



# AI/ML ENGINEER

## Job Description

We are seeking AI/ML engineers to develop, deploy, and optimize advanced AI models. You'll work with cutting-edge tools like OpenAI, API's, Hugging Face, Transformers, ONNX, and LangChain to build scalable AI solution

## EXPERTISE

- Large Language Model
- Transformer
- Open AI and Gen AI
- Hugging face
- Langchain & Llama
- ML Libraries (Pytorch, Tensorflow, etc.)
- Prompt Engineering
- Vector Database
- NLP
- ONNX (Open Neural Network Exchange)
- Client Interaction

- Expertise in Large Language Models (LLMs) and experience with OpenAI API, Hugging Face, Transformers, and other NLP frameworks.
- Proficiency in fine-tuning and deploying pre-trained models for various NLP tasks.
- Strong knowledge of LangChain for building and managing LLM pipelines.
- Experience with Vector Databases (e.g., Pinecone, Milvus, Weaviate) for efficient data retrieval and AI- driven search.
- Proficiency in embedding techniques for search, recommendations, and semantic similarity tasks.
- Solid understanding of Prompt Engineering to optimize model performance.
- Strong Python programming skills and experience with ML libraries (TensorFlow, PyTorch, Scikit-learn)
- Experience deploying ML models using ONNX for scalable and efficient inference.
- Hands-on experience in developing, optimizing and maintaining AI/ML models in production environments.
- Ability to collaborate with Cross- functional teams staying updated on AI advancements to enhance solutions.

## Details:

- CTC: Best in the industry
- Functional Area: Development
- Industry: IT/Software Services
- Office Address: 247/32, Block D, Sector- 63, Noida