	<b>CSE 299: Junior Design Project</b>		
	<b>Department of Electrical and Computer Engineering</b>		
	<b>Semester:</b>	Spring 2025	<b>Section:</b> 18
	<b>Weekly Progress:</b>	1	<b>Date:</b> 31 <sup>st</sup> Jan, 2025

<b>Faculty Name:</b>	Dr. Shafin Rahman
<b>Group Number and Project Title</b>	Group # 7: Project Title: Bank Customer ChatBot
<b>Student Name (ID)</b>	Soleman Hossain 2021682042
<b>Email Address:</b>	soleman.hossain@northsouth.edu

<b>Last week's Target</b>	<b>Follow-up</b>
<ol style="list-style-type: none"> <li>1. Get to know the basic structure of Chatbot's pipeline. Also, how does RAG work?</li> <li>2. We targeted the requirements like the packages, vector database, programming language, dependencies, and versions that we will need for our ChatBot</li> <li>3. We have decided to conduct an online meeting regarding how to go forward with the task that we have selected for the 1<sup>st</sup> week. We also planned an approach for the workflow of the ChatBot Pipeline</li> <li>4. We just planned to implement the vector database for our first week using ChromaDB.</li> </ol>	<p>Write a status update on the tasks mentioned on the left. This could be empty for the first week.</p> <ol style="list-style-type: none"> <li>1. Done,</li> <li>2. Done, but considering other alternatives for better usage in future.</li> <li>3. Done, but they can be changed for the betterment of the ChatBot's Performance.</li> <li>4. Done</li> </ol>


## Work narrative of the current week

- **Tasks completed:** We have used ChromaDB for our vector database to store the PDF in chunk form and use it for retrieval. PDF loader has been implemented using `UnstructuredPDFLoader()` class. After loading the PDF, we have set the chunk size = 4000 with overlapping = 1000, initially. We have used the 'nomic-embed-text' text model of 'Ollama' for embedding. Imported packages and modules are: `chromadb`, `langchain_community.document_loaders`, `langchain_text_splitters` import, `langchain_community.embeddings` and `langchain_community.vectorstores`
- **Challenges faced:** During the task, we faced the problem of dependencies clashing, and version clashing of packages as all the machines do not acquire the exact requirement.
- **Progress made:** For the first week, we only implemented the code as it can load a PDF, then break it into chunks and store it in the database for future usage.

**[We are expecting future update for better workflow and design in the pipeline]**

## Target for next weekly update

1. Looking forward to performing retrievals from the created vector database for a given query by the user.
2. Planning to find better approaches.
3. Exploring more chunking strategies and better models for embedding packages and modules.
4. Planning to implement a minimalist and attractive UI with lots of surprises.

	<b>CSE 299: Junior Design Project</b>			
	<b>Department of Electrical and Computer Engineering</b>			
	<b>Semester:</b>	Spring 2025	<b>Section:</b>	18
	<b>Weekly Progress:</b>	2	<b>Date:</b>	14 <sup>th</sup> Feb, 2025

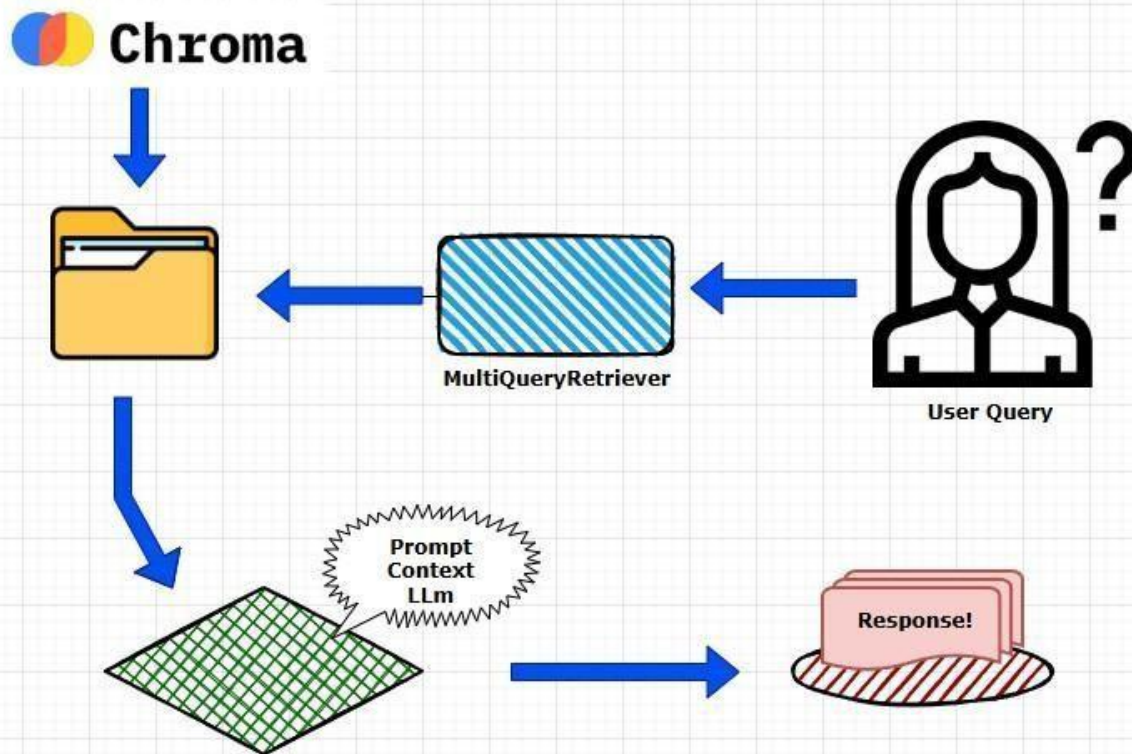
<b>Faculty Name:</b>	Dr. Shafin Rahman
<b>Group Number and Project Title</b>	Group # 7: Project Title: Bank Customer ChatBot
<b>Student Name (ID)</b>	Soleman Hossain 2021682042
<b>Email Address:</b>	soleman.hossain@northsouth.edu

Last week's Target	Follow-up
<ol style="list-style-type: none"> <li>1. Develop retrievers that can access an existing vector database located in the specified directory and utilize it to generate responses based on user queries.</li> <li>2. Optimize the workflow for better efficiency and performance.</li> <li>3. Explore various embedding models, embedding strategies, and vector databases to enhance the chatbot's overall performance and integrate the most effective solutions into its pipeline.</li> </ol>	<ol style="list-style-type: none"> <li>1. Completed.</li> <li>2. Implementing a multimodal RAG chatbot for improved response quality is in progress.</li> <li>3. Done. Explored a few models, strategies, and databases. We have also finalized the model and database we are going to use currently for the project. If we implement a better multimodal Chatbot later, we may need to swap a few elements of the pipeline.</li> </ol>

## Work narrative of the current week

<ol style="list-style-type: none"> <li>1. Set up the retriever using the LLM models "<b>llama3.2</b>" and "<b>mistral</b>" for this week.</li> <li>2. Set the <b>ChatOllama()</b> class to initialize the model.</li> <li>3. Created a query prompt template with <b>PromptTemplate()</b> from the langchain-ChatPromptTemplate, PromptTemplate module.</li> <li>4. Accessed the existing vector database, labeled it for easy reference, and enabled passing it as an argument when needed.</li> <li>5. Configured the retriever using <b>MultiQueryRetriever()</b> from the langchain-MultiQueryRetriever module.</li> <li>6. Developed a RAG prompt template using <b>ChatPromptTemplate()</b> and built the final chain for query processing.</li> <li>7. Integrated "<b>langchain</b>" to seamlessly combine the retriever, prompt, and LLM models into a structured processing chain.</li> </ol> <p><b>[We are expecting future updates for better workflow and design in the pipeline]</b></p>
---


## 8. This week's workflow



[Group Diagram]

## Target for next weekly update

1. Implement an advanced multimodal RAG chatbot to enhance overall performance.
2. Test the pipeline with different LLMs and compare their performance.
3. Develop a user-friendly interface for the chatbot.
4. Search for a suitable knowledge base to create test sets for chatbot evaluation.
5. Gather the necessary resources to conduct a thorough evaluation.

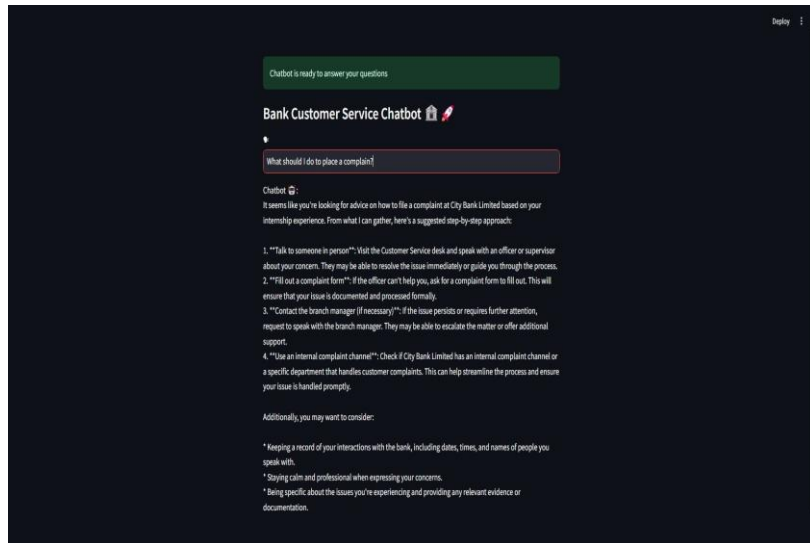
	<b>CSE 299: Junior Design Project</b>		
	<b>Department of Electrical and Computer Engineering</b>		
	<b>Semester:</b>	Spring 2025	<b>Section:</b> 18
	<b>Weekly Progress:</b>	3	<b>Date:</b> 28 <sup>th</sup> Feb, 2025

<b>Faculty Name:</b>	Dr. Shafin Rahman
<b>Group Number and Project Title</b>	Group # 7: Project Title: Bank Customer ChatBot
<b>Student Name (ID)</b>	Soleman Hossain 2021682042
<b>Email Address:</b>	soleman.hossain@northsouth.edu

Last week's Target	Follow-up
<ol style="list-style-type: none"> <li>1. Develop an advanced multimodal RAG chatbot, test it with different LLMs, and compare their performance.</li> <li>2. Develop a user-friendly interface for the chatbot.</li> <li>3. Find a suitable knowledge base and gather all required resources to create comprehensive chatbot evaluation test sets.</li> </ol>	<ol style="list-style-type: none"> <li>1. Completed.</li> <li>2. Created a simple Streamlit UI initially but still looking for a better approach.</li> <li>3. Considered some approach already. Still looking for a better approach to evaluate the Bot</li> </ol>

## Work narrative of the current week

1. Retrieve some queries using both '**llama 3.2**' and '**mistral**' and evaluate the model initially.
2. Developed an advanced multimodal RAG chatbot using **Ollama 3.2** and **Mistral**, testing their performance across various tasks. Ollama 3.2 excelled in retrieval efficiency and handling multimodal inputs, while Mistral provided more coherent and contextually rich responses.
3. Started with a simple **Streamlit** UI to get the chatbot up and running, but I'm still exploring better alternatives. The goal is to find a more flexible and efficient solution that improves usability and overall performance.




4. Researched and considered some relevant **knowledge bases** for creating detailed test sets to evaluate the chatbot. Also gathered the necessary resources, including datasets and tools, to conduct practical assessments of its retrieval accuracy and response quality.

**[We are expecting future updates for better workflow and design in the pipeline]**

## Target for next weekly update

1. Describe the primary metrics and performance indicators to be evaluated before starting the chatbot evaluation phase.
2. Conduct manual testing by asking a wide range of users for thorough input on the accuracy, performance, and user experience of the chatbot.
3. Create a plan for automated testing that includes finding and choosing pertinent terms, information, and requirements to expedite the assessment procedure.
4. Gather and arrange all required materials, including testing tools, datasets, and documentation, to enable a comprehensive, methodical, and repeatable assessment of the chatbot's functionality.

	<b>CSE 299: Junior Design Project</b>		
	<b>Department of Electrical and Computer Engineering</b>		
	<b>Semester:</b>	Spring 2025	<b>Section:</b> 18
	<b>Weekly Progress:</b>	4	<b>Date:</b> 14 <sup>th</sup> March, 2025

<b>Faculty Name:</b>	Dr. Shafin Rahman
<b>Group Number and Project Title</b>	Group # 7: Project Title: Bank Customer ChatBot
<b>Student Name (ID)</b>	Soleman Hossain 2021682
<b>Email Address:</b>	soleman.hossain@northsouth.edu

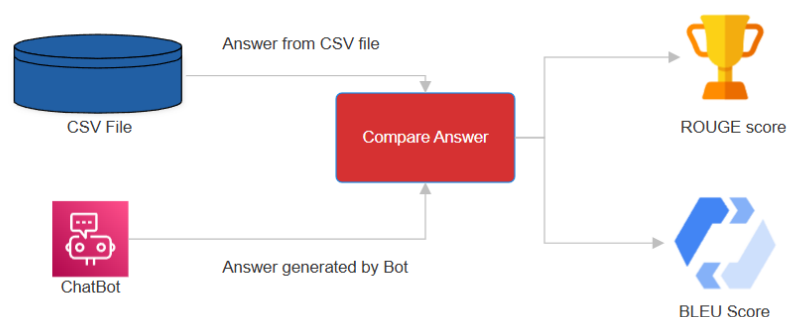
Last week's Target	Follow-up
<ol style="list-style-type: none"> <li>1. Select primary metrics to test the bot's performance and evaluation.</li> <li>2. Conduct manual testing with a diverse use of the bot</li> <li>3. Develop a comprehensive plan for automated testing.</li> <li>4. Gather and organize necessary testing materials and resources.</li> </ol>	<ol style="list-style-type: none"> <li>1. Almost Done.</li> <li>2. Half Done. We are still in the process.</li> <li>3. Still ongoing progress. We have to make a detailed plan. That's what takes time.</li> <li>4. Done.</li> </ol>

## Work narrative of the current week

<p>Here is a summary of the student's work on the project during the current week. This narrative should include:</p> <ul style="list-style-type: none"> <li> <b>Tasks completed:</b> The main evaluation criteria for the chatbot—accuracy, reaction time, and engagement—are almost complete and are presently undergoing validation based on input from the team. Although more users are required for a more varied viewpoint, manual testing has begun with preliminary user feedback on performance and user experience. Although the automated testing plan is being developed and includes a list of key phrases and tools, more work is needed to finish the setup. Every testing resource that is required, such as tools, datasets, and documentation, has been arranged and prepared for the assessment procedure.         </li> <li> <b>Challenges faced:</b> The process probably encountered a number of difficulties. It might have been difficult to define the appropriate performance indicators that would have matched both technical outcomes and user pleasure. It might have been challenging to assemble a varied user base for manual testing in order to guarantee thorough feedback. It may have taken some time to choose the right frameworks and tools for automated testing in order to strike a balance between accuracy and speed. Furthermore, it might have taken more time and effort to coordinate the relevant resources and guarantee compatibility.         </li> </ul>
--



- **Progress made:** Considerable progress has been made thus far. Accuracy, reaction speed, and user satisfaction—the main measures used to assess the chatbot—have almost been finalised and are undergoing final validation. With intentions to broaden testing to a larger user population, manual testing has already begun and useful user feedback regarding performance and experience has been gathered. The strategy for automated testing is well under way; important terms and tools have been selected, and the setup is still being finalised. Furthermore, all necessary testing resources—such as datasets and tools—are arranged and prepared for the assessment stage.



[An initial ongoing **demo** testing phase]


We are thinking of **ROUGE** and **BLEU** as our automated testing to evaluate our ChatBot. But the Plan can be changed if we see a better approach.

[We are expecting future updates for better workflow and design in the pipeline]

## Target for next weekly update

Here are next week's targets:

1. **Finalize Primary Metrics:** Get the team to validate the key performance indicators (KPIs) and get them ready for the assessment stage.
2. **Extend Manual Testing:** To get more varied input on chatbot performance and user experience, expand the user pool for manual testing.
3. **Finish Automated Testing Setup:** Complete the automated testing strategy, including the framework and tool selection, and begin configuring the testing infrastructure.
4. **Implement Voice Interaction System:** Start working on the voice interaction system and ensure it integrates seamlessly with the chatbot's current functionalities.

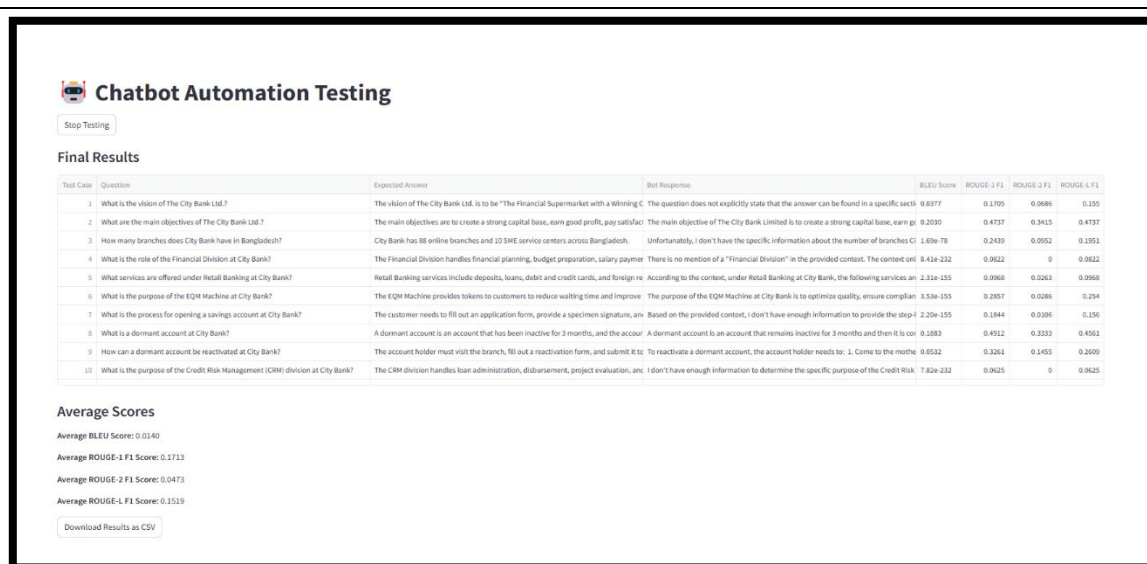
	<b>CSE 299: Junior Design Project</b>		
	<b>Department of Electrical and Computer Engineering</b>		
	<b>Semester:</b>	Spring 2025	<b>Section:</b> 18
	<b>Weekly Progress:</b>	5	<b>Date:</b> 28 <sup>th</sup> March, 2025

<b>Faculty Name:</b>	Dr. Shafin Rahman
<b>Group Number and Project Title</b>	Group # 5: Project Title: Bank Customer ChatBot
<b>Student Name (ID)</b>	Soleman Hossain 2021682
<b>Email Address:</b>	soleman.hossain@northsouth.edu

Last week's Target	Follow-up
1. Finalize primary matrices. 2. Extend manual testing. 3. Fully automated testing setup 4. Implement a voice interaction system.	1. Done. 2. Ongoing 3. Done 4. Ongoing.

## Work narrative of the current week

<p>Here is a summary of the student's work on the project during the current week. This narrative should include:</p> <ul style="list-style-type: none"> <li> <b>Tasks completed:</b> This week, our team made significant progress in evaluating and enhancing our RAG-based chatbot through automated and manual testing. We implemented a comprehensive testing framework using BLEU and ROUGE metrics to objectively measure the chatbot's performance against predefined test cases. The Streamlit dashboard we developed systematically processes each question, compares the chatbot's response with the expected answer, and generates detailed scores, including BLEU for n-gram precision and ROUGE-1, ROUGE-2, and ROUGE-L for recall-based assessment. Alongside automated testing, we conducted manual validation to cross-check the results and identify cases where the chatbot's responses, while technically correct, needed contextual refinement. Additionally, we began exploring voice integration capabilities, experimenting with speech-to-text and text-to-speech libraries to enable seamless voice interactions. Our next steps involve further optimizing the testing process and improving the voice system's responsiveness. Here are some screenshot of automation testing and also we were able to add a feature where the user can download the automation testing result as excel file for better view.           </li> </ul>
---



## Automation Testing

chatbot\_test\_results - Excel

Test Case	Question	Expected Answer	Bot Response	BLEU Score	ROUGE-1 F1	ROUGE-2 F1	ROUGE-L F1
1	What is the vision of The City Bank Ltd.?	The vision of The City Bank Ltd. is to be "The Financial Supermarket with a Winning C...	The question does not explicitly state that the answer can be found in a specific secti...	0.8377	0.170542633	0.068571427	0.155039
2	What are the main objectives of The City Bank Ltd.?	The main objectives are to create a strong capital base. The main objective of The City Bank Limited is to create a strong capital base, earn good profit, pay satisfi...	The main objective of The City Bank Limited is to create a strong capital base, earn g...	0.203	0.473684206	0.341463411	0.473684
3	How many branches does City Bank have in Bangladesh?	City Bank has 88 online branches and 10 SME service c	Unfortunately, I don't have the specific info	1.69E-78	0.24302435	0.095238091	0.195122
4	What is the role of the Financial Division at City Bank?	The Financial Division handles financial planning, budg	There is no mention of a "Financial	8.41E-232	0.962191777	0	0.962192
5	What services are offered under Retail Banking at City Bank?	Retail Banking services include deposits, loans, debit a	According to the context, under Retail	2.31E-155	0.09677419	0.026315787	0.096774
6	What is the purpose of the EQM Machine at City Bank?	The EQM Machine provides tokens to customers to red	The purpose of the EQM Machine at City B	3.53E-155	0.285714282	0.025967142	0.253966
7	What is the process for opening a savings account at City Bank?	The customer needs to fill out an application form, prov	Based on the provided context, I don't	2.20E-155	0.18439716	0.010562008	0.156028
8	What is a dormant account at City Bank?	A dormant account is an account that has been inactive	A dormant account is an account that rem	0.1883	0.491220066	0.333333329	0.45614
9	How can a dormant account be reactivated at City Bank?	The account holder must visit the branch, fill out a reac	To reactivate a dormant account, the	0.0532	0.326086953	0.145454542	0.26087
10	What is the purpose of the Credit Risk Management (CRM) division at City Bank?	The CRM division handles loan administration, disburs	I don't have enough information to	7.82E-232	0.962499997	0	0.9625
11	What types of loans are offered by City Bank?	City Bank offers loans such as City Home Loan, City Scl	I don't have the information to provide a dir	2.43E-155	0.070175435	0.027777774	0.070175
12	What is the role of the Human Resource Division at City Bank?	The HRD handles recruitment, training, employee benef	I don't have information about the specific	7.60E-232	0.065573787	0	0.065574
13	What is the significance of the IT Department at City Bank?	The IT Department manages software development, ne	I don't have information about the IT Depar	2.68E-155	0.107142654	0.032786882	0.107143
14	What is the purpose of the Branch Control & Insp at City Bank?	This division controls branch functions, conducts inter	I don't have the specific information about	6.96E-232	0.072727269	0	0.072727
15	What is the process for closing an account at City Bank?	The customer must submit an application, withdraw th	According to the document, the process	1.13E-78	0.289156623	0.092592589	0.289157
16	What is the purpose of the City Bank Treasury Department?	This division provides treasury solutions, including fore	I don't have information about the specific	7.16E-232	0.051282048	0	0.051282
17	What is the City Bank's approach to SME Banking?	City Bank aims to support SMEs by offering tailored fina	The question requires knowledge of City	2.55E-155	0.130454779	0.019047616	0.108696
18	What is the City Bank's policy on Islamic Banking?	City Bank offers Islamic Banking services under the nar	According to the provided context, City	2.55E-155	0.177777775	0.019047617	0.133333
19	What is the process for applying for a loan at City Bank?	The customer must submit an application, provide nec	I don't have the information needed to	1.01E-231	0.163908042	0	0.16092
20	What is the role of the Customer Service Department at City Bank?	The Customer Service department handles account op	The question cannot be answered based	1.19E-78	0.178571425	0.067796607	0.178571
21	What is the purpose of the City Bank's CRM model?	The CRM model focuses on ensuring customer satisfac	I don't have information about the City Ban	9.59E-232	0.149999995	0	0.15
22	What is the City Bank's strategy for foreign remittance?	City Bank has a strong remittance network with exchan	The City Bank's strategy for foreign remita	0.0544	0.240963852	0.108695648	0.240964
23	What is the City Bank's policy on account nomination?	The account holder can nominate a person who will ha	I don't have any information about City Ban	1.10E-231	0.170212761	0	0.170213
24	What is the City Bank's approach to corporate banking?	City Bank focuses on building long-term relationships w	According to the provided context, City	0.0206	0.232142655	0.040816326	0.232143
25	What is the City Bank's policy on cheque books?	Cheque books are issued to account holders upon req	According to the provided context, accord	9.40E-79	0.24324324	0.068181815	0.243243

## Automation Testing Result in CSV Format

- Challenges faced:** While implementing the automated testing framework and voice integration system, our team encountered several challenges. One major difficulty was handling extremely low BLEU scores for responses that were semantically correct but phrased differently from the reference answers, requiring us to implement decimal precision formatting for better readability. The ROUGE metrics sometimes produced inconsistent results with very short or very long responses, making interpretation tricky. For the voice integration, we faced latency issues in real-time speech processing and accuracy problems with different accents and background noise in the STT (Speech-to-Text) conversion. Additionally, synchronizing the automated testing pipeline with manual validation proved time-consuming, as we had to ensure the metrics aligned with human judgment while maintaining testing efficiency. These challenges highlighted areas for improvement in both evaluation methods and voice system optimization.

- **Progress made:** This week, our team made significant progress in enhancing our RAG-based chatbot's evaluation framework and expanding its capabilities. We successfully implemented an automated testing system using BLEU and ROUGE metrics through a user-friendly Streamlit dashboard, enabling efficient assessment of the chatbot's responses against expected outputs. The system now generates comprehensive reports with formatted scoring, including handling of edge cases like extremely low values. Alongside automation, we conducted thorough manual testing to validate the metrics' reliability. In parallel, we achieved a key milestone by developing a functional prototype for voice integration, establishing the foundation for speech-based interactions. These advancements have strengthened our testing methodology while moving us closer to creating a multimodal conversational interface.

**[We are expecting future updates for better workflow and design in the pipeline]**

## Target for next weekly update

Here are next week's targets:

1. **Extend Manual Testing:** To get more varied input on chatbot performance and user experience, expand the user pool for manual testing.
2. **Implement Voice Interaction System:** Start working on the voice interaction system and ensure it integrates seamlessly with the chatbot's current functionalities.

**[The Team will add further improvement in the pipeline if possible and the result will be shown in the final report]**