LLM (LARGE LANGUAGE MODELS)
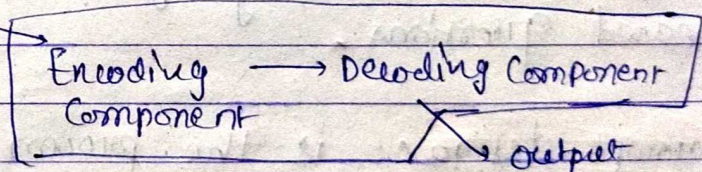
are Subset of deep learning

* LLM models refers to large, general-purpose language model that can be pre trained and the fine tuned for Specific purpose.

are trained to solve Common language problems like, document Summarization, text generation, Quest. Answers.

1) large
   - → dataset
   - → Parameters

2) General purpose
   - → Commonality of human Language
   - → Resource restrictions

3) Pretrained & fine tuned

* Transformer Model

Encoding Component → Decoding Component

→ output

GPT GLM

Generative language Models → LAMDA, Gemini, GPT etc.

LLM Development V/s traditional development

LLM Development (using Pre-trained API's)	Traditional ML Development
• No ML expertise needed	• ML expertise needed
• No training examples	• training examples
• No need to train a model	• Compute time + Hardware
• Thinks about prompt design.	• thinks about minimizing a loss function.

Question Answering in NLP?
Natural language processing?

(QA) → is a Subfield of NLP that deals with the task of automatically answering questions posed in natural language.

QA System → are able to answer a wide range of questions, including factual, definitional, and opinion based questions.

Prompt design :- is the process of creating a prompt that is tailored to the Specific task that the System is being asked to perform.

Prompt Engineering :- is the process of creating a prompt that is designed to improve performance.
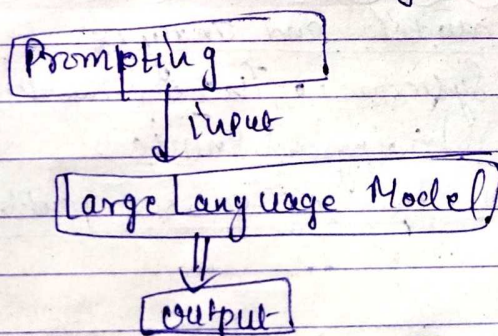
Prompt design → More Jenral

Prompt Engineering → More Specialised for system that require a high degree of accuracy and performance.

Three kind of LLM
i) & Generic LLM
ii) Instruction tuned
iii) dialog tuned.

i) Generic LLM → predict the next Word (technically token) based on the language in the training data.

ii) Instruction tuned → Trained to predict a response to the instructions given in the input.

┌─────────────┐
│ Prompting   │
└─────────────┘
      │ input
      ↓
┌─────────────────────┐
│ Large Language Model │
└─────────────────────┘
      ║
      ↓
┌────────┐
│ output │
└────────┘

iii) Dialogue tuned :- Trained to have a dialogue by predicting the next response

Chain of thought Reasoning ⟶ Models are better at getting the right answer when they first output text that explain the reason for answers.

Tuning ⟹ The process of adapting a model to a new domain or set of custom use cases by training the model on new data.

More efficient Methods of tuning

(PETM) Parameters efficient tuning Methods (PETM) :-
Methods for tuning an LLM on your own custom data without duplicating the model.

Prompt tuning :- one of the easiest parameter-efficient tuning methods.