

Kaviyashre R P

+91-9731863181 kaviyashreragupathy@gmail.com linkedin.com/in/kaviyashre-ragupathy-2103 github.com/RPK2103

Professional Summary

Python AI Engineer with 2.5+ years of experience building LLM-powered applications, RAG pipelines, and distributed backend systems processing 500K+ portfolios. Strong in FastAPI, microservices, and enterprise AI workflows using Azure OpenAI and vector databases.

Technical Skills

Languages: Python, Java, SQL

AI/ML: LLM, RAG, LangChain, Prompt Engineering, Agentic Workflows

Backend: FastAPI, REST APIs, Microservices, Async APIs

Databases: PostgreSQL, MongoDB, Vector DB (FAISS)

Cloud: Azure OpenAI, AWS

Concepts: System Design, Scalability, High-throughput Systems, Caching

Experience

Cloud and AI Engineer

Rapid Circle

March 2026 – Present

- Developed LLM-powered consultant discovery platform using **RAG pipelines and vector embeddings**, improving resource matching efficiency by **40%**.
- Built **FastAPI-based microservices** handling async API calls for skill retrieval across **500+ consultant profiles**.
- Developed Java and FastAPI-based microservices supporting AI-powered consultant discovery workflows, improving backend response efficiency by **35%**.
- Integrated Java and Python backend services with Azure OpenAI APIs enabling intelligent search and recommendation workflows.
- Designed agentic workflows for skill inference and recommendation, reducing manual effort by **50%**.
- Implemented caching strategies improving response time under concurrent requests.

Software Engineer

JP Morgan Chase

June 2023 – Feb 2026

- Built distributed financial computation system processing **500K+ portfolios and millions of daily records**.
- Optimized data pipelines reducing batch processing time by **30–40%**.
- Developed REST APIs supporting high-throughput analytics with **sub-second latency**.
- Applied data modeling techniques on time-series financial data improving reporting accuracy.
- Engineered filtering and transformation logic analogous to feature selection improving computation efficiency.
- Improved financial reporting accuracy by enabling daily transaction-level computation.
- Refactored computation modules reducing latency by **35%**.
- Improved CI/CD pipelines reducing deployment time by **25%**.
- Collaborated with 5+ stakeholders to resolve data inconsistencies impacting analytics outputs.

Software Engineer Intern

JP Morgan Chase

Feb 2023 – May 2023

- Developed REST APIs for benchmark data retrieval improving data access efficiency by **20%**.
- Built backend services supporting financial return computation pipelines.

Projects

Enterprise RAG Knowledge Assistant

Python, LangChain, FAISS

- Built end-to-end **RAG pipeline** improving retrieval relevance and reducing latency by **25%**.
- Implemented document ingestion and embedding pipeline across **1000+ documents**.
- Deployed FastAPI APIs enabling real-time LLM query responses.

AI Resume Matcher

Python, LLM, Embeddings

- Developed semantic matching system improving ranking accuracy using embedding similarity scoring.
- Implemented scoring pipeline for job-resume alignment.

AI Interview Coach

Python, LLM

- Built LLM assistant generating interview questions and structured feedback using prompt engineering.

Achievements and Certifications

SEP Engineering Committee Lead — Selected among 700+ engineers | Code for Good Hackathon — 2nd Runner-Up | AWS Certified Cloud Practitioner | Microsoft Certified: Azure Fundamentals | Oracle Generative AI Professional

Education

SRM Institute of Science and Technology

June 2019 – May 2023

B.Tech in Computer Science

CGPA: 9.2