

FULL-STACK AI ENGINEER

Candidate Reference Number: 202608071556

PROFILE

Full-Stack AI Engineer and senior Python developer with **10+ years** building scalable backend, full-stack, data, and AI systems. Strong capability across **Python/FastAPI, React/TypeScript, PostgreSQL, Generative AI, RAG, vector databases, document intelligence, cloud infrastructure, APIs, microservices, testing, and observability.**

CORE CAPABILITIES

AI / LLM: OpenAI | Azure OpenAI | RAG | LangChain | LangGraph | embeddings | prompt engineering | agents | XAI | OCR | document intelligence

Data: pgVector | Pinecone | Weaviate | Chroma | FAISS | PostgreSQL | SQLAlchemy | MongoDB | Cosmos DB | Redis | PostGIS | semantic/hybrid search

Application: Python 3.10+ | FastAPI | async | REST/OpenAPI | Pydantic | JWT/OAuth2 | Flask | GraphQL | React | TypeScript | JavaScript

Cloud / Delivery: AWS EC2/RDS/S3/SQS/Batch/Lambda/ECS | Azure | Docker | Terraform | GitHub Actions | CI/CD | Pytest | CloudWatch | Grafana | Prometheus

SELECTED DELIVERY EVIDENCE

- Built RAG pipelines with LangChain/LangGraph, OpenAI embeddings, pgVector and PostgreSQL, reporting a **28% improvement in retrieval precision.**
- Optimized PostgreSQL/pgVector indexing for **semantic-search latency under 150 ms.**
- Created prompt-evaluation harnesses that reportedly **reduced hallucination rates by 22%.**
- Migrated Django REST services to FastAPI microservices, reportedly **cutting p95 API latency by 35%.**
- Architected a FastAPI/PostGIS/React location-intelligence platform that reportedly **reduced site-selection time by 70%.**
- Designed a document-analysis platform processing **500,000+ analyses daily at 94% accuracy.**
- Delivered cloud optimizations reporting **30-40% infrastructure-cost reductions** in separate engagements.

ENGAGEMENT FIT

Strong fit for work spanning the **AI-enabled application stack:** Python APIs, React interfaces, relational/vector data, RAG and agentic workflows, document intelligence, cloud deployment, testing, security, batch orchestration, and production observability.

EXPERIENCE EVIDENCE

Candidate Reference Number: 202608071556

2022-2026 | FULL-STACK AI ENGINEERING / GENAI / MLOPS

- Designed production GenAI, LLM, RAG, OCR, semantic-search and document-intelligence workflows using FastAPI, OpenAI/Azure OpenAI, LangChain/LangGraph, PostgreSQL and vector stores.
- Built async FastAPI microservices and workers using AWS SQS and Azure Service Bus for document processing and batch orchestration.
- Developed React/TypeScript audit UIs backed by FastAPI, PostgreSQL/SQLAlchemy and AI-generated document summaries.
- Implemented hybrid retrieval across Pinecone, Weaviate and pgVector; surfaced citation trails for audit/compliance review.
- Deployed Dockerized AI services through GitHub Actions with Pytest gates, security scans, monitoring and logging across AWS and Azure.
- Mentored engineers on FastAPI, LangChain/LangGraph, prompt engineering, vector databases and CI/CD quality gates.

2018-2022 | FULL-STACK ENGINEERING / AI DEVELOPMENT

- Delivered cloud-native Python/FastAPI, PostgreSQL, MongoDB, AWS, Docker and React/TypeScript platforms with OCR and document-processing workflows.
- Created AI-powered PDF/report generation producing **5,000+ reports monthly**.
- Built LangChain/OpenAI agent workflows with PostgreSQL state and Redis caching, reporting **85% accuracy on complex business queries**.
- Implemented RAG-style retrieval and semantic search using embeddings, Pinecone and OpenAI.
- Migrated monoliths to Dockerized AWS ECS microservices with SQS/Batch orchestration.
- Applied LLM summarization over operational runbooks, reportedly **improving support response time by 40%**.

2016-2018 | FULL-STACK / AI ENGINEERING

- Architected high-volume document-analysis systems using Python REST services, PostgreSQL/SQLAlchemy, OCR, semantic search, MongoDB, AWS and Docker.
- Built authorship-determination services processing hundreds of thousands of documents daily.
- Developed document-intelligence workflows with asynchronous workers, Airflow orchestration, OCR extraction and PostgreSQL persistence.
- Designed multi-tenant SaaS patterns with Dockerized Python services, PostgreSQL row-level security, S3 and MongoDB.
- Implemented CI/CD, automated testing, monitoring, caching and secure OAuth2/API-key access.

EDUCATION

Bachelor's Degree - Computer Science