Edison Xu

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

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

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 Arts, Applied Economics

EXPERIENCE

Highmark Inc.   **Feb 2025 - Present**

Software Engineer (Generative AI team)   Pittsburgh, PA

- Built a real-time AI call-center training platform with **React** and **FastAPI**, leveraging **WebSockets** to stream live voice interactions and auto-score sessions via rubric-based evaluation, significantly shortening agent training cycle.
- Architected a Planner–Executor–Critic **multi-agent** system for AI deep research that parallelizes web search and synthesizes multi-source evidence, delivering auditable, citation-backed reports and cutting turnaround time.
- Engineered an enterprise document Q&A workspace on **Vertex AI Search**, leveraging layout-aware chunking and reranking to generate strictly source-grounded answers from access-controlled internal files, eliminating AI hallucinations.
- Deployed GitOps **CI/CD** pipelines on Cloud Run with logging and alerting, reducing MTTR and incident response.

Highmark Inc.   **May 2024 – Dec 2024**

Software Engineer Intern (Generative AI team)   Pittsburgh, PA

- Developed an internal **multi-modal** AI platform on Vertex AI with **React/TypeScript** frontend and **FastAPI** backend, supporting session chats and real-time streaming for text/PDF/image/voice inputs to streamline employee workflows.
- 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 citations, cutting search time 70%.
- Designed and implemented a centralized AI Proxy Service (**REST API**) to securely integrate Gemini models; added authentication, schema validation, and logging to standardize API usage and prevent API key leakage.
- Enforced code quality with **Pytest** suites achieving 90%+ coverage, improving service stability during feature iterations.

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 reduced latency by implementing long/short-term memory with **PostgreSQL**, 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-Based 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	Python, Go, TypeScript, JavaScript, Java, SQL
Frameworks	React.js, Next.js, Vue.js, Spring Boot, Django, 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