

# **Center For Artificial Intelligence**

## **Dr. B.R. Ambedkar National Institute of Technology, Jalandhar**

### **Report on Expert Talk Lecture**

Date: 14/08/2024

Time: 5:00-6:00pm

Topic: Information Retrieval using Large Language Models (LLMs)

Speaker: Dr Kushal Shah, Professor Computer Science Sitare University

The expert lecture on “Information Retrieval using Large Language Models (LLMs)” was organized by the Center for Artificial Intelligence. The talk was given by Dr. Kushal Shah, Professor of Computer Science (Sitare University). He is the founder of Building Self-Shiksha for AI/ML Education. He is currently working on some open-source projects related to Machine Learning and Natural Language Processing. He is also developing a self-learning platform for Data Science & Machine learning ([www.bekushal.com](http://www.bekushal.com)).

It was organized with the focus to transfer the knowledge directly from the people who mastered the technologies to the students of MTech AI.

The presentation was well-structured, beginning with an overview of natural language processing and its applications. The topic of information retrieval was then discussed, which is a basic idea in computer science, notably in the fields of information systems and search engines. The discussion then moves to focus on word embedding models, BERT encoders, GPT decoders, Sentence-BERT models, and the limitations of LSTM-based embeddings. At the end it was a fruitful session for the budding engineers of Nit Jalandhar.

### **Information Retrieval using Large Language Models**

#### **Introduction**

Large language models (LLMs) improve information retrieval by taking advantage of enhanced natural language processing capabilities. The process starts with a user inquiry, which is carefully evaluated to determine its intent and context. This query is then matched against a large collection of papers via indexing, which organizes documents for easy access.

LLMs can improve the search phase by including complex techniques such as semantic understanding and contextual embeddings, resulting in more accurate retrieval of relevant content. The result is a ranked list of papers prioritized depending on their relevance to the query, giving consumers with more precise and contextually suitable data. The integration of LLMs into information retrieval systems improves both the user experience and the effectiveness of data access.

#### **Natural Language Processing**

Natural Language Processing (NLP) is a branch of artificial intelligence that studies the interaction between computers and humans using natural language. It allows

machines to understand, interpret, and synthesize human language, which is useful for applications like translation, sentiment analysis, and conversational agents.

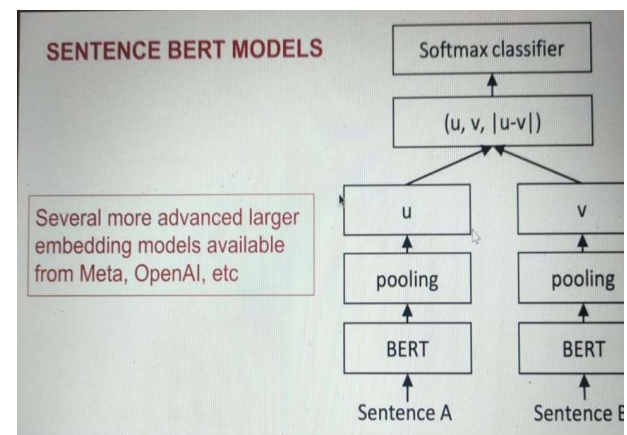
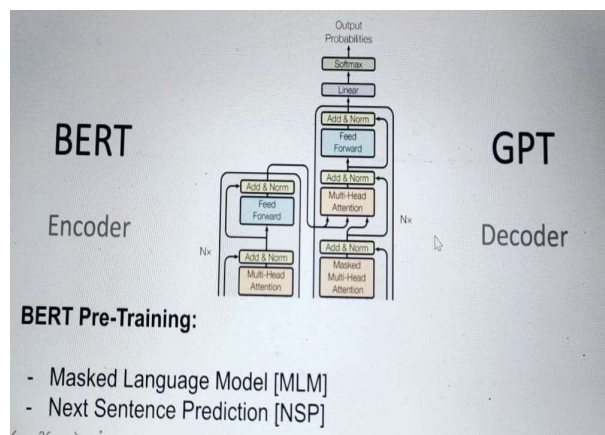
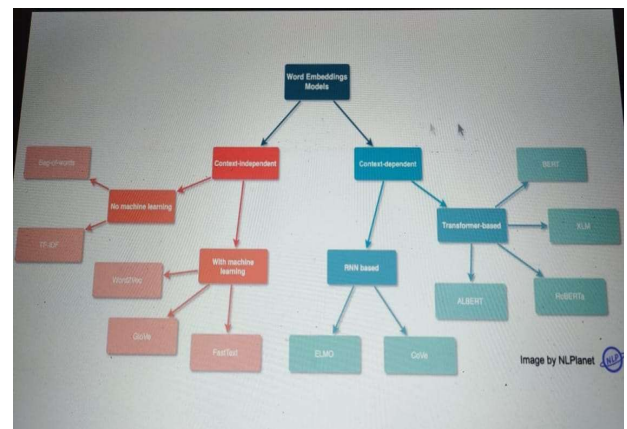
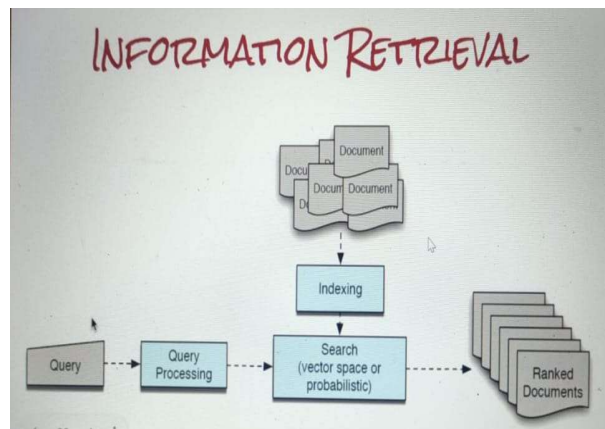
### Language Representation: Feature Extraction and Embedding Models

Language representation is a crucial component of natural language processing (NLP) tasks. It involves converting text data into a format that can be understood by machine learning algorithms. Feature extraction and embedding models are two fundamental techniques used in language representation.

- Feature extraction is the process of extracting relevant features from text data. These features can be used to represent the text in a format that can be processed by machine learning algorithms.
- Embedding models are a type of feature extraction technique that represents words or characters as dense vectors in a high-dimensional space. These vectors can capture the semantic meaning of the words or characters.

### Conclusion

The use of Large Language Models for information retrieval is a key achievement in AI and NLP. IR systems can improve search accuracy, context, and user-friendliness by utilizing LLMs' deep learning capabilities. As technology advances, we should expect better access and interaction with information across multiple areas.



Expert Lecture Slides

Submitted by: Nilanshu Nilay

MTech-AI (24901316)