NLP Engineer Roadmap

Instructor: Rasel sarker

1. Prerequisites

Programming:

- Language: Python
- Libraries: NumPy, Pandas, Matplotlib, Scikit-learn

Mathematics:

- Basic Linear Algebra (vectors, matrices)
- Probability & Statistics
- Calculus (basic understanding of gradients)

2. Core NLP Concepts

Text Processing:

- Tokenization
- Stemming and Lemmatization
- Stopword Removal
- POS Tagging
- Named Entity Recognition (NER)

Language Understanding:

- Bag of Words (BoW)
- TF-IDF
- Word Embeddings: Word2Vec, GloVe, FastText

Practice projects:

- Text classification (spam detection)
- Named entity recognition
- POS tagging visualization

3. Machine Learning for NLP

Supervised Learning Algorithms:

Linear Models

- Linear Regression
- Logistic Regression
- Ridge Regression
- Lasso Regression
- Elastic Net

Classification Algorithms

- Naive Bayes (Gaussian, Multinomial, Bernoulli)
- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)
- Decision Tree
- Random Forest
- Extra Trees Classifier
- Gradient Boosting
- ensemble technique(XGBoost, LightGBM, CatBoost)
- Perceptron

Unsupervised Learning Algorithms

Clustering

- K-Means
- Hierarchical Clustering
- Gaussian Mixture Model (GMM)

Dimensionality Reduction

- PCA
- t-SNE

Libraries:

- NLTK, SpaCy, TextBlob

Practice projects:

- Sentiment analysis on tweets or movie reviews

4. Deep Learning for NLP

Core Concepts:

- CNN, RNNs, LSTMs, GRUs
- Sequence-to-Sequence models
- Attention Mechanism

Frameworks:

- TensorFlow
- PyTorch

Practice projects:

- Chatbot using Seq2Seq
- Text summarization
- Language translation

5. Transformers and Pretrained Language Models

1. you have to learn:
   Transformers Architecture
   Hugging Face Transformers library
   BERT, RoBERTa, GPT, T5, etc.

2. Let’s practice:
   Text classification with BERT
   Question answering
   Named entity recognition using pretrained models
   Fine-tuning a transformer for your custom dataset

6. Advanced Topics in NLP

Foundation-Level:

- **Language Modeling (Auto-regressive & Masked LM)**
- **Zero-shot, One-shot & Few-shot Learning**
- **Prompt Engineering & Prompt Tuning**
- **Fine-tuning LLMs:**
  - Full fine-tuning
  - Parameter-efficient fine-tuning methods:
    - **LoRA (Low-Rank Adaptation)**
    - **QLoRA (Quantized LoRA)**
- PEFT techniques (transformers, peft, trl libs)

Conversational AI:

Chatbot development with:

- Retrieval-Augmented Generation (RAG)
- Memory-augmented chains
- Context-aware multi-turn dialogs
- LangChain or LlamaIndex integrations

AI Agents & Agentic Workflows:

Frameworks:

- CrewAI
- LangGraph
- AutoGen
- OpenAgents
- LangChain Agents
- n8n (Visual agent orchestration)

Multimodal NLP (Text + Image/Audio/Video):

Applications:

- Visual Question Answering (VQA)
- Multimodal Search
- Audio/text agents (Whisper + GPT)
- Image captioning, OCR + LLM pipeline

7. Build & Showcase Real-World Projects

Project Ideas:

- Resume Parser
- Medical Chatbot
- Agricultural Chatbot
- Chatbot for customer service
- Email classification and summarizer
- AI tutor for grammar correction
- Search engine using RAG
- Audio transcript analyzer

Thanks for watching