

Interstellar Intelligence with ReX

Group

Starship Innovators - P4

Starship Innovators

Agenda:	3
Text-Based Interview Chat Bot:	4
User's Interface	5
User Interface Design	6
Visibility of system status	8
Matching between system and the real world	9
Consistency and standards	11
Error prevention	12
Recognition rather than recall	13
Help users recognize, diagnose, and recover from errors	14
Dataset with question and suggested answers	15
PDF Reader	16
Extracting user's input	17
Processing input	18
Agent's feedback	19
Question modules:	20
Situational Judgment Module	21
Technical Question Module	22
Behavior Question Module	23
STAR Question Module	24
Assessments:	25

Agenda:

1. Create a text-based chat bot that will ask a user an interview questions
2. Separate questions by different modules
 - a. Situational Judgment Module
 - b. Technical Questions Module
 - c. Behavior Questions Module
3. Create additional practices for an user
 - a. Concept Explanation Practice
 - b. Team Collaboration Scenario
4. Create Assessment Section
 - a. Personality Assessment
 - b. Communication Skill Assessment
5. Tip Section

Text-Based Interview Chat Bot

To create and implement a chat bot, we are breaking down the process of creation into different sections:

1. User Interface
2. Dataset with question and suggested answers
3. Extracting user's input
4. Processing that input and send to the agent
5. Based on that input, let agent to generate feedback

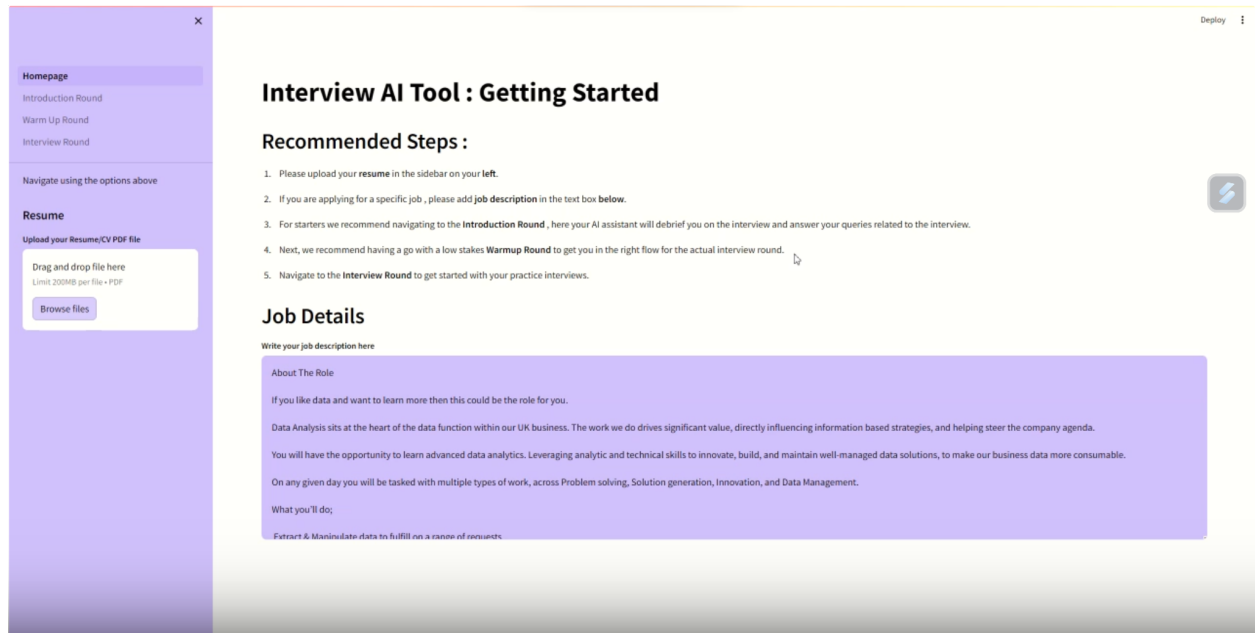
User's Interface

User's Interface is a crucial part of the project. It should contain a usable hierarchy, which is shown below.

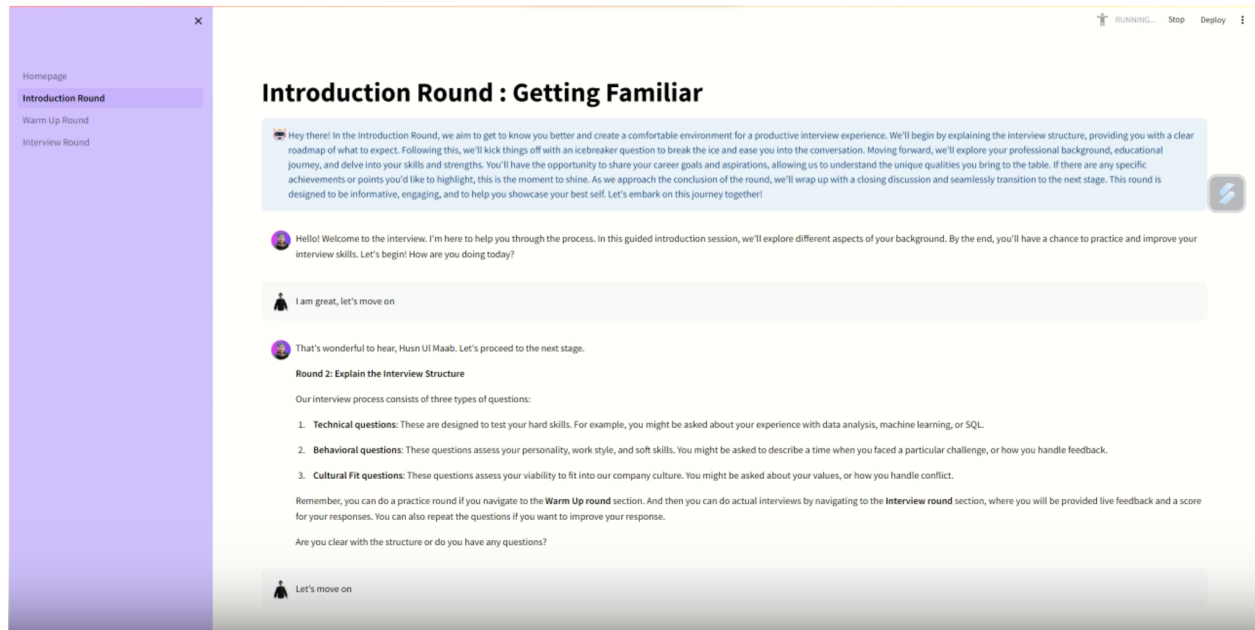
1. User Interface Design
2. Visibility of system status
3. Matching between system and the real world
4. User control and freedom
5. Consistency and standards
6. Error prevention
7. Recognition rather than recall
8. Flexibility and efficiency of use
9. Aesthetic and minimalist design
10. Help user recognize, diagnose, and recover from errors
11. Help and documentation

User Interface Design

Homepage



Introduction Round



Starship Innovators

Warmup Round

×

Homepage

Introduction Round

Warm Up Round

Interview Round

Deploy

👤

Alright, Mr. UI Maab, let's move on to the next question. Given your extensive experience in machine learning and model evaluation, I believe you'll be able to provide a detailed response to this one.

Question:

How do you handle imbalanced datasets in machine learning, and why is it important?

Please share your strategies and the reasoning behind them.

👤

It depends on the specific task. I mostly use SMOTE , to upsample my dataset by inserting artificial samples. But in some cases preserving the integrity of the data is more important, so I might go with undersampling if the data is of reasonable size.

👤

Assessment:

1. Clarity and Conciseness: 8/10

The answer is clear and concise, but could use a bit more explanation on what SMOTE is and why it's used.

2. Tone: 10/10

The tone is professional and formal.

3. Intent Satisfaction: 15/20

The answer addresses the question, but doesn't fully explain why handling imbalanced datasets is important.

4. Depth: 10/15

The response provides strategies used, but lacks depth in explaining the reasoning behind them.

5. Keyword Usage: 7/10

The user mentions key terms like SMOTE, upsample, and undersample, but doesn't explain them.

6. Curiosity: 10/10

The user shows curiosity and willingness to adapt their approach based on the task at hand.

7. Length: 12/15

The answer is of reasonable length, but could be more detailed.

Total Score: 72/100

Feedback:

Your answer is good, but could use more depth. Try to explain why you use specific strategies and why they are effective. Also, explaining key terms like SMOTE, upsample, and undersample would make your answer more comprehensive. Lastly, don't forget to address why handling imbalanced datasets is important.

👤

Thank you for your detailed response, Mr. UI Maab. Your understanding of handling imbalanced datasets is commendable. Let's move on to another important aspect of data analysis.

Type here

➔

Interview Round

×

Homepage

Introduction Round

Warm Up Round

Interview Round

Deploy

Interview Round : Skill Up Your Interviewing Prowess

Which type of questions do you want to practice?

Choose an option

Technical

Behavioural

Culture Fit

STAR

🏃 RUNNING...

Stop

Deploy

Interview Round : Skill Up Your Interviewing Prowess

Which type of questions do you want to practice?

Technical

👤

Hello and welcome! I'm glad to have you here for the data analyst position interview. Let's dive right in, shall we?

Could you describe a situation where you applied clustering techniques in a data analysis project? This could be any project where you used methods such as k-means clustering to group similar data points together.

👤

I have used clustering in market trend analysis for customer segmentation

👤

Clarity of the Answer: 8/10

Your response is clear and easy to understand. However, you could have provided more details about the specific clustering technique you used.

Tone of the Answer: 10/10

Your tone is professional and formal, which is appropriate for an interview setting.

Satisfaction of the Question Intent: 10/20

While you did mention that you used clustering for customer segmentation in market trend analysis, you did not describe a specific situation or project, which was the main intent of the question.

Depth of the Answer: 7/15

Your answer lacks depth. You could have elaborated more on how you applied the clustering technique, what challenges you faced, and what results you achieved.

Use of Keywords: 8/10

You used some relevant keywords such as 'clustering', 'market trend analysis', and 'customer segmentation'. However, mentioning the specific clustering technique used would have been beneficial.

Length of the Answer: 7/15

Your response was quite brief. An interview response typically requires more detail to fully demonstrate your skills and experience.

Total Score: 50/80

Tips for Improvement:

To improve your response, try to provide a specific example of a project where you used clustering. Discuss the clustering technique you used, why you chose it, how you implemented it, and what the results were. This will not only satisfy the intent of the question but also demonstrate your depth of knowledge and experience.

Which question would you like to do?

Next

Repeat

Interstellar Intelligence Implementation

Visibility of system status

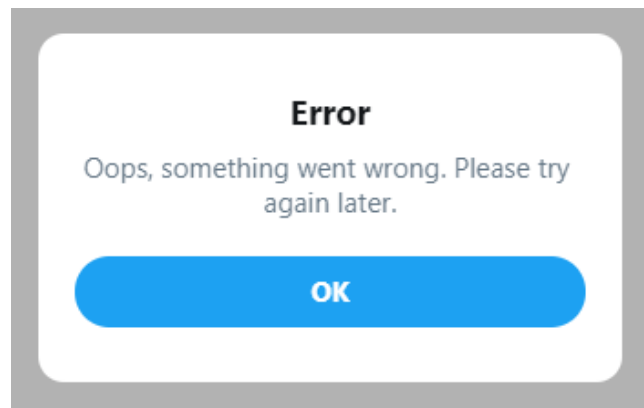
The design should always keep users informed about what is going on, through the appropriate feedback within a reasonable amount of time (ideally - immediately)

When users know the current system status, they learn the outcome of their previous interactions and determine next steps.

In our case, with the chat bot, it should be clear for a user to understand if:

First, the bot has taken the answer and is processing it.

Second, if some errors occur, the state and nature of an error is displayed.



Matching between system and the real world

The design should speak the user's language. We have to follow the real-world conventions, making information appear in a natural and logical order. That way, it is easier for a user to navigate and remember how the interface works.

In our case, we have to use icons that can be corresponded to real-world objects.

In our case:

The question that corresponds to the bot, should have a robot icon in front, when the user's response should have a human icon. That way, it is easier for a user, to separate and distinguish the correspondence of the lines in a dialog, especially, in a chat based.



Can you tell me a bit about your experience with data analysis, specifically with what tools and programming languages you've worked with?



I have graduated from NYU as a system analyst

User control and freedom

As it always happens, users tend to make mistakes. They need a way to undo their actions, the way it is not going to give them a headache in the future. In other words, when it's easy for people to back out of a process or undo an action, it gives them a sense of freedom and confidence. Exits allow users to remain in control of the system and avoid getting stuck and feeling frustrated.

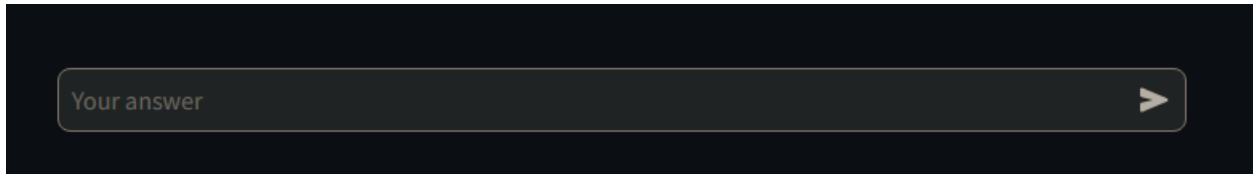
In our bot design, users should have a “back” button, in case they want to either leave an existing conversation, or if they entered the wrong, unintended module.



Consistency and standards

Users should not have to wonder whether different words, situations, or actions mean the same thing.

The main features, that should be clear for our users, are the fields where the questions and responses will be displayed, as well as the field where the user will have to type and answer.

A dark-themed input field with a light gray border. Inside the field, the text "Your answer" is displayed in a light gray font. On the right side of the field, there is a light gray right-pointing arrow icon.

Error prevention

Good error messages are important. If the error occurs, we should be able to present that to the user with the confirmation option before they commit to the action.

Errors should be prevented in the first place, by handling them with helpful constraints and good defaults.

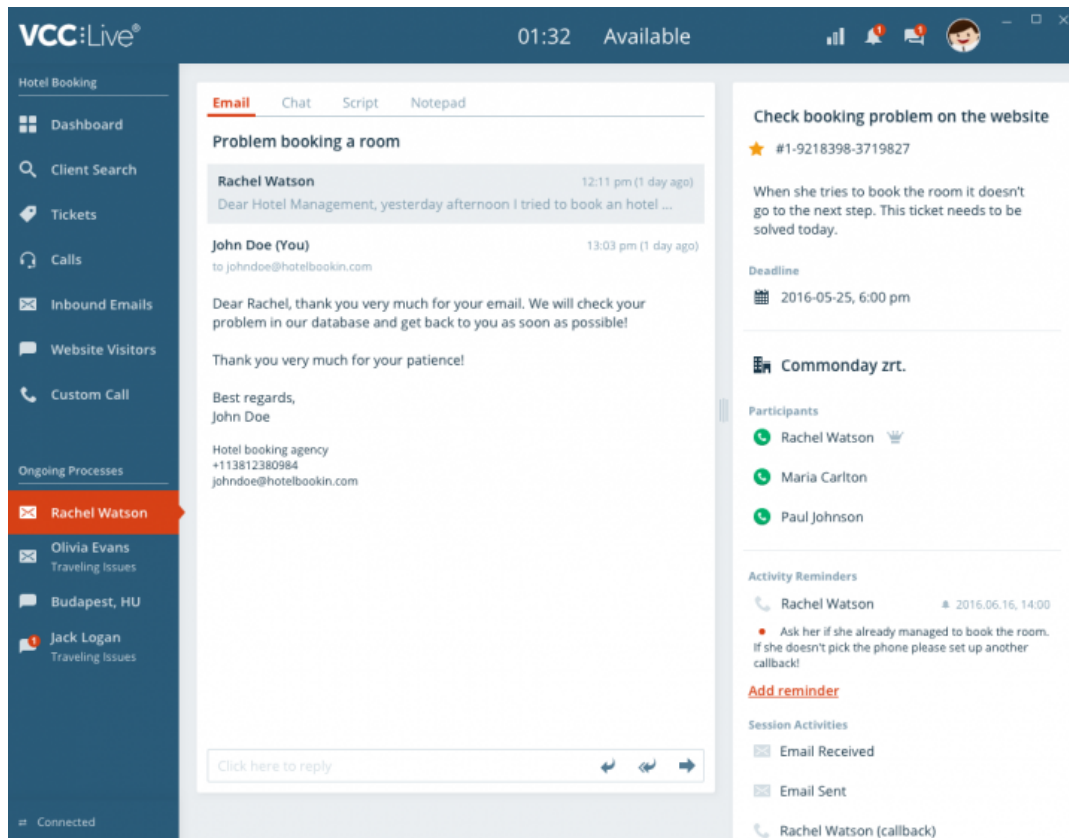
```
# Check for potential errors
if response.status_code != 200:
    raise ValueError(f"Error from OpenAI API: {response.content}")
```

Recognition rather than recall

We have to minimize the user's memory load by making elements, actions, and options visible. The user should not have to remember the information from one part of the interface to another. The information should be visible or easily retrievable when needed.

We should have to implement not only the easy access for each module, but also a smooth transition from one to another, or to the main menu.

Here is an example:

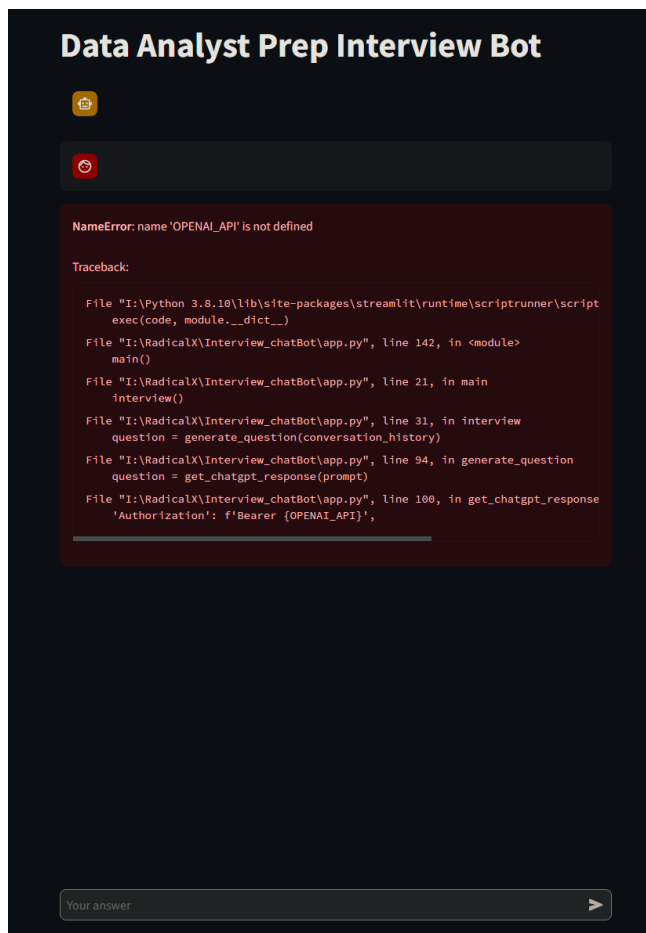


Help users recognize, diagnose, and recover from errors

Error messages should be expressed in plain language (no error codes), precisely indicate the problem, and constructively suggest a solution.

Important note: The error should not crash/stop users from using the service. The user should still remain on the page.

This should not be the case ever:



Dataset with question and suggested answers

The dataset used will be made using a hybrid approach of adding questions and answers through research and asking Generative AI models to generate questions for Data Analytics Interview and will then be stored in an appropriate format. The questions will be categorized into various categories such as technical questions related to hard skills, behavioral questions, and cultural fit assessment questions. An example dataset might look something like this :

```
{ "question": "What is the best way to handle null values in SQL?" }  
{ "question": "How can I secure my SQL database?" }  
{ "question": "Can you explain the concept of normal forms in SQL?" }  
{ "question": "How can I use Python with Big Data frameworks like Hadoop or Spark?" }  
{ "question": "What are some advanced data visualization techniques in Python?" }  
{ "question": "What are some complex SQL queries I should practice?" }  
{ "question": "Can you explain the differences between different SQL dialects, like MySQL, PostgreSQL, and SQL Server?" }  
{ "question": "What are some advanced SQL concepts I should understand deeply?" }  
{ "question": "Can you give me some tips on optimizing SQL queries for large databases?" }  
{ "question": "How can I ensure my SQL queries are secure and prevent SQL injection?" }  
{ "question": "What are some real-world scenarios where complex SQL queries are used?" }  
{ "question": "Can you explain how data warehousing works with SQL?" }  
{ "question": "How can I use SQL for data analysis and reporting?" }  
{ "question": "What are the most efficient ways to use indexing in SQL?" }  
{ "question": "Can you explain transaction control statements in SQL?" }
```

PDF Reader

For a better user's experience, we have included the ability to upload a resume. All information regarding the user's skills, past-experience, certificates, major and attended schools will be stored and used. That way, the Interview Prep Bot will be more personalized to each user independently, and adjust questions and behavior in favor of that.

To extract the information from the resume, we are converting the pdf into the plain text. After that we are dividing it into chunks, and feed those chunks with embedding into the langchain vectorsotre. Then we use similarity search using chains to store the necessary information.

```
if resume:
    pdf = PdfReader(resume)

    text = ""
    for page in pdf.pages:
        text += page.extract_text()

    text_splitter = CharacterTextSplitter(
        separator="\n",
        chunk_size=1000,
        chunk_overlap=200,
        length_function=len
    )
    chunks = text_splitter.split_text(text)
    embeddings = OpenAIEmbeddings()
    doc = FAISS.from_texts(chunks, embeddings)

    chain = load_qa_chain(llm, chain_type="stuff")

    name = chain.run(input_documents=doc.similarity_search("What is the person's name?"), question="What is the person's name")
    exp = chain.run(input_documents=doc.similarity_search("What is the professional experience?"), question="What is the professional experience?")
    skills = chain.run(input_documents=doc.similarity_search("What are the person's skills?"), question="What are the person's skills?")
    certs = chain.run(input_documents=doc.similarity_search("What is the person's courses/certifications?"), question="What is the person's courses/certifications?")
    projects = chain.run(input_documents=doc.similarity_search("What are the person's projects"), question="What are the person's projects")

    resume_info = {
        "Name": name,
        "Experience": exp,
        "Skills": skills,
        "Certifications": certs,
        "Projects": projects
    }

    st.session_state["Resume Info"] = resume_info
```


Extracting user's input

By extracting the user's input, we can send it to the agent as a *{response}*. By doing so, we are providing the LLM with some data, with which it first can work, and second, it can analyze it, give feedback, and generate the next question. The most important aspect of that, is the fact that the question will be generated based on the user's response, and the role of the LLM agent.

```
message.write(question)
if answer:
    user_response.write(f"{answer}")
```

Processing input

To process user's input we will use the Langchain framework. It will be connected to an OpenAi language model - either gpt-3.5-turbo, or gpt-4.

Using a PromptTemplate, the agent will be able to act and respond as we tell him.

The main concept is to get an agent to a state, where it will satisfy our needs. First, by consistency, and second, by the need. For example, based on the question that is assigned to a specific module, the agent will correlate to that role, and behave accordingly.

```
✓ choose_template = """ Based on the action : {action} , choose a question and corresponding intent from the the dataset : {data}.

If action is "Next" , the chosen question should NOT be the same as the ones in this list of questions: {questions}. One question should not be asked more than once.
If the action is "Repeat", just provide the last question in the list of questions again. Do not choose a new question in this case.
Try to make the questions relevant to the user's details : {details} and job description : {description}

Your response should be like :

Question : <<<chosen question, that should be asked , do not mention the question that shouldn't be asked here>>>
Intent : <<<the intent provided alongside the question in the dataset, complete with all the keywords and everything>>>
Logic : <<< Provide the logic of choosing the question based on the action , the previous questions and the relevant job and user details>>>

You MUST choose a question, and provide intent from dataset for it.
"""
```

Agent's feedback

The same concept will be applied to another agent, that will only be focusing on providing feedback to the user, based on his response. However, an extra layer of caution will be taken, since we need consistency from the agent. So, the agent will compare the user's response to the responses that are given in a dataset. After that, it will evaluate and give non-bios, and precise feedback.

```
interview_feedback_template = """Provide an critical assessment to the response : {response} to the question : {question} of an
interview.
You can use the following steps to guide your final assessment :

1. Is the answer in clear, concise wording?
2. Is the a good attempt to answer the question?
2. Is the tone of the answer professional and formal?
3. Does the answer satisfy the intent of the question?
4. Is the response of sufficient depth?
5. Does the user mention the right key words in their answer
6. Does the answer show curiosity to learn if the answer is unknown?
7. Is the answer of a sufficient length?

Your response MUST contain BOTH :

1. First give final assessments explaining how the response can be improved.
2. Then provide an example improved response
3. Be sure to highlight the most important parts

Give the response a score out of 100. Be critical. Do not be afraid to give 0 if the candidate's response does not
satisfy the intent, is not even related to the question. Answers that are too abrupt or short should also be marked
poorly.

Before giving the feedback , make sure that the response is relevant to the question. If it seems like the user gave an ambiguous response, or it seems.
like the user did not understand the question. For example , "I don't know" , "Not Sure" , and other responses like these are not to be considered proper responses.
Tell the user you are not sure if the user understood the question correctly , and ask them to choose the repeat option if that
is the case.
| | | | | """
```

Question modules

The conversational model will consist of a questioning and feedback loop , using agents and tools from frameworks such as Langchain that will be prompt engineered to achieve a desired behavior. As an additional functionality , the tool can use a Resume reading feature , along with a job description to extract appropriate questions from the dataset, to tailor the experience for the user.

There will be 3 main question modules, based on which, the user will be asked a question.

1. Situational Judgment Module
2. Technical Questions Module
3. Behavior Questions Module

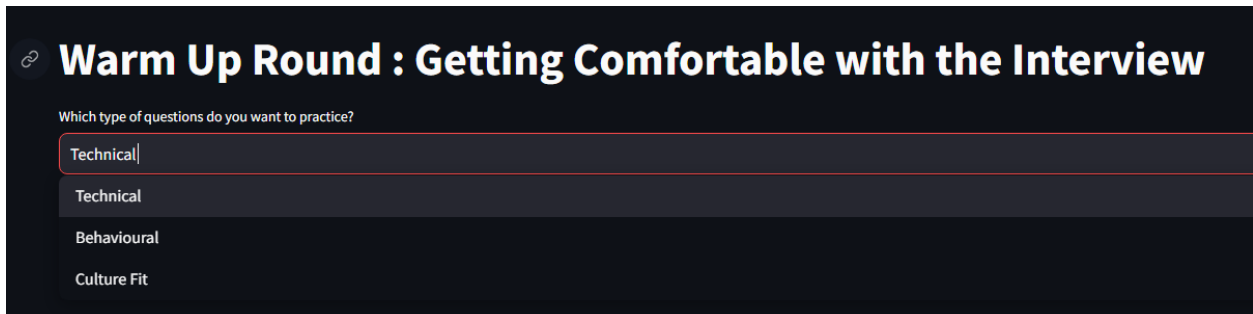
For each module, we will have its own dataset of questions and possible answers.

Situational Judgment Module

As for all question types, we are having the exact same prompt, to extract and represent them. However, we have a separate, pre-build json file with pre-existing questions with the ideal answers as a kick off.

```
{  
  "question": "How do you contribute to creating a positive and collaborative team environment?",  
  "type": "Cultural Fit",  
  "answer/intent": "The intent behind this question is to assess the candidate's teamwork and collaboration skills. Look for keywords such as positive team environment, collaboration, and open communication."  
}
```

The user has an ability to choose the desired type of question through the interface.



Warm Up Round : Getting Comfortable with the Interview

Which type of questions do you want to practice?

- Technical
- Technical
- Behavioural
- Culture Fit

Technical Question Module

As for Technical Questions, we follow the same procedure with the pre-existing dataset of technical questions with the ideal answers. The user response would be compared with the ideal answer. Based on that, the user will see the feedback.

```
{
  "question": "Explain the concept of 'p-value' in statistics and its significance in hypothesis testing.",
  "type": "Technical",
  "answer/intent": "The p-value is the probability of obtaining results as extreme or more extreme than the observed results under the assumption that the null hypothesis is true. A lower p-value suggests stronger evidence against the null hypothesis."
},
{
  "question": "Describe a situation where you applied clustering techniques in a data analysis project.",
  "type": "Technical",
  "answer/intent": "I applied clustering to group similar data points together based on certain features. For example, in customer segmentation, I used k-means clustering to identify distinct customer groups. Keywords: clustering techniques, grouping, k-mean"
},
{
  "question": "How do you handle imbalanced datasets in machine learning, and why is it important?",
  "type": "Technical",
  "answer/intent": "I handle imbalanced datasets by using techniques like oversampling, undersampling, or using algorithms that handle imbalanced classes. It's important because imbalanced datasets can lead to biased models, and these techniques help in achiv
```

Behavior Question Module

Behavior questions are more like general questions that can be asked to any IT specialties.

```
{
  "question": "How do you approach explaining complex data findings to non-technical stakeholders?",
  "type": "Behavioral",
  "answer/intent": "The intent behind this question is to evaluate the candidate's communication skills and ability to convey complex information to non-technical stakeholders. Look for keywords such as explain complex data, visualizations, storytelling, and"
}
```

```
{
  "question": "Tell me about a time when you had to prioritize tasks in a data analysis project to meet tight deadlines.",
  "type": "Behavioral",
  "answer/intent": "The intent behind this question is to assess the candidate's ability to prioritize tasks effectively in a time-sensitive environment. Look for keywords such as prioritize tasks, impact on goals, and tight deadlines."
}
```

Based on the user's choice, which is selected in the dropdown menu. That choice will be passed into the langchain chain, specifying what is the user's intention for the interview prep rounds. And what type of questions he wants to practice.

```
▼ choose_template = """ Based on the action : {action} , choose a question and corresponding intent from the the dataset : {data}.

If action is "Next" , the chosen question should NOT be the same as the ones in this list of questions: {questions}. One question should not be asked more than once.
If the action is "Repeat", just provide the last question in the list of questions again. Do not choose a new question in this case.
Try to make the questions relevant to the user's details : {details} and job description : {description}

Your response should be like :

Question : <<<chosen question, that should be asked , do not mention the question that shouldn't be asked here>>>
Intent : <<<the intent provided alongside the question in the dataset, complete with all the keywords and everything>>>
Logic : <<< Provide the logic of choosing the question based on the action , the previous questions and the relevant job and user details>>>

You MUST choose a question, and provide intent from dataset for it.
"""
```

STAR Question Module

In a STAR question module, users will have to respond to a behavior-based interview question by discussing the specific situation, task, action, and result of that particular situation.

```
{
  {
    "question": "Using the STAR format, describe a specific situation where you successfully implemented a machine learning model to solve a business problem. Outline the key tasks you undertook, actions you implemented, and the measurable results achieved.",
    "type": "STAR",
    "answer/intent": "This STAR question assesses the candidate's hands-on experience with machine learning in a business context. Look for responses that clearly present the Situation, Task, Action, and Result, highlighting their role, specific tasks performed, actions taken, and the measurable impact of the project.",
  },
  {
    "question": "In a previous role, share a situation where you had to present complex data findings to a non-technical audience. Utilizing the STAR format, detail the specific tasks involved, actions you took, and the overall impact on stakeholder understanding and decision-making.",
    "type": "STAR",
    "answer/intent": "The intent behind this STAR question is to evaluate the candidate's communication skills and ability to convey complex information. Look for responses that structure their answer using the Situation, Task, Action, and Result framework, emphasizing the challenges faced, specific tasks performed, and the impact on stakeholders.",
  },
  {
    "question": "Describe a challenging data preprocessing task you encountered in a data analytics project, applying the STAR format. Outline the steps you took to address the challenge, your role in the task, and the impact on the overall data analysis.",
    "type": "STAR",
    "answer/intent": "This STAR question aims to assess the candidate's data preprocessing skills and problem-solving approach. Look for responses that clearly delineate the Situation, Task, Action, and Result, providing insights into the specific challenges faced, the candidate's role, actions taken to address the challenge, and the impact on the data analysis.",
  },
  {
    "question": "Using the STAR format, describe a scenario where you had to lead a data analytics project. Outline the key tasks you undertook, actions you implemented to ensure project success, and the measurable results achieved.",
    "type": "STAR",
    "answer/intent": "This STAR question evaluates the candidate's project leadership skills in a data analytics context. Look for responses that follow the Situation, Task, Action, and Result structure, highlighting the candidate's leadership role, specific tasks performed, actions taken to ensure project success, and the measurable results achieved.",
  },
  {
    "question": "Share an instance where you identified a critical error in a dataset during a data analysis project. Applying the STAR format, detail the Situation, the specific Tasks you performed to rectify the error, Actions you took, and the overall impact on the project outcomes.",
    "type": "STAR",
    "answer/intent": "This STAR question assesses the candidate's attention to detail and problem-solving skills when dealing with data errors. Look for responses that present a clear Situation, Task, Action, and Result, outlining the specific error identified, tasks performed to rectify it, actions taken, and the impact on the project outcomes.",
  }
}
```


Assessments

The user can then be asked to carry out a proper interview where they'll be asked questions from various categories and they'll be provided feedback and tips based on their responses, and how they can better tailor their responses to answer the interviewer's questions based on the intent and requirements. This way, the user's performance can be tracked and the user can see the areas they need to work on in order to improve their responses.