Description

We're looking for a **real software engineer** who knows how to **build scalable, production-grade apps** that use **AI models** like GPT-4, Claude, or open-source LLMs. You should be able to go from business problem → code → AI-powered solution, not just play with prompt templates.

What You'll Be Doing:

- Building internal tools, APIs, and full-stack apps where LLMs are a **core part of the logic**
- Integrating models like GPT-4 into real workflows (classification, summarization, question answering, RAG, etc.)
- Working with embeddings, vector search, and tools like Pinecone, Weaviate, or FAISS
- Designing services that are modular, debuggable, and scalable
- Collaborating with automation engineers and operators to productize ideas quickly
- Owning the lifecycle of AI-powered features — from design to deployment to monitoring

You Should Apply If You:

- Have **strong backend skills** (Python, Node, Go, PHP.) and can write clean, production-grade code
- Have shipped something real using **OpenAI, Anthropic**, or similar APIs
- Understand **prompt engineering** beyond surface-level hacks
- Know your way around **APIs, databases, cloud infra**, and **testing frameworks**
- Are curious, fast, and biased toward building over theorizing
- Can explain the difference between a cool demo and a durable system

Bonus Points For:

- Experience with full-stack frameworks (React, Svelte, etc.)
- Experience with fine-tuning or using open-source models (e.g. Mistral, LLaMA, Zephyr)
- Familiarity with n8n, LangChain, or retrieval-augmented generation pipelines

Compensation & Benefits:

- **$140,000 – $220,000 base salary (USD)**
- **Bonus + Competitive Benefits**
- Remote-first culture with async-friendly workflows
- Direct access to leadership & fast decision cycles

- Freedom to ship, test, and own high-impact projects

Next Steps:

Send us links to things you've built (GitHub, blog posts, demos, etc.). Tell us why this excites you. Show us code you're proud of.