Edison Xu

Edisonapply@gmail.com | Atlanta | 626-552-1696

www.edisonwhale.com | linkedin.com/in/Edisonwhale | github.com/edisonwhale

Objective: Seeking Software Development Engineer or AI/ML Engineer position as a new graduate/entry level

EDUCATION

Georgia Institute of Technology

Aug 2022 - Dec 2024

Master of Science in Computational Science & Engineering

University of California - Santa Barbara

Oct 2019 - Mar 2021

Bachelor of Art, Applied economics

EXPERIENCE

Highmark Inc. May 2024 - Present

Software Engineer (Generative AI team)

Pittsburgh, PA

- Architected an autonomous AI deep research agent that decomposes queries for parallel search and synthesizes multisource data, using a **Critic model** in a self-correction loop to ensure factually accurate and coherent reports.
- Engineered an enterprise knowledge workspace with Vertex AI Search and Chain-of-RAG, using layout-aware chunking and indexing to retrieve and rerank evidence, producing accurate, source-grounded answers from private files.
- Developed an internal **multi-modal** chatbot using Vertex AI (Gemini API) that processes text, PDF, image, and voice inputs via **TypeScript/React** frontend and **FastAPI** backend, enhancing employee efficiency for routine tasks.
- Built a **Retrieval-Augmented Generation** (RAG) system using Flask, integrating embedding models and Gemini LLM to answer employee queries on internal policies, summarize documents with sources, and cut search time 70%.
- Deployed GitOps CI/CD pipelines with Cloud Run and Kubernetes, optimizing deployment workflows for efficiency.

ChineseInvestors.com Inc. Sep 2022 – Nov 2022

Data Analyst San Gabriel, CA

- Developed a market cycle prediction model using multiple machine learning algorithms to analyze price strength, trend direction, and momentum indicators, increasing investment strategy performance and decision-making accuracy.
- Optimized a **PySpark** and Azure Databricks data pipeline by implementing data partitioning and caching, reducing query execution time and streamlining processing for over 1 million stock records.

PROJECTS

FigBrain: Figma Multi-Agent Plugin

Apr 2025 – May 2025

- Built a Figma AI agent at the Microsoft AI Hackathon, enabling natural-language prompts to generate and populate FigJam components via the **Model Context Protocol** (MCP) and automated Figma API/plugin flows.
- Designed a supervisor **multi-agent** system using **LangGraph**, where a master orchestrator routed tasks to planner, executor, and verifier agents with **Chain-of-Thought** prompts and **ReAct** loops, validated at ~85% first-try accuracy.
- Cut API costs ~28% and lowered latency through **long/short-term memory** mechanism, plan/result caching, and dynamic model switching between lightweight and advanced LLMs.

Golang-based High-Efficiency Search Engine

Aug 2023 - Dec 2023

- Developed a Go-based search engine with text and vector search capabilities, using gRPC for microservices architecture
- Implemented a distributed web crawler with Kafka cluster integration for efficient data collection and processing
- Enhanced search relevance using TF-IDF and BM25 ranking algorithms, with Trie trees for optimized query suggestions

CloudBin: Cloud-Base Distributed File Storage System

Jan 2023 - Mar 2023

- Designed a storage system using Spring Boot, Vue3 and AWS S3, featuring chunk-based file uploads for reliability
- Leveraged **Redis** with a Bloom Filter to prevent cache penetration, boosting cache hit rate and reducing database load
- Integrated ElasticSearch for efficient fuzzy file name searches, reducing search time and improving accuracy

SKILLS

Language	Java, Python, Go, JavaScript, Typescript, C, HTML, CSS, R, SQL
Frameworks	React.js, Vue.js, Spring Boot, Django, Flask, Node.js, Express.js, Selenium
Database/Big Data	MySQL, PostgreSQL, MongoDB, Redis, PySpark, DynamoDB, Firebase
Developer Tools	Docker, Kubernetes, Jira, Git, AWS, GCP, Postman, Nginx, Linux