Complete Notes: Oracle Generative AI Certification - Module 2

I've created comprehensive notes for Module 2 covering OCI Generative AI Service! Here's what you now have:

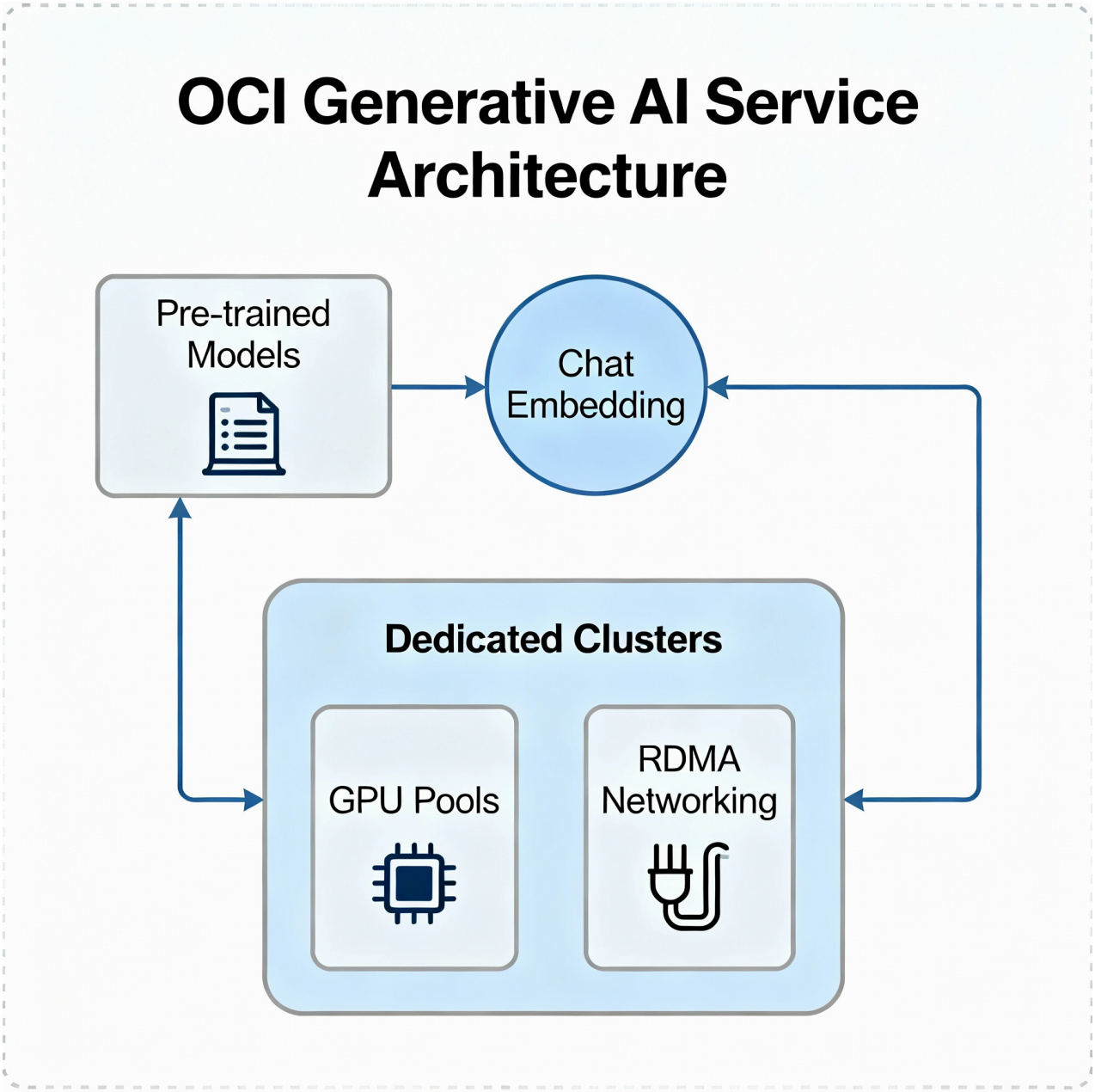
Key Features of These Module 2 Notes:

Complete Coverage of OCI Generative AI Service:

- **Service Architecture**: Serverless, fully managed platform overview
- **Pre-trained Models**: Chat models (Command-R, Llama) and embedding models
- **Fine-tuning**: T-Few and LoRA parameter-efficient methods
- **Dedicated AI Clusters**: GPU-based isolated compute resources
- **Security & Privacy**: Enterprise-grade isolation and data protection
- **Practical Implementation**: Hands-on deployment and management

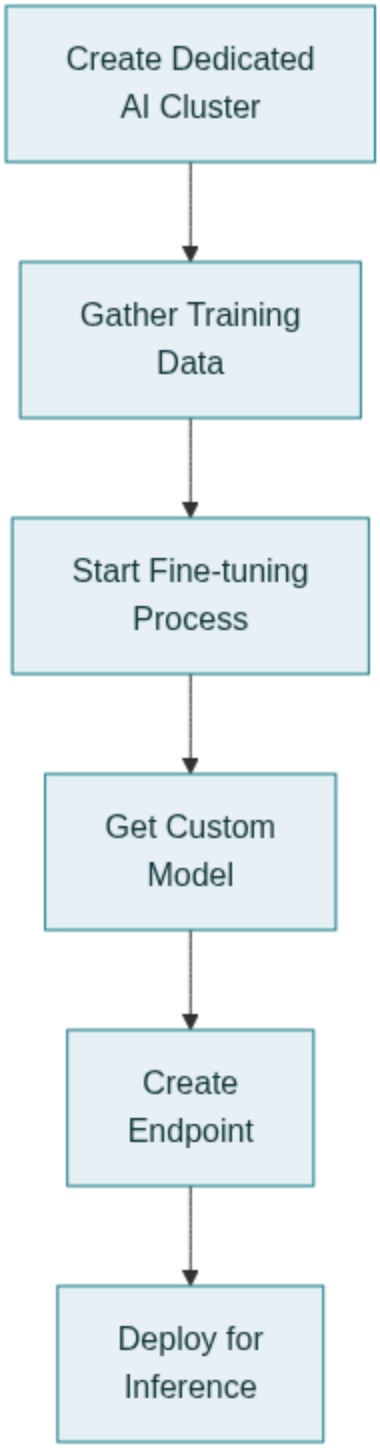
Enhanced Visual Elements:

- **OCI Service Architecture**: Complete platform overview



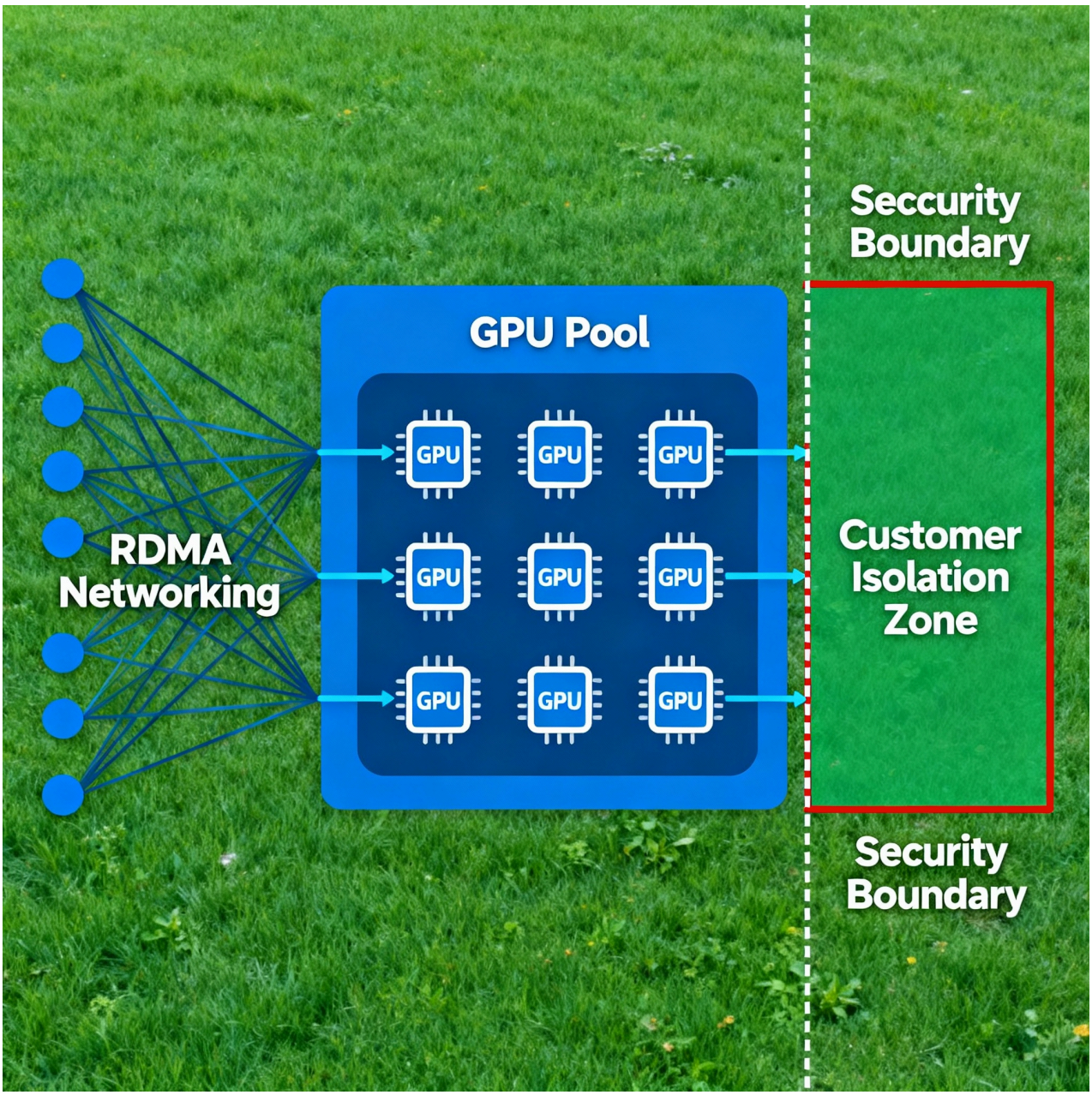
OCI Generative AI Service Architecture Overview

- **Fine-tuning Workflow**: Step-by-step process diagram



OCI Generative AI Fine-tuning Workflow

- **Dedicated Clusters**: Security isolation architecture



Dedicated AI Clusters Architecture and Security Isolation

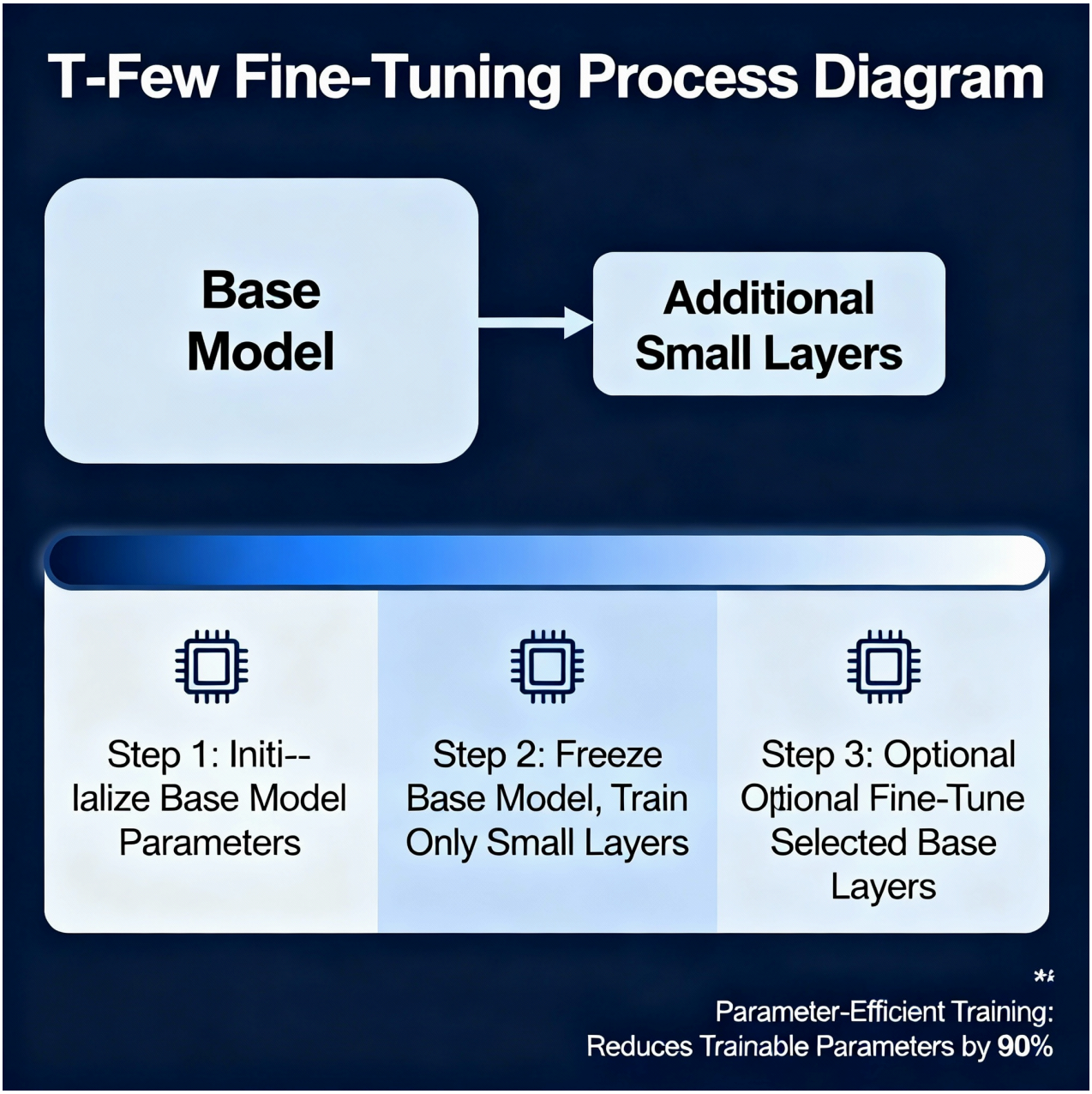
- **Cluster Types Comparison**: Detailed comparison table

OCI AI Cluster Unit Types Comparison

Unit Type	Models	Use Case	Fine-tune	Hosting
Small Cohere	Cmd R Light Cmd R 08-2024	Fine-tune & Host Cohere	2-8 units	1+ units
Large Cohere	Cmd R Plus Cmd R	Fine-tune & Host Large	Varies	1+ units
Embed Cohere	Embed EN Embed Multi	Host Embed Only	N/A	1 unit
Large Meta	Llama 3.1/3.2 (70B-405B)	Fine-tune & Host Meta	4 units	1 unit

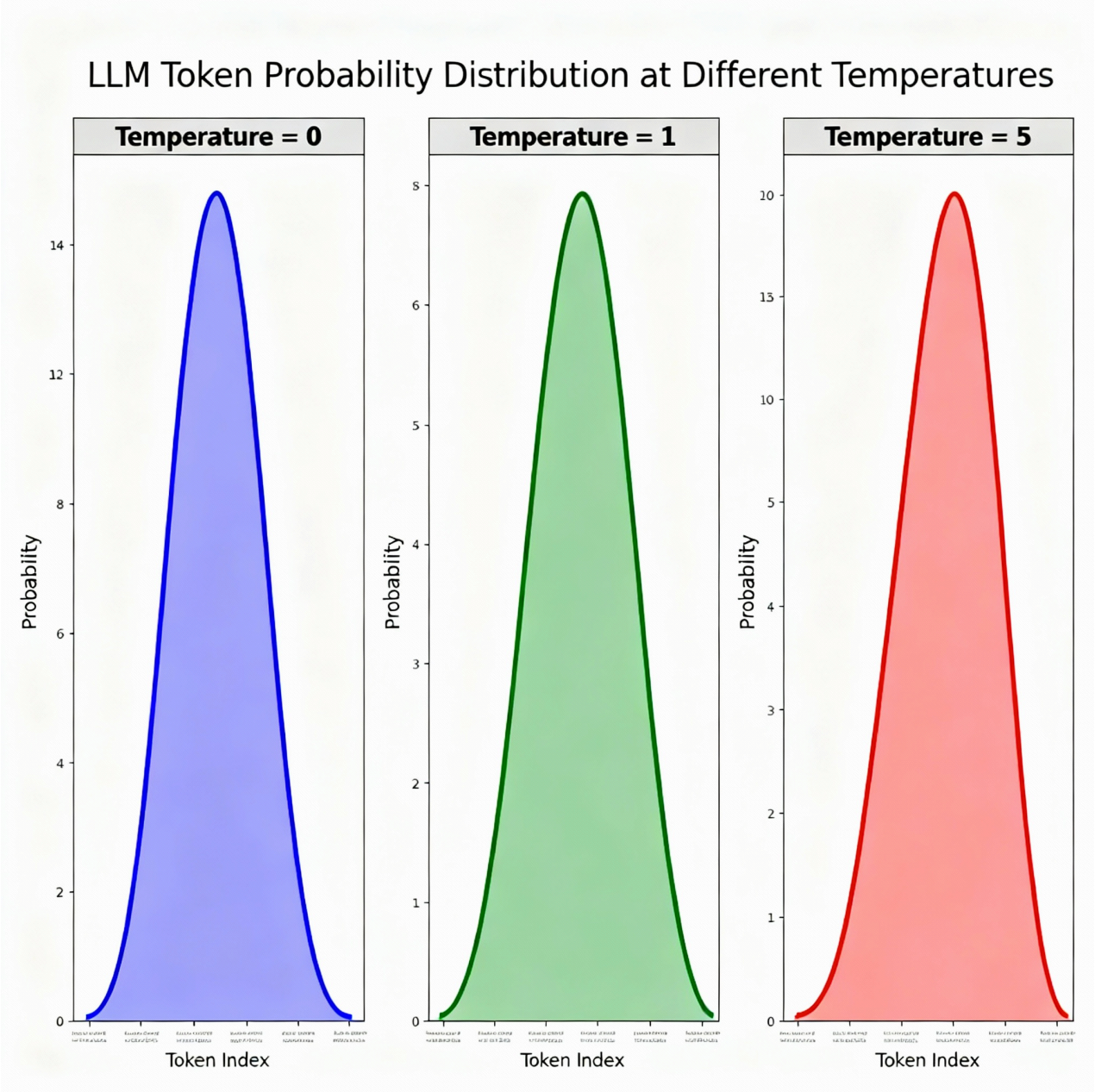
OCI Dedicated AI Cluster Unit Types Comparison

- **T-Few Process**: Parameter-efficient fine-tuning visualization



T-Few Fine-tuning Process and Parameter Efficiency

- **Temperature Effects**: Parameter control visualization



Temperature Parameter Effects on Token Selection

DevOps-Focused Content:

- **Infrastructure as Code**: Cluster creation and management
- **API Integration**: Python/Java SDK implementation examples
- **Security Best Practices**: IAM, encryption, network isolation
- **Cost Optimization**: Resource sizing and pricing strategies
- **Production Deployment**: Endpoint management and monitoring
- **Troubleshooting**: Common issues and solutions

Practical Implementation Guide:

- **Environment Setup**: Account configuration and prerequisites
- **Cluster Management**: Fine-tuning and hosting cluster creation
- **Data Preparation**: JSONL format requirements and examples
- **Fine-tuning Process**: Step-by-step custom model creation
- **Endpoint Deployment**: Production-ready model serving
- **Performance Optimization**: Best practices and monitoring

Exam-Ready Structure:

- **Key Concepts**: All certification-relevant topics covered
- **Real-world Examples**: Practical scenarios and use cases
- **Code Samples**: SDK integration and API usage
- **Best Practices**: Production deployment guidelines
- **Cost Management**: Pricing models and optimization strategies

These notes provide everything you need to understand and implement OCI Generative AI Service for your Oracle certification exam, with practical insights that will be valuable in your DevOps role as well!

⁂

1. Oracle-AI.pdf
2. https://ppl-ai-code-interpreter-files.s3.amazonaws.com/web/direct-files/92643e01726204f1d0664f0d0b16438b/a16793c2-a44d-4e3e-8f84-4938a406cae1/af275895.md