SARAH CARTER

Generative AI Engineer

sarah.carter@email.com (123) 456-7890 Boston, MA

LinkedIn

WORK EXPERIENCE

Generative AI Engineer

State Street

January 2024 - current Boston, MA

- Designed and deployed enterprise RAG pipelines using Python, LangChain, and Pinecone, **processing 1.2M+ internal document records** for knowledge retrieval applications.
- Implemented hybrid search and metadata filtering strategies, improving document retrieval precision by 29% across internal evaluation datasets.
- Developed multi-agent workflows using LangGraph and LangChain to automate research and document analysis tasks, expediting manual processing time by 7 minutes.
- Fine-tuned open-source large language models using LoRA and QLoRA techniques, reducing GPU memory requirements by 41% during model adaptation experiments.

Machine Learning Engineer

Repligen Corporation

June 2021 - December 2023 Waltham, MA

- Built semantic search pipelines with vector embeddings and Qdrant, indexing 314,822 technical and operational records for faster information search.
- Implemented document preprocessing, chunking, and embedding generation workflows, **shortening indexing time by 16 seconds** across knowledge bases.
- Developed REST APIs to integrate machine learning models with existing business applications, improving model accessibility across three internal teams.
- Optimized model inference workflows through batching and quantization experiments, reducing average response latency by 28%.

Software Engineer

HubSpot

July 2018 - July 2021 Cambridge, MA

- Developed Python-based backend services and REST APIs supporting customer-facing software applications.
- Built automated data processing workflows to transform and validate 122,528 records monthly, lowering manual data preparation by 61%.
- Optimized SQL queries and database operations, improving average API response times by 14 seconds.
- Implemented unit and integration testing workflows to identify application defects and improve release reliability.

EDUCATION

B.S.

Computer Science

Boston University

October 2014 - May 2018

Boston, MA

SKILLS

- LangChain & LlamaIndex
- Pinecone & Qdrant
- Retrieval-Augmented Generation (RAG)
- Python (PyTorch)
- LoRA & QLoRA Fine-Tuning
- LangGraph
- Hugging Face Transformers
- Vector Embeddings & Semantic Search
- Agentic AI Workflows

CERTIFICATIONS

- Databricks Certified Generative AI Engineer Associate | Databricks — 2026