

# Prompt Engineer / AI Integrations

AI / ML

Startup

SMB

[Company Name] is hiring a Prompt Engineer / AI Integrations specialist to design, build, and optimize AI-powered features using large language models. You will craft prompts, build retrieval-augmented pipelines, and integrate LLM capabilities into our product. This role is ideal for someone who understands both the capabilities and limitations of modern AI models and loves turning them into reliable product features.

## Key Responsibilities

- Design, test, and iterate on prompts and prompt chains for production LLM features
- Build and maintain retrieval-augmented generation (RAG) pipelines and integrations
- Develop evaluation frameworks to measure prompt quality, accuracy, and safety
- Integrate LLM APIs (OpenAI, Anthropic, open-source models) into product workflows
- Collaborate with product managers to define AI feature requirements and success criteria
- Monitor and improve AI feature performance, latency, and cost efficiency
- Stay current with rapidly evolving LLM capabilities, tools, and best practices

## Required Skills & Experience

- 2+ years of experience building software, with at least 1 year working with LLMs or NLP
- Proficiency with Python and experience calling LLM APIs programmatically
- Hands-on experience with prompt engineering techniques (chain-of-thought, few-shot, system prompts)
- Familiarity with RAG architectures including vector databases and embedding models
- Understanding of LLM limitations (hallucinations, context windows, token costs)
- Experience with evaluation and testing of AI outputs (both automated and human-in-the-loop)
- Strong written communication skills for crafting clear, effective prompts

## Nice-to-Have

- Experience with LLM orchestration frameworks (LangChain, LlamaIndex, or Semantic Kernel)
- Familiarity with fine-tuning or distillation of language models
- Background in NLP, information retrieval, or computational linguistics
- Experience with AI safety, content moderation, or responsible AI practices
- Knowledge of cost optimization strategies for LLM-based applications

## Tech Stack

Python

OpenAI API / Anthropic API

LangChain / LlamaIndex

Pinecone / Weaviate / pgvector

FastAPI

Docker

Git

Weights & Biases

Streamlit

## What We Offer

- Competitive salary and equity package
- Flexible remote or hybrid work arrangement
- Health, dental, and vision insurance
- Annual learning and development budget
- Generous PTO policy

## Interview Process

---

1. Recruiter phone screen (30 min)
2. Technical screen: prompt engineering concepts and LLM fundamentals (45 min)
3. Take-home exercise: design a prompt system for a given use case (2-3 hours)
4. Live walkthrough and iteration on take-home solution (60 min)
5. Product thinking and culture fit interview (30 min)