

# Syllabus

## Generative AI Engineer Roadmap

---

### Phase 1: Foundations (Computer Science + AI Basics)

Skill	Tools/Libraries
Python (Advanced)	Python , typing , asyncio
Data Structures & Algorithms	Core CS concepts
Mathematics for ML	Linear Algebra, Probability, Calculus
Statistics	Descriptive & Inferential Statistics
Machine Learning Basics	scikit-learn , pandas , numpy
Deep Learning Foundations	PyTorch / TensorFlow , CNNs, RNNs, Transformers

---

### Phase 2: Core Generative AI Concepts

Concept	Tools/Libraries
Transformers Architecture	Understand attention, positional encoding
Language Models (LLMs)	GPT, BERT, T5, LLaMA
Diffusion Models	Stable Diffusion, DALL·E
Prompt Engineering	CoT, ReAct, few-shot prompting
RAG (Retrieval-Augmented Generation)	LangChain, LlamaIndex, FAISS
Fine-tuning & PEFT	LoRA, QLoRA, DPO, SFT with Hugging Face , PEFT
Embeddings	OpenAI Embeddings , SentenceTransformers , HuggingFace

---

## ✖ Phase 3: Frameworks & Tools

### ◆ LLM Frameworks

Framework	Use
LangChain	LLM chaining, agents, tools, memory
LlamaIndex	RAG pipelines with document loaders, vector stores
Haystack	NLP pipelines with search
AutoGen / CrewAI / AgentOps	Multi-agent systems
PromptLayer / Guidance / DSPy	Prompt logging & control

### ◆ Model Access & Inference

Tool	Use
Hugging Face Transformers	LLM model loading & training
Ollama	Run LLaMA, Mistral locally
vLLM	Fast inference engine for LLMs
LM Studio / Kobold	UI for local inference
Transformers.js / LLM-RS	Running models in browser / WASM

---

## 📄 Phase 4: Retrieval-Augmented Generation (RAG)

Step	Tool
Document Loading	Unstructured , LangChain , PyMuPDF , pdfplumber
Chunking	RecursiveCharacterTextSplitter , NodeParser
Embedding	OpenAI , InstructorEmbedding , Hugging Face
Vector Stores	FAISS , Chroma , Weaviate , Pinecone
Retrieval	Similarity Search, Hybrid Search
Prompt Injection	LangChain PromptTemplate or custom
Query Engine	LlamaIndex, LangChain Chains

---

## Phase 5: Agentic AI (Autonomous LLMs)

Component	Tool
Agent Execution Loop	LangChain Agents, AutoGen
Planning & Tool Use	ReAct, Plan-and-Execute, Tree-of-Thoughts
Tool Integration	APIs, Python functions, browser, search
Multi-Agent Collaboration	CrewAI, AutoGen, OpenAgents
Memory & State	LangChain Memory, Redis, Weaviate

---

## Phase 6: Evaluation & Testing

Type	Tool
Prompt Evaluation	PromptLayer , Ragas , TruLens
Output Evaluation	BLEU , ROUGE , GPT Judge , Rubrics
Model Comparison	A/B Testing, LLM-as-a-judge

---

## Phase 7: Deployment & Production

Area	Tool
Web App UI	Streamlit , Gradio , LangFlow , Next.js
Model Deployment	FastAPI , Flask , Docker , vLLM , TGI
API Serving	OpenAI , AWS Bedrock , Together.ai , Replicate
CI/CD	GitHub Actions , DockerHub , Terraform
Monitoring	PromptLayer , TruLens , Weights & Biases
Scaling	Kubernetes , Ray Serve , vLLM , Cloud Run