

AI Engineer Complete Roadmap (Start → Finish → Job Ready)

A fully structured, detailed, beginner-friendly learning timeline covering everything from Python foundations to production-grade AI engineering.

● PHASE 1: Programming Foundations (Python)

Duration: 3–6 weeks

Goals

- Build a strong foundation in Python programming.
- Understand how to work with data structures and functions.
- Learn Object-Oriented Programming.
- Get comfortable with coding environments (Anaconda, Jupyter).

What to Learn

- Python syntax and variables
- Lists, tuples, sets, dictionaries
- Conditional statements and loops
- Functions and modules
- Object-Oriented concepts (classes, objects, inheritance)
- File handling (CSV, text files)
- Virtual environments (Conda)
- NumPy basics
- Basic visualization with Matplotlib

Why This Matters

You cannot understand NLP, LangChain, or LLMs without solid Python skills.

● PHASE 2: Data Handling & Visualization

Duration: 3–4 weeks

Goals

- Start working with real-world datasets.
- Learn how to clean and manipulate data.
- Understand data analysis and visualization.

What to Learn

- Pandas (core skill)
- NumPy (intermediate concepts)
- Matplotlib + Seaborn
- Data cleaning & preprocessing
- Exploratory Data Analysis (EDA)

Skills You Should Have

- Loading CSV, Excel, JSON files
 - Handling missing values
 - Dealing with categorical text and numeric data
 - Plotting heatmaps, histograms, box plots, line charts
-

PHASE 3: Core Machine Learning

Duration: 6–10 weeks

What to Learn

- Supervised Learning
 - Classification
 - Regression
- Unsupervised Learning
 - Clustering
 - PCA (dimensionality reduction)
- Train/test split & cross-validation
- Feature engineering
- Hyperparameter tuning
- Model evaluation metrics

Essential Algorithms

- Linear Regression
- Logistic Regression
- KNN
- Decision Trees
- Random Forest
- Gradient Boosting (XGBoost/LightGBM optional)

- SVM
- K-means clustering
- PCA

Tools

- Scikit-learn
- Jupyter Notebook / Google Colab

Projects to Build

- House price prediction (regression)
 - Spam classifier (classification)
 - Customer segmentation (clustering)
-

● PHASE 4: Deep Learning (DL)

Duration: 5–8 weeks

What to Learn

- Basic neural network foundations
- Feedforward networks (ANNs)
- Loss functions & optimizers
- CNNs for image tasks
- RNNs/LSTMs basics
- Training/validation/testing
- Regularization (dropout, L2, batchnorm)

Tools

- TensorFlow 2.x
- Keras
- PyTorch (optional but beneficial)

Recommended Beginner Projects

- MNIST digit classification
 - Dog breed classifier
 - Heart disease prediction
 - Image classifier using CNNs
-

● PHASE 5: Natural Language Processing (NLP)

Duration: 4–6 weeks

What to Learn

- Tokenization & text preprocessing
- Stemming & lemmatization
- Bag of Words
- TF-IDF
- Basic NLP pipelines
- Text classification

Tools

- NLTK
- SpaCy
- Scikit-learn NLP features

Projects

- Sentiment analysis
- Spam detection
- Topic classification

● PHASE 6: Large Language Models (LLMs)

Duration: 4–6 weeks

Concepts You Learn

- Transformers architecture
- Attention mechanisms
- GPT, BERT, RoBERTa, XLNet basics
- Embeddings (why they matter)
- Prompt engineering
- Fine-tuning small transformer models

Tools

- HuggingFace Transformers
- Google Colab (GPU)
- T5, DistilBERT, BERT pretrained models

Skills You Gain

- Running pretrained models
 - Using pipelines (summarization, translation, Q&A)
 - Model inference
 - Fine-tuning transformer-based models
-

PHASE 7: Vector Databases (Pinecone) + LangChain

Duration: 4–8 weeks

Vector Database Concepts

- Embeddings
- Vectorization
- Similarity search (cosine similarity)
- Indexing
- Retrieval-Augmented Generation (RAG)

Pinecone Skills

- Creating Pinecone indexes
- Storing/upserting vectors
- Querying semantic similarity
- Integrating Pinecone with LLMs

LangChain Concepts

- LLMChain
- Sequential chains
- Tools & Agents
- Memory
- RAG pipelines
- Connecting databases + LLMs

Projects

- ChatGPT-style chatbot
- Document Q&A system
- AI assistant for websites
- PDF chat assistant using RAG

● PHASE 8: Real AI Engineering

Duration: 4–8 weeks

You Learn to Use

- APIs (OpenAI, Anthropic, Gemini, Groq)
- Building production-ready AI applications
- Deployment & hosting

Must-Know Skills

- Streamlit or FastAPI
- Basic Docker
- Git/GitHub workflow
- CI/CD basics (GitHub Actions optional)

Advanced AI Projects

- RAG chatbot with LangChain
 - Voice assistant with speech recognition
 - AI content generator
 - Customer support automation bot
 - AI search engine for PDF documents
-

● PHASE 9: Portfolio + Resume + Job Prep

Duration: 2–3 weeks

What You Need to Be Job-Ready

- 6–10 high-quality ML/AI projects
- At least 3–4 advanced AI engineering projects (RAG, LLM apps)
- Polished GitHub repositories
- Detailed READMEs
- Kaggle submissions (optional but helpful)
- Resume tailored for ML/AI roles

What to Prepare

- LinkedIn profile optimized for AI roles
- Portfolio website (optional but powerful)

- Start applying for internships & entry-level roles
-

FULL TIMELINE: How Long Will It Take?

Study Pace Total Duration

2–3 hrs/day 8–12 months

4–6 hrs/day 5–7 months

Full time 3–4 months

This is the exact learning journey followed by:

- AI Engineers
 - Machine Learning Engineers
 - Data Scientists
 - LLM App Developers
-

Where YOU Should Start

Begin with **Phase 1: Python Foundations**.

Don't jump to NLP, LLMs, or LangChain until your coding basics are strong.

If you want, I can also add:

- A **weekly study plan**
- A **daily task breakdown**
- A **project list with difficulty levels**
- A **resource list (free + paid)**
- A **checklist to track progress**

Just tell me and I'll include it!