

About the job

Company Name: RACO

Location: Dhaka

Job Type: Full-Time

Workplace Type: On-Site

Job ID: IJOB202500533

Context

As an AI Engineer at Raco AI, you will design, build, and optimize state-of-the-art AI systems, focusing on Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), Vision-Language Models (VLM), Speech-to-Text (STT), Text-to-Speech (TTS), and Multimodal Analysis. You will create, fine-tune, adapt, and train models for production-level applications, while collaborating with product and backend teams to deploy AI-driven solutions. Your work will directly impact how RacoAI delivers cutting-edge AI capabilities to our customers.

Responsibilities

- Design, fine-tune, and train LLMs, RAG, VLM, STT, TTS, and multimodal AI models.
- Implement reinforcement learning pipelines and RLHF (Reinforcement Learning with Human Feedback) for model optimization.
- Develop scalable AI APIs and services using FastAPI.
- Design and maintain database schemas and model metadata in PostgreSQL and MongoDB.
- Evaluate model performance, implement improvements, and ensure reliability and robustness.
- Collaborate with backend and product teams to integrate AI models into production systems.
- Research and prototype new AI architectures, frameworks, and training techniques.
- Optimize training pipelines for performance, cost, and scalability.
- Maintain clear documentation of AI models, experiments, and training workflows.

About The Job

Last Date of Application: 31 January 2026

Desired Skills and Experience

Object-Oriented Programming (OOP), Python Programming