

ATS OPTIMIZED CV - prompt_engineer

PROFESSIONAL SUMMARY

Prompt Engineer with 6+ years software engineering experience specializing in LLM application development, RAG systems, and AI integration. Expert in OpenAI API, GPT-4, Claude, prompt optimization, vector databases, and production AI system architecture. Proven track record building scalable APIs serving 200K+ users, with demonstrated experience in Python, FastAPI, system design, and cloud deployment.

SKILLS

****LLM & Prompt Engineering:**** OpenAI GPT-4, Claude API, Prompt Optimization, RAG Systems, Vector Databases, LangChain, Semantic Search, Few-shot Learning, Chain-of-Thought Prompting
****Programming Languages:**** Python, Java, JavaScript, SQL ****AI/ML Frameworks:**** LangChain, OpenAI API, Anthropic Claude, Hugging Face, Vector Embeddings ****Backend Technologies:**** FastAPI, Django, Spring Boot, Node.js, REST API, Microservices ****Databases & Storage:**** PostgreSQL, MongoDB, Redis, Vector Databases (Pinecone, Chroma), Elasticsearch ****Cloud & DevOps:**** AWS, Docker, CI/CD, API Gateway, Lambda Functions ****Integration:**** API Integration, Webhook Implementation, Third-party Services, Authentication (OAuth, JWT)

EXPERIENCE

Senior Software Engineer | Orion Digital Systems | 2021 – Present

- Developed production LLM applications using OpenAI GPT-4 API, implementing RAG architecture with vector embeddings and semantic search. Reduced documentation generation time by 85% through optimized prompt engineering and API integration.
- Built intelligent chatbot using GPT-4, LangChain, and custom prompt templates. Implemented retrieval-augmented generation (RAG) system with PostgreSQL vector storage. Achieved 60% automation rate for customer support queries, reducing response time by 70%.
- Architected scalable backend systems using Python FastAPI, Java Spring Boot, and microservices. Designed REST APIs serving 200K+ concurrent users with 99.9% uptime. Optimized system performance by 40% through caching strategies and database optimization.
- Implemented CI/CD pipelines using Docker and AWS for LLM application deployment. Integrated third-party APIs including OpenAI, Anthropic Claude, and vector databases (Pinecone, Chroma).
- Led prompt optimization initiatives resulting in 35% improvement in LLM response accuracy and 50% reduction in token usage through few-shot learning and chain-of-thought prompting techniques.

Software Engineer | Bluewave Technologies | 2018 – 2021

- Developed full-stack applications using JavaScript (React.js, Node.js), Python Django, and REST API integration. Built microservices architecture handling 50K+ daily API requests.
- Implemented automated testing frameworks with 90% code coverage. Integrated 15+ third-party APIs including payment gateways, authentication services, and data providers.
- Optimized database performance using PostgreSQL, MongoDB, and Redis caching. Reduced query response time by 60% through indexing strategies and connection pooling.

- Built real-time data processing pipelines using Python and message queues. Implemented error handling and monitoring systems for production reliability.

EDUCATION

B.Tech in Computer Science – Jawaharlal Nehru Technological University (2014 – 2018)

Certifications: AWS Certified Developer – Associate, Oracle Certified Java Programmer **AI/ML Learning:** Prompt Engineering Specialization, LangChain Development, Vector Database Implementation

PROJECTS

Production RAG System (2024): Architected retrieval-augmented generation system using OpenAI embeddings, Pinecone vector database, and custom prompt templates. Implemented semantic search with 92% accuracy for document retrieval and question answering.

LLM API Documentation Generator (2024): Built automated documentation system using GPT-4 API, OpenAPI specification parsing, and optimized prompts. Reduced manual documentation time by 80% while maintaining consistent quality and formatting.