

Alibaba Cloud Certified Professional: LLM Engineer

OEM: Alibaba Cloud • Code: LMP-C01

COURSE DESCRIPTION

An LLM Engineer designs, fine-tunes, and deploys large language models (LLMs) on Alibaba Cloud to solve complex business problems. This role requires expertise in techniques like RAG, LoRA, and prompt engineering, alongside cloud-native integration for scalable, secure, and cost-efficient solutions.

AUDIENCE PROFILE

- This section doesn't contain any content.

PREREQUISITES

- This section doesn't contain any content.

COURSE OBJECTIVES

By the end of this course, participants will be able to:

- LLM Fundamentals: Understand transformer architecture, attention mechanisms, and parameter tuning (temperature, top_p, seed, etc.).
- Retrieval-Augmented Generation (RAG): Build pipelines for dynamic knowledge integration.
- Fine-Tuning Techniques: Apply LoRA, adapter layers, and full parameter tuning for domain-specific adaptation.
- Model Evaluation: Learn quantitative (BLEU, ROUGE) and qualitative methods for performance validation.
- Cloud Deployment: Deploy LLMs via APIs using Alibaba Cloud services (PAI, ECS, NAS, OSS).
- Ethics & Compliance: Address bias, data privacy, and regulatory requirements in LLM workflows.

COURSE OUTLINE

Module 1: LLM Fundamentals & Alibaba Cloud Integration

- Introduction to LLMs
- LLM Parameters and Behavior

Module 2: Building Retrieval-Augmented Generation (RAG) Pipelines

- RAG Architecture
- Implementing RAG on Alibaba Cloud

Module 3: LLM Fine-Tuning with LoRA

- LoRA Theory
- Training LoRA Models on Alibaba Cloud

Module 4: Model Evaluation & Optimization

- Evaluation Metrics & Methods
- Optimization Techniques

Module 5: Deploying LLMs on Alibaba Cloud

- API Deployment
- Scalability and Monitoring

Module 6: Ethics, Compliance, and Advanced Topics

- Ethical AI Practices
- Cloud-Native Security
- Case Studies