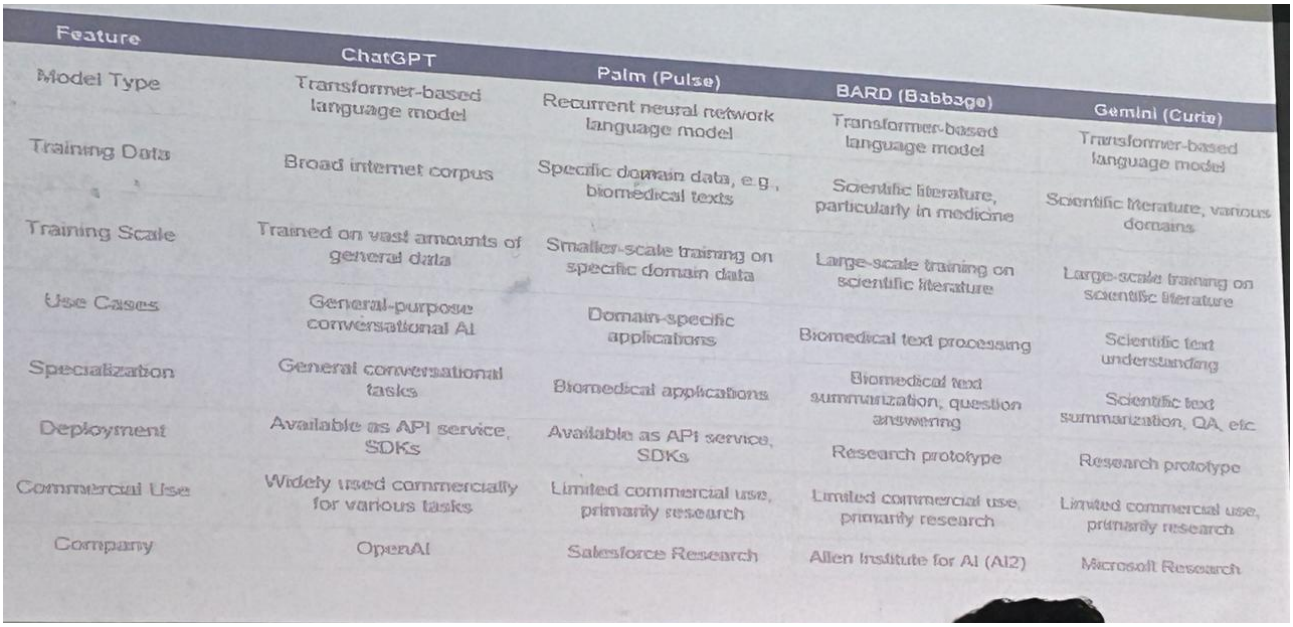
INTRODUCTION TO LLM

- LLM is a Large Language Model used to read, write and chat.
- Traditional AI works in predicting where structured dataset is available but when there is no an unstructured data, LLM is used
- LLM can have any number of parameters. The worry about the mean square error can be eliminated when using LLM.
- In scope of NLP, it will interpret, analyse and summarise the data whereas LLM's can generate new data using the pre-existing data
- NLP required structured data whereas LLM can work with unstructured data
- NLP requires so much of time to process huge data but LLM works very faster even with huge datasets

ChatGPT : Transformed based chat model
Palm : Recurrent neural model
Bard : Transformer based , requires so much of time
Gemini : Transformer based



Feature	ChatGPT	Palm (Pulse)	BARD (Babbage)	Gemini (Curie)
Model Type	Transformer-based language model	Recurrent neural network language model	Transformer-based language model	Transformer-based language model
Training Data	Broad internet corpus	Specific domain data, e.g., biomedical texts	Scientific literature, particularly in medicine	Scientific literature, various domains
Training Scale	Trained on vast amounts of general data	Smaller-scale training on specific domain data	Large-scale training on scientific literature	Large-scale training on scientific literature
Use Cases	General-purpose conversational AI	Domain-specific applications	Biomedical text processing	Scientific text understanding
Specialization	General conversational tasks	Biomedical applications	Biomedical text summarization, question answering	Scientific text summarization, QA, etc.
Deployment	Available as API service, SDKs	Available as API service, SDKs	Research prototype	Research prototype
Commercial Use	Widely used commercially for various tasks	Limited commercial use, primarily research	Limited commercial use, primarily research	Limited commercial use, primarily research
Company	OpenAI	Salesforce Research	Allen Institute for AI (AI2)	Microsoft Research

Prompt design VS Prompt Engineering

Prompt design - designing or structuring a prompt
Prompt Engineering - optimise the prompt preference efficiently for desired outputs

Tuning and Fine Tuning

Tuning : process of adjusting hyper parameters or model configurations during training to optimize the LLM's performance on a specific task or dataset.

Fine-Tuning : process of taking a pre-trained LM and further training it on a task-specific dataset or for a specific task. Fine tuning is a very easy task (not so expensive)

It can be done in two ways ,
1) Reinforcement learning platform
2) Retrieval Argument Generation - (RAG)

Life Cycle of LLM

Scope of the project -> Build / improve your model -> interpret evaluation -> deploy and monitor

LLM Models

Closed source model : Very expensive, based on cloud and requires an API key
Open sourced model : can be installed using pip and can be easily accessed. Requires fine tuning