

# Backend Developer with Generative AI Focus Assignment

## Main Task

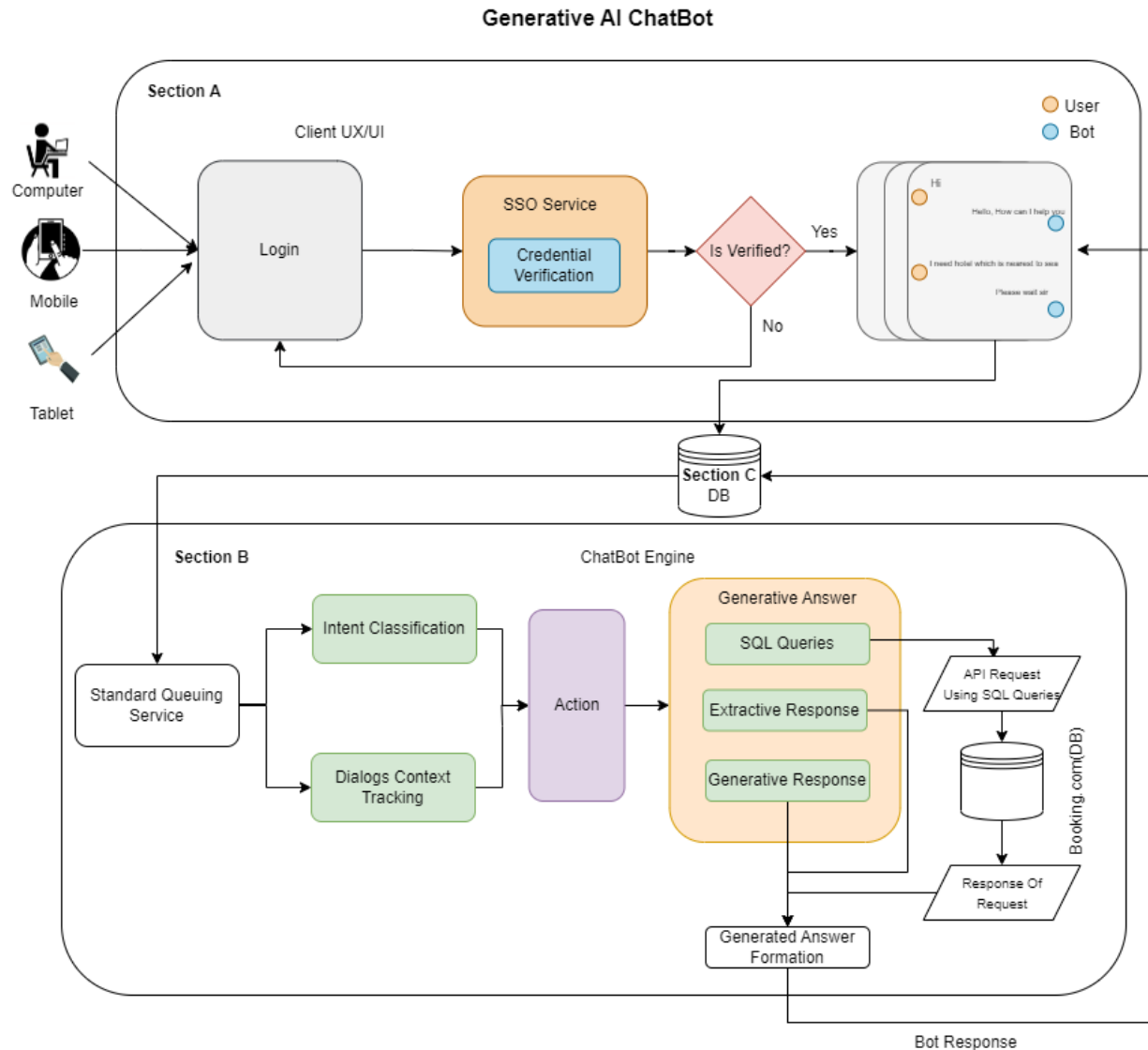


Figure: Generative AI chatbot

In this figure, I try to elaborate on the overall process flow of the chatbot diagram, which are separated into Three sections.

1. Section A is the client server or website for the client.

2. Section B is the chatbot engine or generative AI chatbot system.
3. Section C is the Database of Chatbot System

## **Section A:**

1. Login and authentication system
2. Conversation Page

### **Login and authentication system:**

Basically login system uses for isolate individual user from each other and maintain client privacy. So for the login system here used username, password or SSO system for user verification, this SSO may have "aws cognito".

### **Conversation Page:**

Successful login user can access chat with the generative AI chatbot and get Hotel booking information as well as client require queries response.

## **Section B:**

For the Enterprise Generative AI Chatbot must be hosted into a cloud server. So cloud server can be prepared two different ways.

1. Amazon Lex - Conversational Solutions

2. Customize Architecture – Client are hosted into ECS and Generated AI Chatbot hosted into ECR. So That client and Chatbot communicate with each other using API with Standard queuing service.

Here I will describe 2 Architecture for the understanding a chatbot engine component how they connect to each other and the role.

**Here I have listed all of component:**

1. Intent Detection
2. Entity Extraction
3. Dialogs Context Tracking
4. Action and Generate Answer
  1. Rule based (FAQ)
  2. SQL queries generation and data retrieving from Booking.com database
  3. Extractive Response
  4. Generative Response
5. Generate answer formatting(Validation)

**Intent Detection**

Intent detection in a hotel booking chatbot refers to the ability of the system to accurately understand and interpret the user's intent and extract relevant information from their messages to provide appropriate responses. For the better performance can be used **Transformer** or **Bert** based model for detection the user Intent.

Such as example of intent:

- i. greeting
- ii. thanks
- iii. location
- iv. payments

- v. booking
- vi. cancellation

### **Dialog context tracking:**

Dialog context tracking is an essential aspect of a chatbot's design and functionality. It refers to the ability of the chatbot to maintain and utilize the history of the conversation with a user, including previous messages, user inputs, and system responses.

### **Action and Generate Answer:**

Action and Answer generation depend on intent and entity extraction. The action and answer generation for the chatbot use conversational LLM(Large Language model) as well as some rule based solution.

So for the perspective here try to describe the process of Generative AI chatbot process steps

### **SQL queries generation and data retrieving from Booking.com database:**

SQL queries generation perform when user want to extract some hotel information like price range, near the sea beach or sea view, or give some preference. So Need to generate sql query from the given text, entity extraction and it's Intent. Using the SQL query information from booking.com database.

### **Extractive Response:**

Extractive Response plays a crucial role in the functionality and effectiveness of a chatbot and it also call common sensing of generative model. It is the process of identifying and extracting specific pieces of information or entities from user inputs, such as names, dates, locations, rent of room price, or any other relevant details.

Such as example of extractive response:

**User Input :** Please give some hotel which are range \$150 to \$160 per night and sea view.

**Extractive Response Output is :**

**Room rent range:** \$150 to \$160

**Room type :** sea view

**Generative Response:**

Generative Response feature is a text generation process of the model which are generate from the knowledge graph. If we give a example like chatgpt or other LLM Bard. So here can build a LLM model like chatgpt using prompt and comprehension or fine-tune the chatgpt using custom dataset.

Or if want to build LLM model using open source proven model like "**stanford\_alpaca**" it give better performance like chatgpt3.5.

Link of codebase : [https://github.com/tatsu-lab/stanford\\_alpaca](https://github.com/tatsu-lab/stanford_alpaca)

**Generate answer formatting(Validation):**

Generate answer format functionality means maintain the coherence given the input and response output. This is the importance things because if the response are not coherent of the input that is called like randomness. So this type of response need to penalized so the it can be perform better given by the feedback or rewarding. So this method I want about talk about reinforcement learning which technique use in ChatGPT.

**Section C**

**Database:** Used the database for store the user and bot conversation. So that in the future it can be use for build SFT(supervised fine-tune) model and then using the dataset try to improve Chabot performance.

## **Summary of the discussion:**

In the discussion we have some confusion and conflict about the Chatbot Engine with other generative models. Basically, Generative AI Chatbot is not a Chatbot engine but this is the brain of chatbot. So the initial pre-train generative model was trained a large amount of unlabeled data and then it would fine-tune the SFT(supervised fine-tune) model of the label data. And the last one uses a reinforcement learning model for rewarding the model prediction to get a more generalized model.

So, in the generative answer section, it is like a large language model which was trained in prompt and comprehension.

For the chatbot purpose here, some listed LLM Model or open source Chatbot library give support for building chat bots for closed domain purposes like:

1. RASA
2. Dialogflow
3. ChatGPT-[Fine-tune Custom dataset]
4. Bard model