PRESENTATION

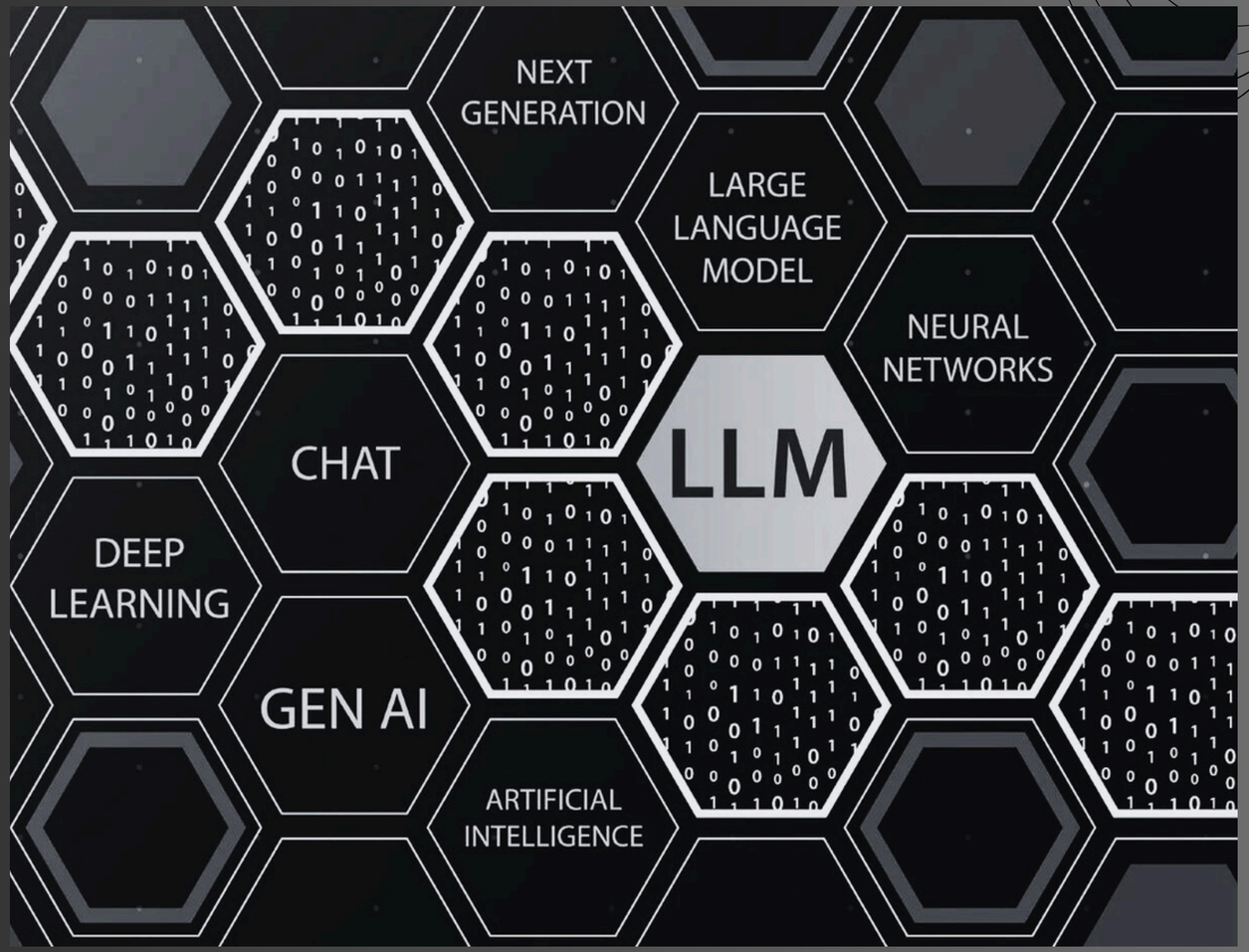
SummaShift

Delivers swift summaries, allowing you to digest information quickly and efficiently.

LLM (Large Language Model) :-

An example of artificial intelligence (AI) that has been trained on enormous volumes of text data is the LLM model.
With the use of large datasets, these models are trained to comprehend the complexities and structure of language, allowing them to carry out activities like question answering, content creation, and language translation support.
Their ability to comprehend and produce human language makes them perfect for activities involving presentations.

Examples of popular LLMs include **GPT-3, T5,** and **Jurassic-1 Jumbo.**



Text Summarisation using NLP models

Text summarization using natural language processing (NLP) techniques involves condensing a piece of text, such as an article, document, or webpage, into a shorter version while preserving its key information and meaning.

There are two main approaches to text summarization :-

- **Extractive summarization**: Selects the most important sentences or phrases from the original text.
- **Abstractive summarization:** Generates new text that captures the essence of the original content.

Various Techniques of Text Summarization using NLPs

Text summarization is a vital task in natural language processing (NLP) aimed at condensing large volumes of text while retaining the most important information.

Here are some common techniques which have we used for text summarization :-

WORD FREQUECNY

Word frequency refers to the number of times a word appears in a given text or corpus.

TF-IDF

TF-IDF, is a numerical statistic used in natural language processing (NLP) to measure the importance of a word in a document relative to a collection of documents.

BART

BART is pre-trained on large text corpora using techniques like masked language modeling and denoising autoencoding.

Word Frequency Techniques

Word frequency refers to the number of times a word appears in a given text or corpus. The working of word frequency involves counting how often each word appears in a given text or corpus.

This process typically includes the following steps :-

- **Tokenization :-** The text is divided into individual words or tokens, usually by splitting the text at spaces or punctuation marks.
- **Counting :-** Each unique word is counted to determine its frequency of occurrence in the text.
- **Ranking :-** The words are ranked based on their frequency, with the most frequently occurring words appearing at the top of the list.
- **Analysis :-** The frequency distribution of words can be analyzed to identify patterns, trends, or important terms in the text.

Term Frequency - Inverse Document Frequency

The TF-IDF (Term Frequency-Inverse Document Frequency) algorithm is a technique used in natural language processing (NLP) to measure the importance of a word in a document relative to a collection of documents.

This process typically includes the following steps :-

- **Term Frequency (TF) :-** Term frequency measures how often a term appears in a document.
  TF = (Number of times term appears in a document) / (Total number of terms in the document)
- **Inverse Document Frequency (IDF) :-** Inverse document frequency measures how unique or rare a term is across a collection of documents.
  IDF = log_e(Total number of documents / Number of documents containing the term)
- **TF-IDF Calculation :-** TF-IDF = TF * IDF
- **TF-IDF is commonly used in information retrieval, text mining, and document classification tasks.**

Bidirectional and Auto-Regressive Transformers

BART is a state-of-the-art model in natural language processing (NLP) developed by Facebook AI. It combines the strengths of bidirectional transformers with auto-regressive models, making it proficient in both understanding and generating natural language text.

This process typically includes the following steps :-

- **Architecture :-** BART is based on the transformer architecture, which consists of encoder and decoder layers.
- **Bidirectional Learning :- I**t can understand the context of a word or phrase by considering both preceding and succeeding words.
- **Auto-Regressive Decoding :-** During the decoding phase, BART generates text sequentially, predicting one token at a time based on the previously generated tokens.
- **Pre-training :-** Pre-training involves tasks such as masked language modeling and denoising autoencoding, which help BART learn rich representations of language.
- **Fine-tuning :-** After pre-training, BART can be fine-tuned for various downstream tasks, such as text summarization, translation, or question answering.

Evaluation of the generated Summaries :-

Evolutionary techniques are commonly used in natural language processing (NLP) tasks to improve the performance of machine translation systems.

Let's break down how these techniques are applied to three popular evaluation metrics: Rouge, BLEU, and METEOR.

ROUGE

Rouge, standing for "Recall-Oriented Understudy for Gisting Evaluation," measures the quality of summaries by comparing them to reference summaries

BLEU

BLEU, or "Bilingual Evaluation Understudy," evaluates the quality of machine-translated text by comparing it to one or more reference translations.

METEOR

METEOR, which stands for "Metric for Evaluation of Translation with Explicit Ordering," is another metric used for evaluating machine translation output.

THANK YOU

FOR YOUR NICE ATTENTION

Telephone
+123-456-7890

Github Repository :-
https://github.com/arusix-46/4th-Sem-Project