ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

DAY – 17

Date: Jul 15, 2025

Natural Language Processing (NLP)

Natural Language Processing (NLP) is a field that combines computer science, artificial intelligence and language studies. It helps computers understand, process and create human language in a way that makes sense and is useful. With the growing amount of text data from social media, websites and other sources, NLP is becoming a key tool to gain insights and automate tasks like analyzing text or translating languages.

NLP Techniques

NLP encompasses a wide array of techniques that aimed at enabling computers to process and understand human language. These tasks can be categorized into several broad areas, each addressing different aspects of language processing. Here are some of the key NLP techniques:

1. Text Processing and Preprocessing

- **Tokenization:** Dividing text into smaller units, such as words or sentences.
- **Stemming and Lemmatization:** Reducing words to their base or root forms.
- **Stopword Removal:** Removing common words (like "and", "the", "is") that may not carry significant meaning.
- **Text Normalization:** Standardizing text, including case normalization, removing punctuation and correcting spelling errors.

2. Syntax and Parsing

- **Part-of-Speech (POS) Tagging:** Assigning parts of speech to each word in a sentence (e.g., noun, verb, adjective).
- **Dependency Parsing:** Analyzing the grammatical structure of a sentence to identify relationships between words.
- **Constituency Parsing:** Breaking down a sentence into its constituent parts or phrases (e.g., noun phrases, verb phrases).

3. Semantic Analysis

- **Named Entity Recognition (NER):** Identifying and classifying entities in text, such as names of people organizations, locations, dates, etc.
- **Word Sense Disambiguation (WSD):** Determining which meaning of a word is used in a given context.
- **Coreference Resolution:** Identifying when different words refer to the same entity in a text (e.g., "he" refers to "John").

4. Information Extraction

- **Entity Extraction:** Identifying specific entities and their relationships within the text.
- **Relation Extraction:** Identifying and categorizing the relationships between entities in a text.

5. Text Classification in NLP

- **Sentiment Analysis:** Determining the sentiment or emotional tone expressed in a text (e.g., positive, negative, neutral).
- **Topic Modeling:** Identifying topics or themes within a large collection of documents.
- **Spam Detection:** Classifying text as spam or not spam.

6. Language Generation

- **Machine Translation:** Translating text from one language to another.
- **Text Summarization:** Producing a concise summary of a larger text.
- **Text Generation:** Automatically generating coherent and contextually relevant text.

7. Speech Processing

- **Speech Recognition:** Converting spoken language into text.
- **Text-to-Speech (TTS) Synthesis:** Converting written text into spoken language.

8. Question Answering

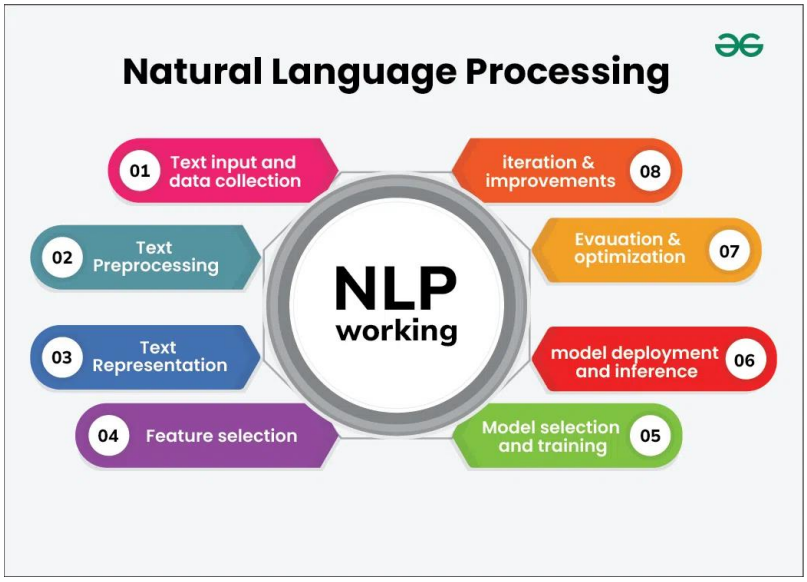
- **Retrieval-Based QA:** Finding and returning the most relevant text passage in response to a query.
- **Generative QA:** Generating an answer based on the information available in a text corpus.

9. Dialogue Systems

- **Chatbots and Virtual Assistants:** Enabling systems to engage in conversations with users, providing responses and performing tasks based on user input.

10. Sentiment and Emotion Analysis in NLP

- **Emotion Detection:** Identifying and categorizing emotions expressed in text.
- **Opinion Mining:** Analyzing opinions or reviews to understand public sentiment toward products, services or topics.



Applications of Natural Language Processing (NLP)

- **Spam Filters:** One of the most irritating things about email is spam. Gmail uses natural language processing (NLP) to discern which emails are legitimate and which are spam. These spam filters look at the text in all the emails you receive and try to figure out what it means to see if it's spam or not.
- **Algorithmic Trading:** Algorithmic trading is used for predicting stock market conditions. Using NLP, this technology examines news headlines about companies and stocks and attempts to comprehend their meaning in order to determine if you should buy, sell or hold certain stocks.
- **Questions Answering:** NLP can be seen in action by using Google Search or Siri Services. A major use of NLP is to make search engines understand the meaning of what we are asking and generate natural language in return to give us the answers.
- **Summarizing Information:** On the internet, there is a lot of information and a lot of it comes in the form of long documents or articles. NLP is used to decipher the meaning of the data and then provides shorter summaries of the data so that humans can comprehend it more quickly.