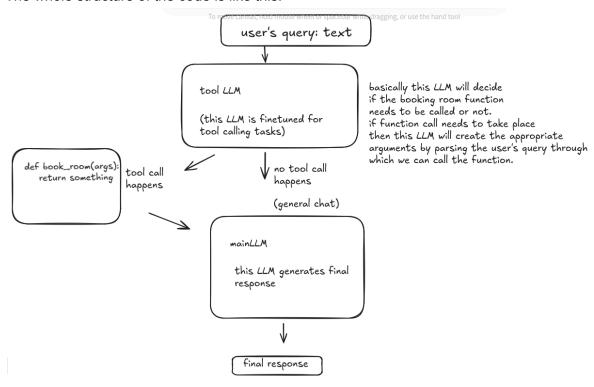
## **Assignment Report (Booking Agent):**

Our objective is to create an LLM integrated with external APIs/functions which can fetch real time data/insights and based on that information and then our LLM can generate responses based on the information. LLMs with this type of flow are known as AI Agents.

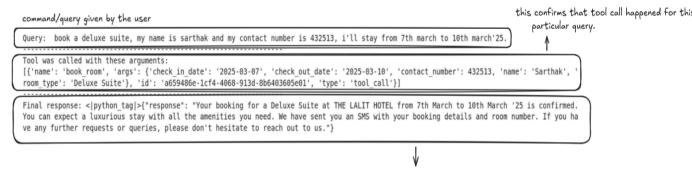
I've used ollama to get the model weights which can be used for inference. Apart from that I've used langchain to integrate our LLM (which is downloaded from ollama), with API/function call, which is commonly known as tool calling.

Also the system prompt given to these LLMs also play a key role in how the final response will look like. These models are prompt sensitive.

The whole structure of the code is like this:



## Output if tool calling happens:



this is the final response which the user will be able to see.

## Output for general chat:

## **Future scope of improvement:**

- Making the LLM more robust by adding the chat history functionality i.e. LLMs now remembers all the previous conversations which have happened with a particular user.
- Currently the tool/function is a dummy function (code snippet attached below for reference), we can connect it with DB and can share the latest information with the user.

```
# Function which will be called if tool calling happens.
@tool
def book_room(name, contact_number, check_in_date, check_out_date,
room_type)-> str:
    """Book a hotel's room on behalf of the user."""
    if(payment_done()):
        # update the DB that a room is booked.
        return f"Thanks {name}. Your {room_type} room has been booked
and details will be shared to your contact number:
{contact_number}."
    return "Payment was not successful. Room wasn't booked."
```

Integration with website or hotel documents via RAG so that even staff can use it.