

## **HW-Topic-5**

Data Acquisition, Modeling and Analysis: Big Data Analytics

Submitted By – **Sudhanshu Kakkar**  
CWID – **20036779**

# Large Language Models (LLMs)

Sudhanshu Kakkar | ID - 20036779

## WHAT IS IT?

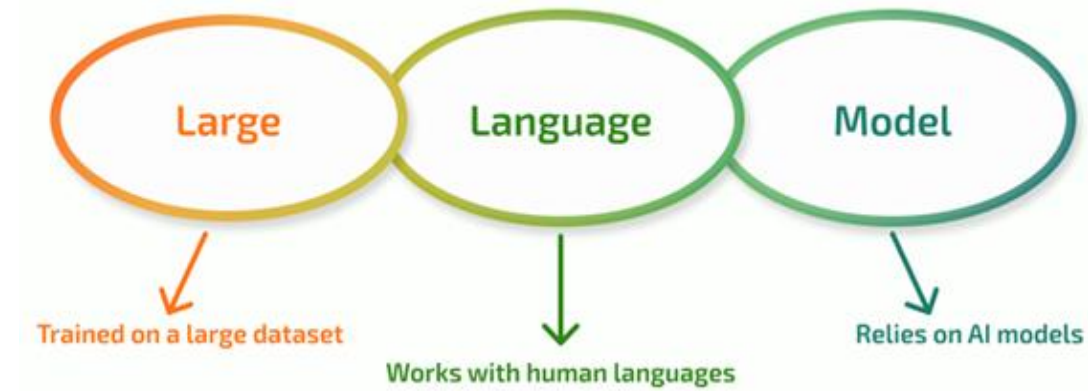
An LLM is a type of **deep learning model** designed to understand, generate, and interact with human language.

## KEY CONCEPTS

- **Transformer Architecture:** The neural network design enabling parallel data processing.
- **Attention Mechanism:** A process that weighs the importance of input tokens to understand context.
- **Tokenization:** Breaking down raw text into smaller, readable units (tokens).
- **Context Window:** The maximum number of tokens the model can process or "remember" in one interaction.

## APPLICATIONS

- Code Generation
- Human Like Chatbots
- Content Creation
- Information Summarization



## Advantages

1. Increased Productivity in daily tasks
2. Natural Language Understanding
3. Code Generation
4. Highly Scalable

## Limitations

1. Hallucinations and inaccuracies
2. Limited or outdated knowledge
3. Expensive training
4. Lack of transparency

# Fine-Tuning LLM

Sudhanshu Kakkar | ID - 20036779

## WHAT IS IT?

Fine tuning a Large Language Model (LLM) is the process of taking a pre-trained general-purpose model and adapting it for a specific task or domain by continuing to train it on a smaller, targeted dataset

## KEY CONCEPTS

- **Starting Point:** You begin with a pre-trained LLM, which has already learned extensive language and grammar, and a wide from huge amounts of public data.
- **Specialization:** You then introduce a **smaller, task specific dataset** (often a collection of labeled input/output examples).
- **Process:** The model continues its training process, adjusting its internal parameters (weights) based on the specialization dataset.

## APPLICATIONS

- **Healthcare:** Summarize clinical notes, generate patient reports.
- **Finance:** Perform sentiment analysis on financial reports and news for market trend prediction
- **Legal:** Analyze, summarize, and extract key clauses from legal documents



## Advantages

1. Industry Specific Tasks
2. Productivity in daily tasks
3. Reduced Hallucinations

## Limitations

1. Hard to Maintain with latest data
2. Finding optimum data for training becomes difficult some times.