

Q1. Explain Natural Language Processing (NLP). Discuss its main components and applications.

Ans. Natural Language Processing (NLP) is a branch of Artificial Intelligence that focuses on enabling computers to understand, interpret, and generate human language in a meaningful way. It acts as an interface between human communication and computer systems by processing both spoken and written language.

The main components of NLP include:

- Lexical Analysis, which deals with breaking text into words or tokens.
- Syntactic Analysis, which analyzes the grammatical structure of sentences.
- Semantic Analysis, which focuses on understanding the meaning of words and sentences.
- Pragmatic Analysis, which interprets language based on context and real-world knowledge.

NLP is widely used in applications such as machine translation, sentiment analysis, speech recognition, chatbots, information retrieval systems, and text summarization. These applications help automate tasks that require understanding natural language, improving human-computer interaction.

Q2. Describe the Bag-of-Words (BoW) model. What are its advantages and limitations in text representation?

Ans. The Bag-of-Words (BoW) model is a simple and widely used technique for representing text data in Natural Language Processing. In this model, a document is represented as a collection of words without considering the order or grammatical structure of the words. Each document is converted into a vector based on the frequency of words from a predefined vocabulary.

One major advantage of the Bag-of-Words model is its simplicity and ease of implementation. It is computationally efficient and works well for basic text classification and information retrieval tasks. It also allows machine learning algorithms to process textual data numerically.

However, the Bag-of-Words model has several limitations. It ignores word order and context, which can lead to loss of semantic meaning. It also cannot handle synonyms or polysemy effectively and often results in high-dimensional sparse vectors. Due to these limitations, BoW is less effective for complex NLP tasks that require deeper language understanding.

Q3. Explain how transformers work in Natural Language Processing. Mention the role of the attention mechanism.

Ans. Transformers are a modern deep learning architecture used extensively in Natural Language Processing tasks such as translation, text generation, and question answering. Unlike traditional sequence-based models such as RNNs and LSTMs, transformers process the entire input sequence simultaneously, enabling better parallelization and efficiency.

The core component of the transformer architecture is the attention mechanism, specifically self-attention. Self-attention allows the model to determine the importance of each word in a sentence relative to other words. This helps the model capture long-range dependencies and contextual relationships more effectively.

Transformers consist of encoder and decoder blocks, each containing multi-head attention layers and feed-forward neural networks. The attention mechanism assigns weights to words based on relevance, allowing the model to focus on key information. Due to this design, transformers achieve superior performance and scalability, forming the foundation of models like BERT, GPT, and TrOCR.