

## **Job Title: Software Engineer – AI**

### **Job Summary:**

- We're looking for a AI Software Engineer who thrives at the intersection of cutting-edge AI and robust backend systems. This role is perfect for someone with strong Python development experience and a passion for building scalable GenAI applications. You'll lead the design, development, and deployment of backend services that power LLM-based systems and AI-first user experiences. While your primary focus will be backend and AI infrastructure, familiarity with frontend and cloud technologies is a plus.

### **What You'll Do**

#### **Generative AI & Backend Engineering:**

- Design, implement, and deploy GenAI pipelines (LLM orchestration, RAG, multi-agent workflows) using frameworks like LangChain, Agno, LlamaIndex, ReAct, AutoGen, or CrewAI.
- Build and maintain robust Python backend services using FastAPI or Flask to serve AI-powered features.
- Develop APIs and microservices to expose LLM capabilities to applications and users.
- Integrate external data sources (text, structured, unstructured) into AI workflows and retrieval pipelines.
- Handle model selection, fine-tuning, performance evaluation, and optimization for production use.
- Build and manage model versioning, monitoring, retraining, and evaluation pipelines as part of the ML Ops lifecycle.
- Collaborate closely with product and design teams to translate user needs into backend AI features.

#### **Nice to Have – Frontend & Cloud:**

- Interest or experience in building intuitive interfaces using React, Next.js, or Vue.
- Exposure to deploying applications in cloud environments (AWS/DigitalOcean/Azure) using Docker, Kubernetes, and CI/CD pipelines.
- Familiarity with authentication systems, API gateways, and basic DevOps workflows.
- Exposure to Python libraries for ML/AI: Transformers, scikit-learn, PyTorch, TensorFlow, etc.

#### **You Should Have:**

- 0–4 years of professional experience in Python backend development and machine learning.
- Experience building and deploying GenAI applications using LLMs and related orchestration frameworks.
- Strong understanding of RESTful API design and backend architecture.
- Familiarity with vector databases (e.g., FAISS, Pinecone, Weaviate) and semantic search.
- Experience with FastAPI or Flask for production-grade services.
- Good understanding of software engineering principles, testing, version control, and deployment practices.
- A builder's mindset – someone who thrives in fast-paced, ambiguous environments and takes initiative.

#### **What We Value:**

- **Ownership:** You take full responsibility for your work and outcomes.
- **Pragmatism:** You balance innovation with reliability and scale.
- **Curiosity:** You're always learning and experimenting with new AI tools and trends.
- **Collaboration:** You work well across disciplines and communicate ideas clearly.

#### **Employment Type:**

Contract-to-Hire — initial contract with strong potential for full-time conversion based on performance and fit.

### **CORPORATE OFFICE:**

No.390, 1st Floor,7th Main, 9th Cross Rd, 2nd Block, Jayanagar, Bengaluru, Karnataka 560011

[www.vedtechbio.com](http://www.vedtechbio.com)