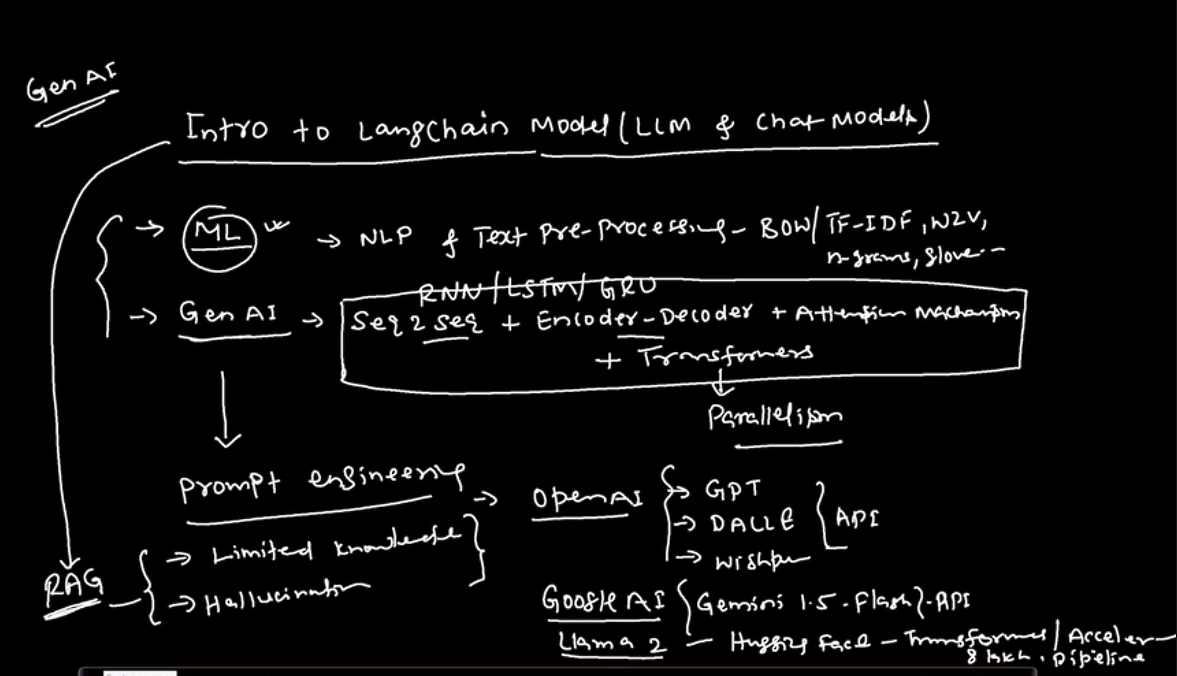
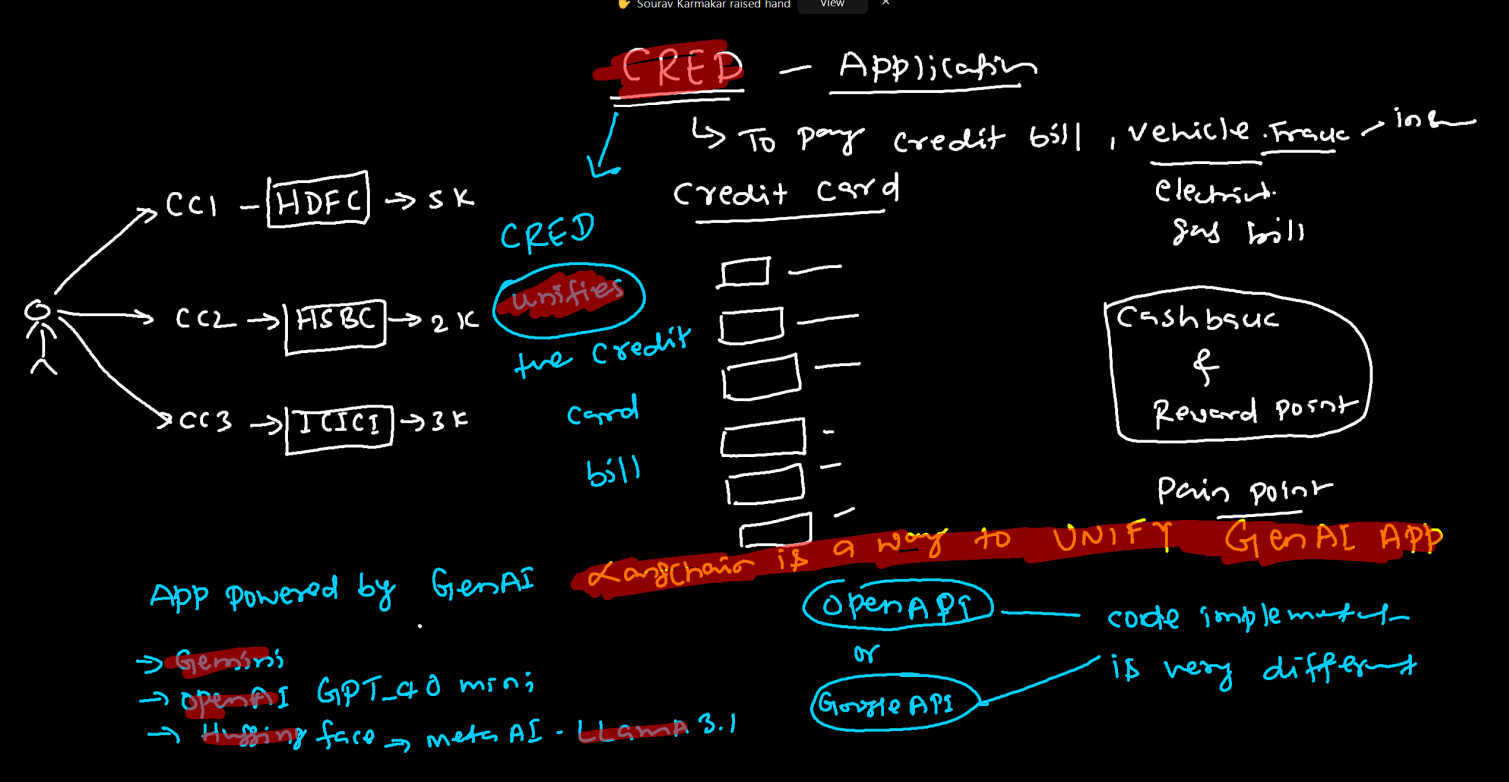
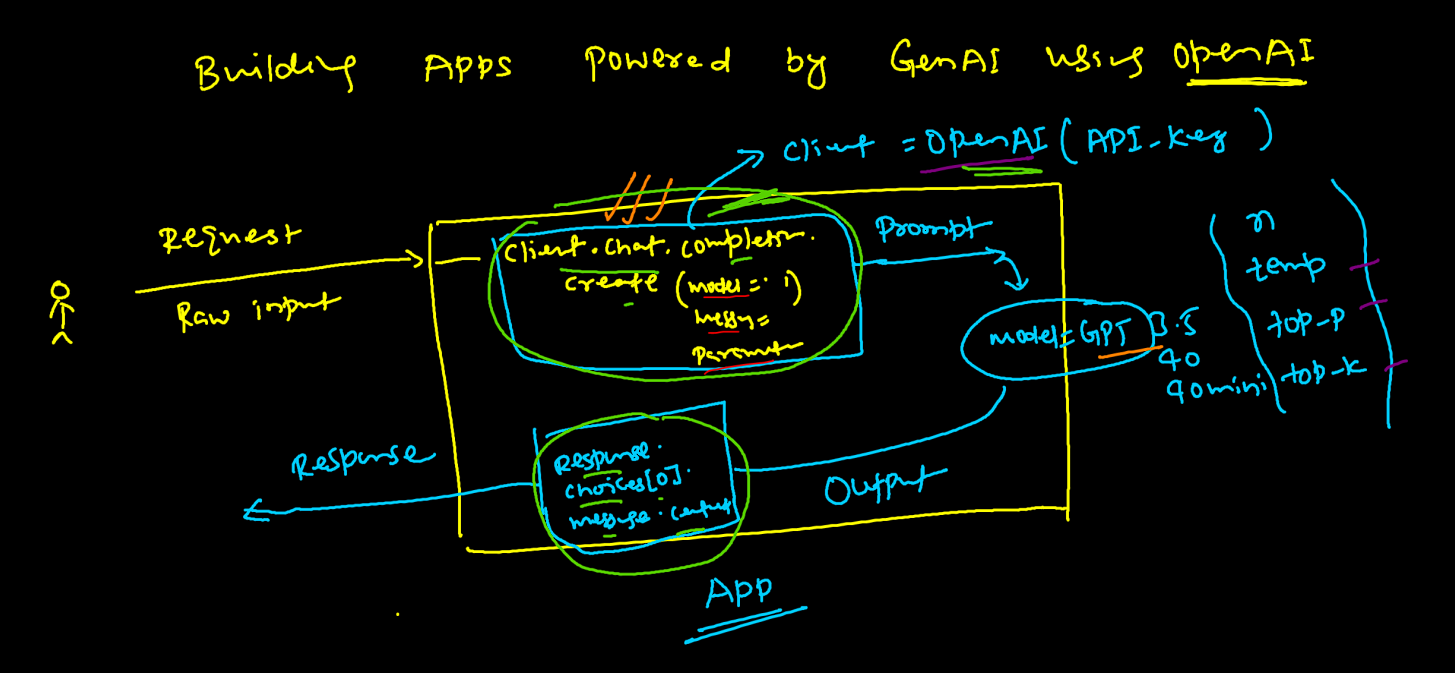
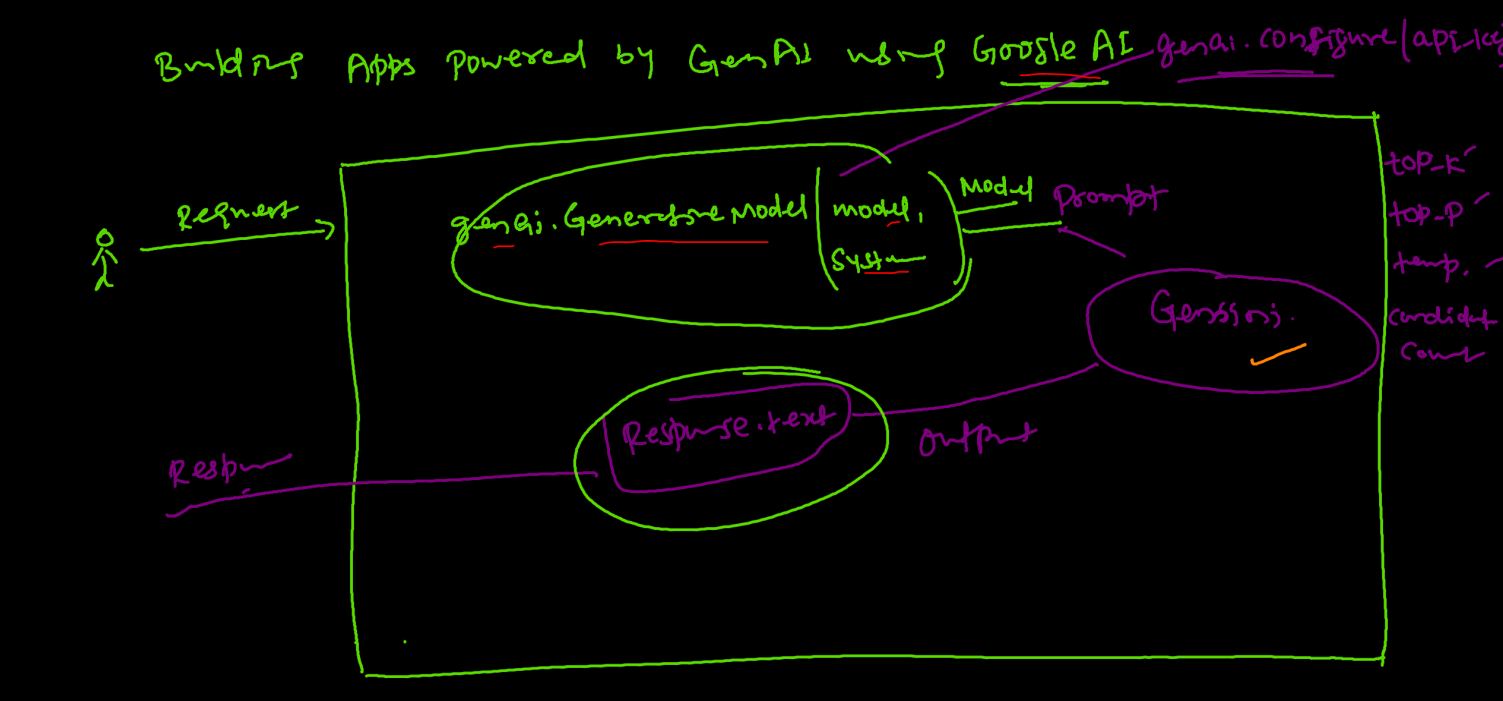
**LangChain**



Langchain is a way to unify the GenAI applications.



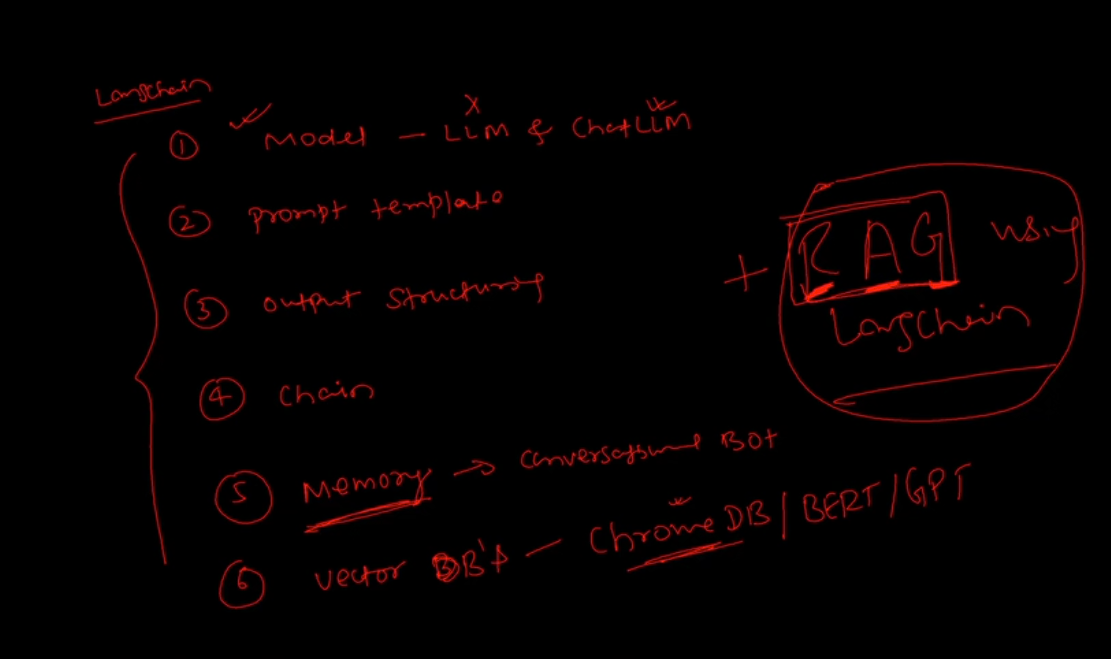


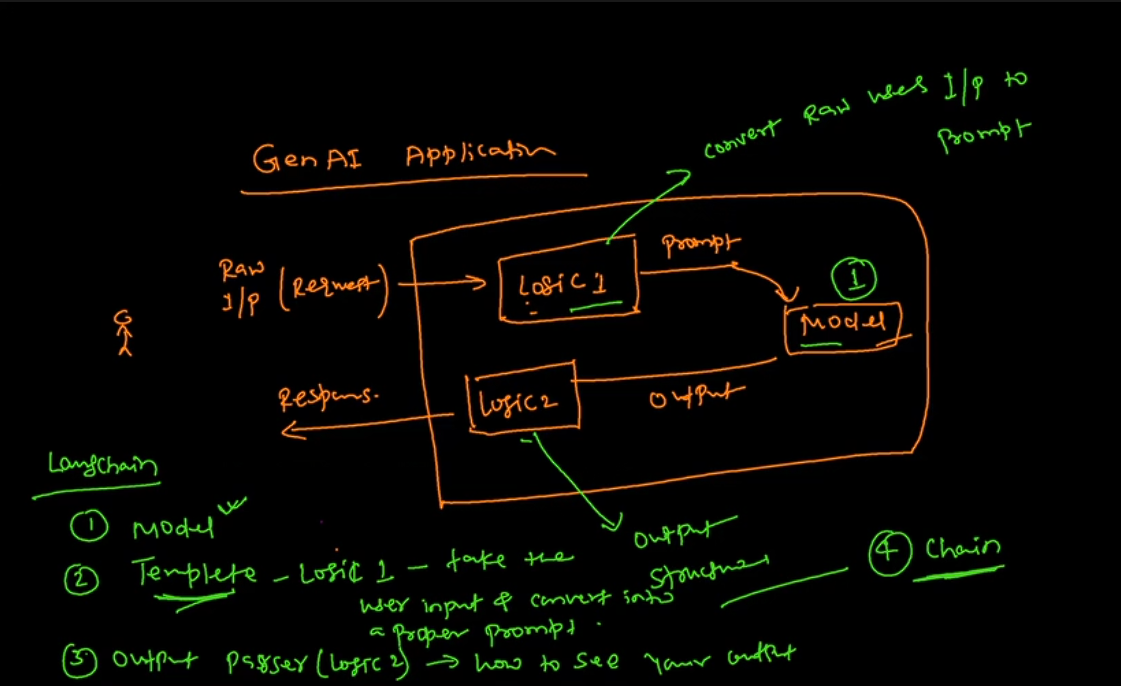


Developing Langchain Model:-

To develop a langchain application or model, we need:-

* Model – Model name that is to be used.
* Prompt Template or Logic 1 – We need to define a template in the beginning itself, example – take the use input and convert into a proper prompt.
* Output parse (Output Structuring) also called Logic 2 – how we want to see your output.
* Then we build a **chain** combining these
* Then we have **Memory** (to build a conversational chat bot)
* Vector DBs – chroma DB, BERT, GPT





The new version of langchain that is version 2 came in march 2024.

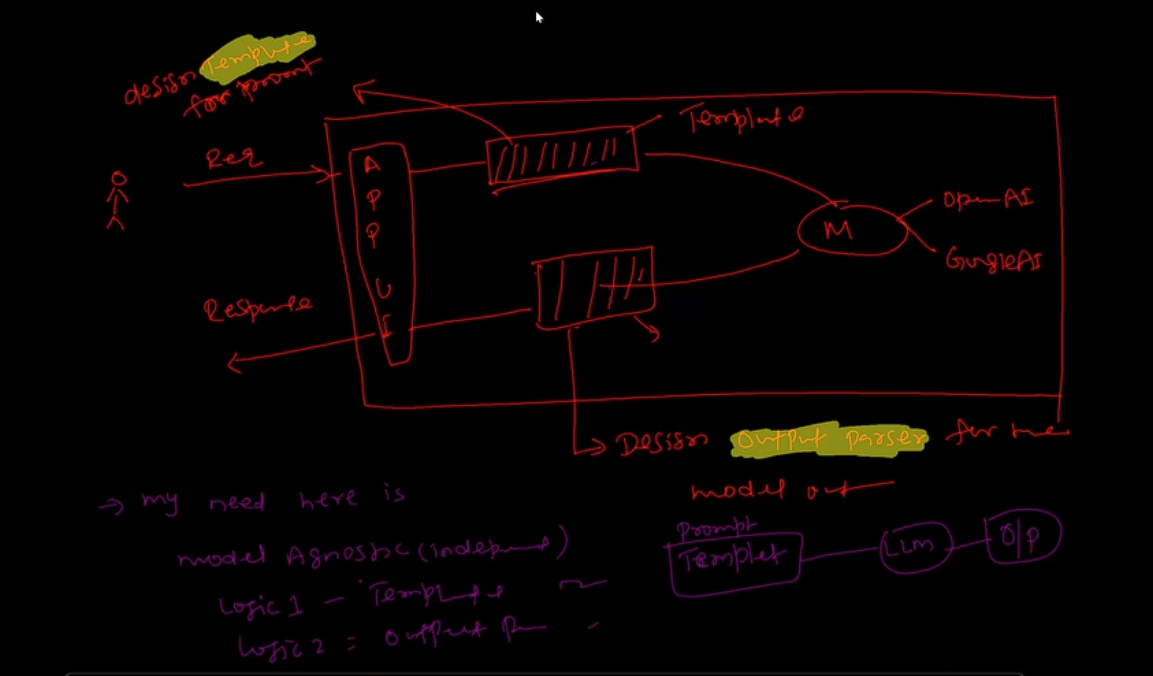
Summary of the Langchain architecture:-

A user sends a request to an APP GUI, that can be Streamlit, Flask, AWS, Azure, etc.,

The request then goes to the Prompt template or logic 1 which is the design **template** (Logic 1) for our prompt. This then goes to our Model.

Then again the output of the model goes to the **output parser** (Logic 2) which the design output parser for our model output. Then we get the response back.

Our need here it to develop an Agnostic (Independent) model

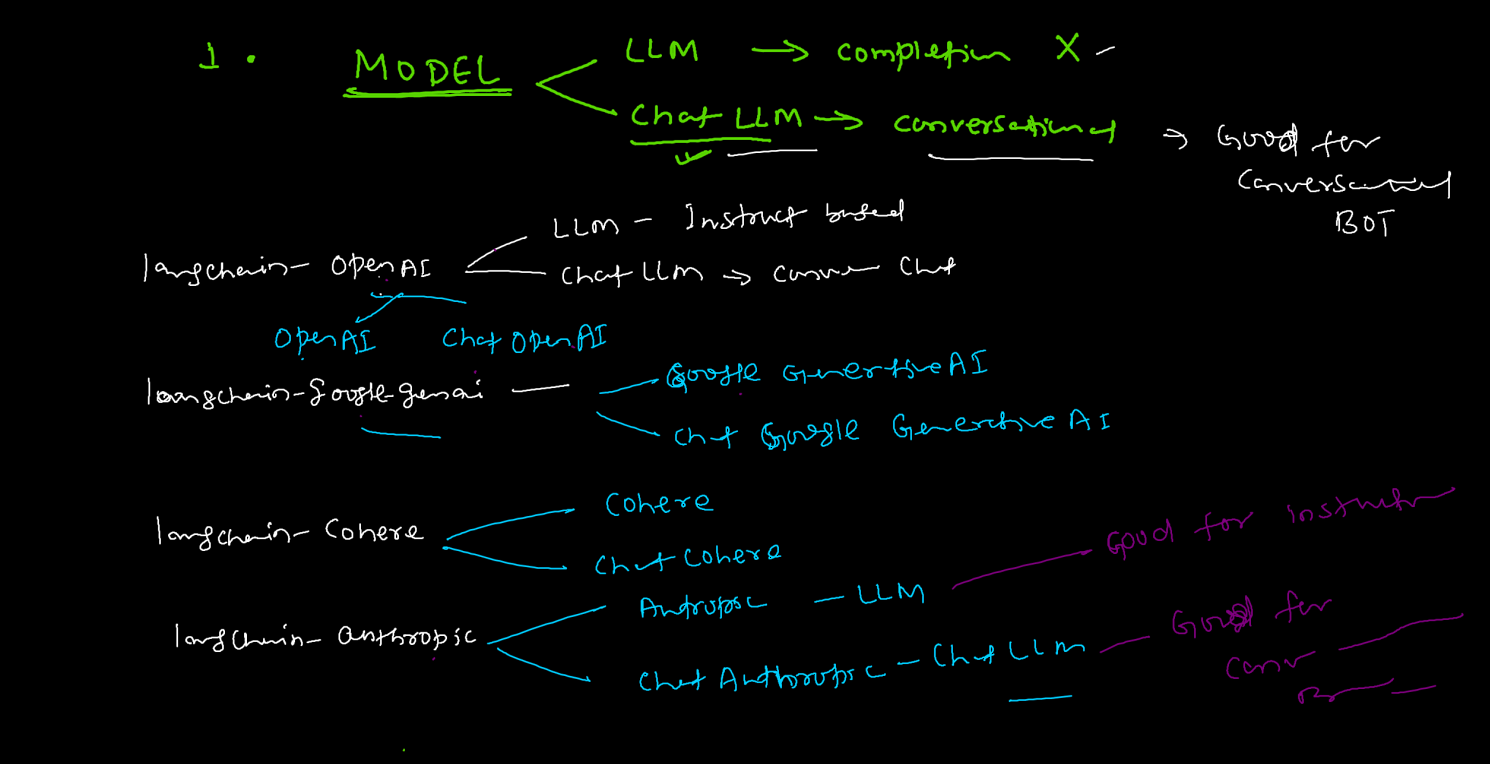


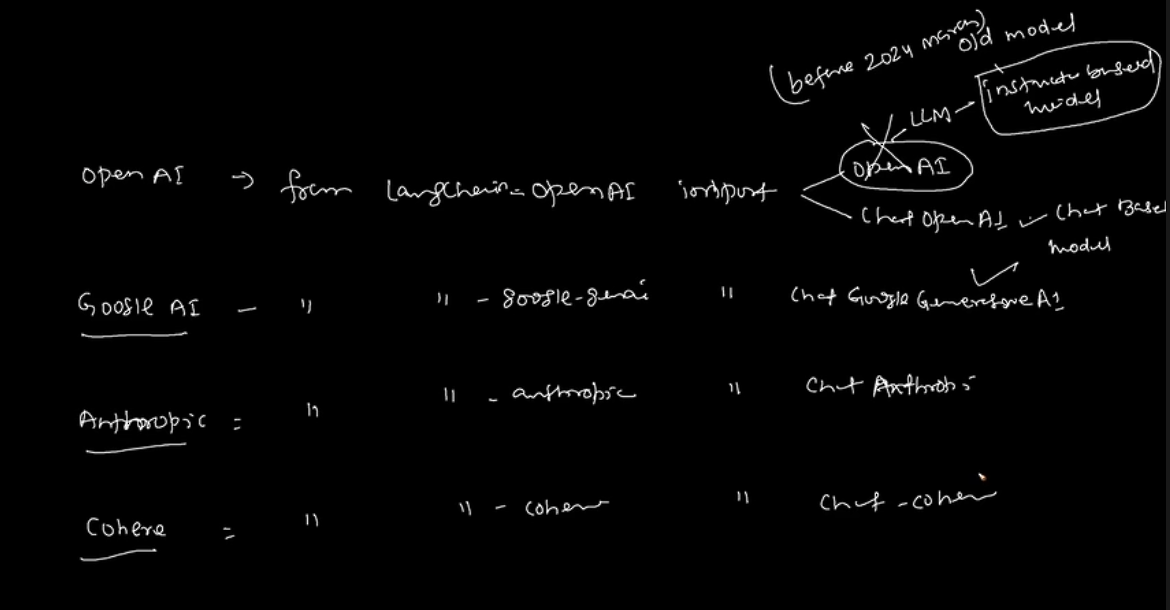
Lets talk about each point in detail:-

1. Model – langchain model has 2 components – LLM and chat based LLM.

The LLM is called Completion model and the chat based LLM is Conversational model. The LLM one is used rarely as it is the instruct based model (fill in the blank kind of problems)

While the chat based LLM is extensively used (conversational AI is in demand)

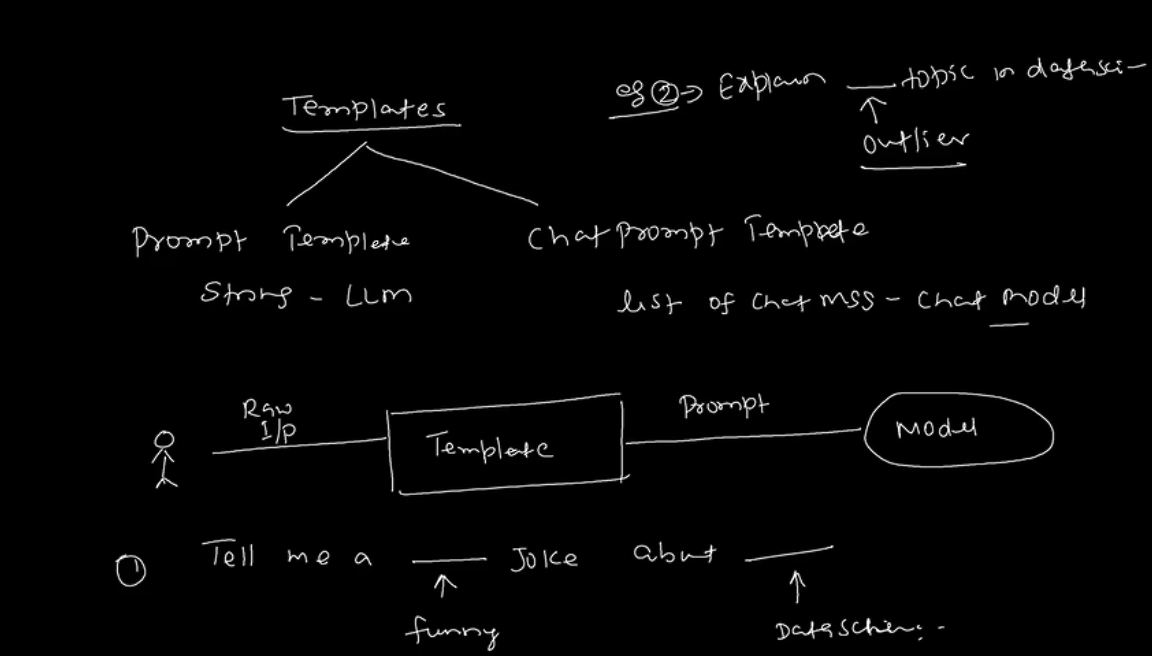


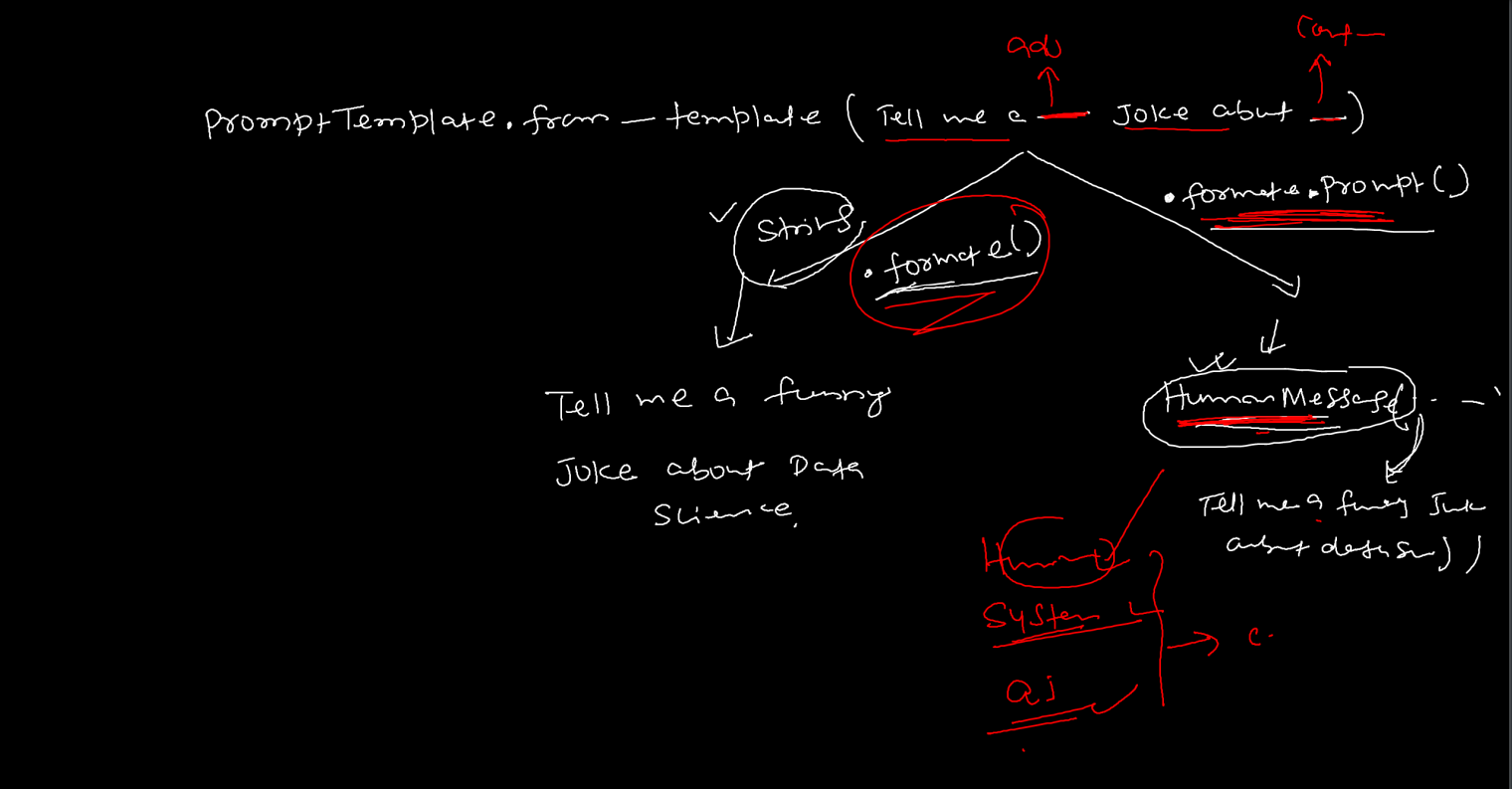


Templates – we have 2 templates,

Prompt template (instruction based model) and chat prompt template (conversational based model)

The prompt template generates the string while the chat prompt template generates the list of chat messages.





The main benefit of using langchain ecosystem is that logic 1 and logic 2 is fixed just we need to change the name of the model so that we can use this chain as per our need. So once we define our logic 1 which is our prompt template and logic 2 which is our output parser, we can just change the name of the model and use it anywhere we like.

Memory:-

The memory is kind of a storage inside the model where the raw input or prompt is stored along with the parsed output generated by the output\_parser.

## the memory or history gets stored in the temporary folder and gets erased once the system is shut. Hence we need to crate a pickle file to store the memory.

