intern, Austin, TX   *Jun 2025 – Aug 2025*

• Designed an **AI agent** in **C# & .NET 8** with **Semantic Kernel**, deployed on **Amazon EC2**.
• Supports natural-language queries on **AWS Timestream** data, with **MCP** integration for Bitcoin mining ops.
• Engineered a **hybrid RAG** search using **BM25 + vectors**, cutting SQL-generation tokens by **95%**.
• Deployed an **Agentic workflow** with **GPT-4o** for planning and **Claude 4 Sonnet** for complex SQL generation.
• Refined the agent via iterative prompt tuning, expanding accurate query coverage by **10x**.
• Delivered **90%** higher query accuracy, **70%** faster response time, and **80%** lower operational cost.

C-STAR   *Python, FastAPI, MLOPs, AWS, MySQL, React*

Software Engineering Intern, New York City, NY   *Oct 2024 – Mar 2025*

• Built a web app with **FastAPI**, **React**, and **MySQL**, reducing manual real estate processes by 95%.
• Engineered data pipeline using **AWS S3**, **EC2**, and **Lambda** for automated batch ingestion.
• Deployed **SageMaker** with **Random Forest Regressor** to predict property prices with 92% accuracy.
• Implemented real-time data sync between **MySQL** and **Google Sheets API** for dynamic property listings.
• Created rolling storage solution to reduce **Google Sheets** storage usage by 90% while preserving data integrity.

LuxPricer: AI-Powered Luxury Bag Valuation App (app.luxpricer.com)   *React, React Native, Python, LangGraph*

Software Engineer   *Apr 2025 – Present*

• Built an **AI appraisal platform for luxury handbags**; users upload photos to get instant valuations & resale insights.
• Engineered **FastAPI** orchestration with **LangGraph** parallel workflows, reducing appraisal latency by 40%.
• Integrated **Gemini 3 Pro** vision models for image-based brand/model identification, achieving 95%+ detection accuracy.
• Built **Node.js + Puppeteer** scraper aggregating real-time pricing from 3 luxury resale platforms with CAPTCHA handling.
• Leveraged **Perplexity API** for market analysis & built **Plotly** charts with outlier removal & similarity scoring.
• Developed **React** web & **React Native** mobile apps with **Supabase** auth & data; deployed via **Docker** & **Coolify**.

PROJECTS

ChengAI: My AI Digital Twin [chengai.me]   *TypeScript, Next.js, React, PostgreSQL, pgvector, OpenAI, Docker*

• Built an **AI portfolio** enabling recruiters to explore candidate background via **RAG chat** & automated JD matching.
• Developed **hybrid retrieval** combining pgvector similarity search & full-text search with **sub-2.5s p95 latency**.
• Implemented **JD parsing engine** that extracts requirements, maps to skill graph, and generates analysis reports.
• Engineered **streaming chat** via SSE with 3 modes: Auto, Tech Deep Dive, and Behavioral (STAR).
• Built **knowledge ingestion pipeline** for PDF/DOCX/TXT with **smart chunking & batch embedding**.
• Designed **Admin CMS** with CRUD, CSRF protection, analytics dashboard; deployed at **chengai.me**.