FULL-STACK AI ENGINEER

Candidate Reference Number: 202608081823

PROFILE

AI/Python Engineer and Full-Stack Developer with approximately **5 years of production software experience** across Python, .NET, JavaScript/TypeScript, and cloud-native stacks. Strong focus on **LLM application development, RAG pipelines, agent orchestration, evaluation harnesses, model serving, and production AI operations**. Experienced taking AI features from notebook-level experimentation through deployed, monitored services with guardrails, tests, retrieval grounding, and measurable evaluation.

CORE CAPABILITIES

AI / LLM: RAG | LangChain | LangGraph | LlamaIndex | embeddings | prompt/context engineering | function calling | structured outputs | guardrails | OpenAI | Anthropic Claude | Gemini | Groq | Hugging Face
Agentic Systems: multi-agent orchestration | ReAct | planner-executor patterns | tool routing | memory | MCP | human-in-the-loop workflows
LLMOps / Evaluation: RAGAS | LangSmith | Langfuse | golden datasets | regression testing | hallucination/drift monitoring | MLflow | Weights & Biases | A/B testing
Model Serving / Optimization: Ollama | vLLM | llama.cpp | ONNX | TorchServe | quantization | LoRA/QLoRA | PEFT | batching | KV-cache tuning
Application Stack: Python | FastAPI | Flask | async Python | C#/.NET Core | React | TypeScript | Next.js | Node.js | Spring Boot | REST | gRPC | WebSockets | OAuth2/JWT
Data / Cloud: pgvector | Pinecone | ChromaDB | FAISS | PostgreSQL | SQL Server | MySQL | MongoDB | Redis | Azure | AWS | Docker | Kubernetes | Helm | Terraform

SELECTED DELIVERY EVIDENCE

- Builds **agentic AI workflows in Python** with tool use, structured outputs, retrieval grounding, guardrails, and evaluation harnesses.
- Owns features end to end across **data models, retrieval layers, API design, backend services, and React/TypeScript interfaces**.
- Designs RAG systems using **chunking, hybrid search, reranking, embeddings, and vector stores** with evaluation and regression testing.
- Implements agent workflows using **LangGraph, ReAct, planner-executor patterns, tool routing, memory, MCP, and human-in-the-loop controls**.
- Has hands-on model-serving and optimization experience with **Ollama, vLLM, llama.cpp, quantization, LoRA/QLoRA, and PEFT**.
- Builds accessible, responsive web front ends with semantic HTML, React/Next.js, TypeScript, modern CSS, WCAG-aware patterns, and end-to-end testing.
- Designs documented REST integrations and works directly with product managers and business stakeholders to turn ambiguous requests into scoped delivery.

ENGAGEMENT FIT

Strong fit for Full-Stack AI engagements requiring a practitioner across **RAG, agentic workflows, LLM evaluation, model serving, secure APIs, responsive frontend delivery, vector/data systems, and cloud deployment**. Particularly useful where the client needs one engineer who can move from AI architecture and retrieval design through implementation, testing, monitoring, and user-facing delivery.

EXPERIENCE EVIDENCE

Candidate Reference Number: 202608081823

2026-PRESENT | SOFTWARE ENGINEERING / AGENTIC AI

- Builds agentic AI workflows in Python using tools, structured outputs, retrieval grounding, guardrails, and evaluation harnesses.
- Owns features from schema and retrieval design through API/backend services and React front ends.
- Designs versioned REST APIs and third-party integrations, contributes architecture decisions, writes design documentation, reviews pull requests, and resolves production defects.
- Develops responsive, accessible React/TypeScript interfaces and maintains unit, integration, and end-to-end test coverage.

2025-2026 | AI & AUTOMATION ENGINEERING

- Built Python/Flask services and REST APIs powering internal tools, automation workflows, reporting dashboards, and scheduled/webhook-based processes.
- Used SQL and Pandas for data cleaning, joins, aggregation, and API-serving workflows.
- Built React/TypeScript dashboards with reusable components and performance optimization through code splitting, lazy loading, and render tuning.
- Containerized services with Docker, deployed to Azure Kubernetes Service, and maintained GitHub Actions pipelines for tests, linting, builds, and deployments.

2023-2024 | FULL-STACK DEVELOPMENT / RESEARCH

- Built a real-time tracking and attendance application using Python/Flask, PostgreSQL, React, REST APIs, and WebSockets for hundreds of daily users.
- Implemented OAuth2/JWT with role-based access control and improved slow-query performance through indexing, query rewrites, and schema changes.
- Deployed Dockerized services to AWS EC2 and Azure App Services through GitHub Actions.

2020-2022 | SOFTWARE ENGINEERING

- Developed enterprise applications in C#/.NET Core, ASP.NET Core Web API, MVC, Java/Spring Boot, Python, React, PostgreSQL, MySQL, and SQL Server.
- Built Python automation scripts, REST/JSON integrations, Redis caching, responsive React interfaces, and Jenkins/Docker CI/CD.
- Tuned T-SQL queries, stored procedures, and indexes and worked in Agile delivery with project managers and QA.

SELECTED PROJECTS

- **AI Assistant Platform:** Python, FastAPI, LangGraph, OpenAI/Anthropic APIs, pgvector, React/TypeScript, Docker.
- **Agentic RAG Assistant:** Python, LangGraph, FastAPI, pgvector, Ollama/vLLM, RAGAS evaluation harness.
- **Ticketing / Workflow API:** .NET 8, ASP.NET Core Web API, EF Core, SQL Server, Docker, Azure, React/TypeScript.

EDUCATION

M.S. in Computer Science
B.Tech in Computer Science