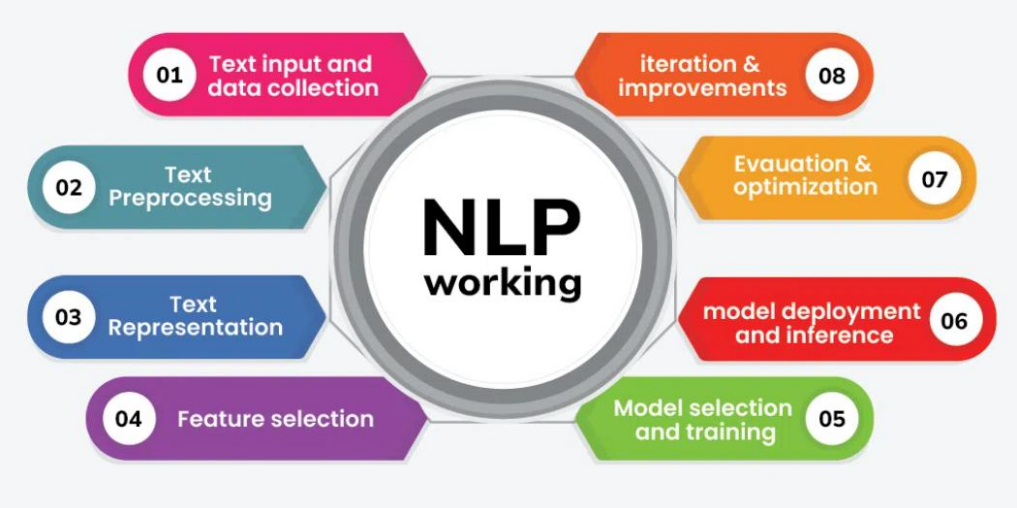
ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

DAY – 17 **15 July 2025**

Natural Language Processing (NLP)

Natural Language Processing (NLP) is a field of artificial intelligence that focuses on enabling computers to understand, interpret, generate, and respond to human language in a meaningful way. It combines computational linguistics with machine learning and deep learning to process language data. At its core, NLP allows machines to read and comprehend text or speech similarly to how humans do. The basic building blocks of NLP include text preprocessing (such as cleaning, tokenization, and normalization), syntax analysis (like part-of-speech tagging and parsing), and semantic analysis (understanding meaning and context). Common NLP tasks include sentiment analysis, machine translation, text summarization, named entity recognition (NER), and question answering. NLP is widely used in applications like chatbots, search engines, voice assistants (like Siri or Alexa), and spam detection. By bridging the gap between human communication and computer understanding, NLP plays a vital role in making technology more intuitive and interactive.



Text Processing in NLP

1. **Definition**:
   Text processing is the initial and essential step in NLP that transforms raw, unstructured text into a clean, structured, and machine-readable format.

2. **Purpose**:
   Prepares text for deeper linguistic or statistical analysis by NLP models.
3. **Key Steps in Text Processing**:
   - **Text Cleaning**:
     - Removes unwanted characters like punctuation, numbers, special symbols, and extra spaces.
     - Converts all text to lowercase for consistency.
   - **Tokenization**:
     - Splits text into smaller units (tokens) such as words or sentences.
   - **Stopword Removal**:
     - Eliminates common words like “is,” “the,” and “and” that typically carry little semantic value.
   - **Normalization**:
     - **Stemming** and **lemmatization** reduce words to their root forms.
     - Ensures words like “running,” “ran,” and “runs” are treated the same (→ “run”).
   - **Part-of-Speech (POS) Tagging** (optional but useful):
     - Identifies the grammatical role of each word (e.g., noun, verb, adjective).
4. **Applications**:
   - Text classification
   - Machine translation
   - Sentiment analysis
   - Named entity recognition
   - Information extraction
5. **Importance**:
   - Simplifies and standardizes input data.
   - Enhances accuracy and performance of NLP models.