Introduction to NLP

- It is the **technology that helps computers understand and work with human language** (like English, Hindi, Marathi, etc.)

- It's used in **chatbots, Google Translate, Siri, spam filters, grammar checkers**, etc.

- It mixes **computer science + linguistics (study of language) + AI (Artificial Intelligence)**

📌 What Is NLP?

Natural Language = The way **humans naturally speak or write**

Processing = The way **computers analyze, understand, or generate** that language

🔁 So, NLP is:

> Teaching computers to understand, interpret, and respond to human languages like humans do.

📌 Why Is NLP Needed?

Imagine trying to talk to a robot using English. Computers speak only in **0s and 1s (binary)**. But we speak in complex sentences like:

> "Can you book a flight for tomorrow evening to Delhi?"

This has:

- Ambiguity (what time is "evening"?)

- Context (from where to where?)

- Grammar

- Idioms

- Misspellings sometimes

⚠️ **Challenge**: Human language is **messy, emotional, and filled with multiple meanings.**

✅ **Goal of NLP**: Bridge the gap between **human language** and **computer understanding**.

What Can NLP Do? (Real-World Examples)

Application	What NLP Does
🔍 Google Search	Understands your search even with spelling errors
💬 Chatbots (e.g., Alexa)	Understands your questions and responds naturally
📧 Spam Filters	Detects unwanted emails based on text patterns
🗣️ Voice Assistants	Converts spoken language to text and vice versa
🌐 Google Translate	Translates text between languages
📝 Grammarly, MS Word	Checks grammar, spelling, sentence structure
👁️ Sentiment Analysis	Finds out if a review is happy, sad, angry, etc.

📌 How NLP Works Internally (Step-by-Step Like a Machine)

Let's say the sentence is:

"I love pizza."

Step 1: Text Input

Input from user: "I love pizza."

Step 2: Tokenization

Break sentence into smaller pieces called **tokens**:

```
["I", "love", "pizza", "."]
```

Paragraph → Sentences → Words

> 💡 Words & sentences both can be vecors

Step 3: Text Cleaning

Remove punctuation, convert to lowercase, etc.:

```
["i", "love", "pizza"]
```

Step 4: Part of Speech Tagging (POS)

Label words:

```
i (pronoun), love (verb), pizza (noun)
```

Step 5: Named Entity Recognition (NER)

Find special words like names, places, dates:

> No named entity here, but in:
>
> "I live in Mumbai." → **Mumbai = location**

Step 6: Dependency Parsing

Understand grammar relationships:

- "love" is the main verb
- "I" is subject
- "pizza" is object

Step 7: Sentiment Analysis

Detect emotion:

> "I love pizza." → Positive emotion ✅

Step 8: Language Modeling / Response

If a chatbot, generate a reply like:

> "Pizza is great! What's your favorite topping?"

📌 Techniques in NLP (Two Main Types)

Type	Explanation
Rule-Based NLP	Based on grammar rules written by humans
Machine Learning NLP	Learns patterns from large amounts of text (data-driven)

Modern NLP mostly uses **machine learning**, especially **deep learning** with models like:

> 💡 Modern NLP mostly uses **machine learning**, especially **deep learning** with models like:
>
> - **BERT**, **GPT**, **T5**, **LLaMA**, etc.

📌 NLP Building Blocks

Concept	What It Does
Tokenization	Split sentences into words or subwords
Stop Words Removal	Remove common words like "the", "is", "and"
Stemming/Lemmatization	Reduce words to base/root form ("running" → "run")
POS Tagging	Label each word with its role (noun, verb, etc.)
NER	Identify names of people, places, etc.
Parsing	Analyze grammar structure
Embeddings	Turn words into numbers computers understand (like vectors)

Terminologies

Corpus → Paragraph

Document → Sentence

Vocabulary → Unique Words

Words → All the words present in corpus

🛠️ Common Tools and Libraries

- **NLTK** – Natural Language Toolkit (great for learning)

- **spaCy** – Industrial-strength NLP in Python

- **TextBlob** – Simple API for common NLP tasks

- **Transformers (Hugging Face)** – State-of-the-art models like BERT, GPT

- **OpenAI APIs** – Access to advanced language models like GPT-4