# Yash Desai

Generated by JobOs AI • Gen AI Engineer • Portfolio Link

## Professional Summary

Highly motivated and detail-oriented Generative AI Engineer with a strong foundation in AI principles and a keen interest in developing end-to-end AI-powered applications. Proficient in designing and implementing robust AI architectures, with a focus on scalability, security, and cost efficiency. Seeking to leverage my skills and knowledge in bridging the gap between cutting-edge AI research and production-grade software.

## Technical Skills

Python • TypeScript • Go • PyTorch • TensorFlow • LangChain • AutoGen • LangGraph • SQL/NoSQL • Vector DBs • AWS (SageMaker)

## Selected Projects

### Personal Project - Retrieval-Augmented Generation (RAG) Pipeline

- Designed and implemented a basic RAG pipeline to explore the capabilities of retrieval-augmented generation.

- Utilized vector databases for efficient information retrieval and integrated with a simple language model for text generation.

- Explored techniques for fine-tuning the model to improve domain-specific accuracy.

### Theoretical Study on LLM Optimization

- Conducted a theoretical study on optimizing large language models (LLMs) using techniques such as LoRA and QLoRA.

- Explored the impact of these optimization techniques on model performance and efficiency.

- Documented findings in a personal blog post to share knowledge with the community.

### CI/CD Pipeline for AI Model Deployment

- Developed a simple CI/CD pipeline for deploying AI models, focusing on automated evaluation and cost tracking.

- Utilized containerization and orchestration tools to ensure scalable and reliable model deployment.

- Implemented basic monitoring for model performance and latency.

## Certifications

- **AWS** - AWS (2025-02-23 00:00:00.000Z)