

Generative AI

Generative AI Overview

- What is Generative AI
 - Types of Generative Models
 - Gen AI vs Traditional ML
 - Real-world Applications
-

Modern Generative Models

- Transformers Overview
 - Diffusion Models
 - GANs (Quick Recap)
 - VAEs (Quick Recap)
-

Transformers & Large Language Models

- Transformer Architecture Deep Dive
 - Attention Mechanism (Self & Cross Attention)
 - Encoder vs Decoder vs Encoder-Decoder
 - GPT, BERT, T5 Models
 - How LLMs Work (Training & Inference)
-

Prompt Engineering (Core Skill)

- Prompt Design Techniques
 - Zero-shot, Few-shot Learning
 - Chain-of-Thought Prompting
 - Role-based Prompting
 - Prompt Optimization
-

Advanced LLM Concepts

- Fine-Tuning vs Prompting
 - Retrieval-Augmented Generation (RAG)
 - Embeddings & Similarity Search
 - Context Window & Token Management
-

Vector Databases & Retrieval

- Embeddings Generation
 - Vector Databases (FAISS, Pinecone, Chroma)
 - Semantic Search
 - Document Indexing
-

Gen AI Frameworks & Tools

- OpenAI API
- Hugging Face Transformers
- LangChain
- CrewAI
- PhiData

Multimodal Generative AI

- Text-to-Image (Stable Diffusion, DALL·E)
 - Image-to-Text (Captioning Models)
 - Audio Generation
 - Vision-Language Models
-

Building Gen AI Applications

- Chatbot Development (LLM-based)
 - Document Q&A Systems (RAG)
 - AI Content Generators
 - AI Assistants
 - Workflow Automation with LLMs
-

Deployment & Production

- FastAPI / Flask for LLM Apps
 - Model Deployment Strategies
 - Cloud Integration (AWS / Azure / GCP)
 - Docker for AI Applications
 - API Integration
-

Evaluation & Optimization

- LLM Evaluation Techniques
- Prompt Evaluation
- Cost Optimization (Token Usage)
- Latency & Performance Tuning

Responsible & Safe AI

- Hallucination Reduction Techniques
 - Bias & Fairness
 - Data Privacy
 - Safe AI Practices
-

Real-world Projects

- GPT-based AI Chatbot
- Document Q&A System (RAG)
- AI Resume & Content Generator
- Multimodal AI App
- Autonomous AI Assistant