Evaluation and Testing Report: University Chatbot Project

# Chatbot 1 Evaluation and Testing (ASKDAH\_CHATBOT.ipynb)

## 1. Overview

Chatbot 1 utilizes a Retrieval-Augmented Generation (RAG) system with LangChain and GPT-4. It retrieves answers strictly from uploaded PDF documents using Chroma vector database, and supports interactive Q&A in a looped input format.

## 2. Test Plan

The chatbot was tested with realistic student queries focusing on functionality, robustness, and relevance. Each test checked alignment with intended document-based response behavior.

## 3. Sample Interaction Output

Sample inputs included: 'how to reset my email password?', 'where is the IT department?', 'what is 5\*4'. The responses were relevant and grounded in the student documents.

## 4. Test Results Summary

|  |  |  |  |
| --- | --- | --- | --- |
| Test Case | Expected Outcome | Actual Result | Result |
| Reset email password | Should provide steps from DAH document | Provided accurate instructions | ✅ Pass |
| Location of IT department | Should return location info | Correctly referenced HRC | ✅ Pass |
| Typo greeting 'helli' | Should handle input | Responded appropriately | ✅ Pass |
| Math exam help | Only if info is in source | Gave generic advice | ⚠️ Partial |
| Simple math 5\*4 | Give correct result if accepted | Returned 20 | ✅ Pass |

## 5. Quantitative Evaluation

|  |  |
| --- | --- |
| Criteria | Rating (1–5) |
| Functionality | 5 |
| Accuracy | 4 |
| Robustness | 5 |
| Relevance | 4 |
| User Experience | 5 |
| Security & Ethics | 5 |

## 6. Findings and Recommendations

- High retrieval accuracy  
- Great user experience  
- Needs improved constraint to avoid generic fallback answers

## 7. Conclusion

Chatbot 1 is effective and suitable for real student use. Its interactive design and document focus offer a reliable virtual assistant experience.

# Chatbot 2 Evaluation and Testing(CHATBOT\_ASKDAH.ipynb)

## 1. Overview

Chatbot 2 is a document QA bot built using FAISS and LangChain’s OpenAI integration. It processes a student handbook PDF and answers one-off questions using similarity search and a LangChain QA chain.

## 2. Test Plan

The chatbot was tested with common student queries. Its goal is to provide answers found within the PDF.

## 3. Sample Interaction Output

Test inputs: 'Vision for university', 'How can I find my semester dues?'. It successfully returned document-based answers.

## 4. Test Results Summary

|  |  |  |  |
| --- | --- | --- | --- |
| Test Case | Expected Outcome | Actual Output | Result |
| Vision for university | Return info from handbook | Gave relevant excerpt | ✅ Pass |
| Semester dues inquiry | Explain how to check | Provided process | ✅ Pass |
| Misspelled query | Handle typos | Partial handling | ⚠️ Partial |
| Out-of-scope question | Should avoid hallucination | Gave unrelated response | ❌ Fail |
| Sensitive info | Avoid personal data | No explicit guardrails | ⚠️ Partial |

## 5. Quantitative Evaluation

|  |  |
| --- | --- |
| Criteria | Rating (1–5) |
| Functionality | 5 |
| Accuracy | 4 |
| Robustness | 3 |
| Relevance | 3 |
| User Experience | 4 |
| Security & Ethics | 3 |

## 6. Findings and Recommendations

- Fast document QA  
- Needs conversation loop for usability  
- Lacks constraints on irrelevant questions

## 7. Conclusion

Chatbot 2 performs well as a backend engine for document retrieval. However, without constraints or user interface features, it's better suited as a prototype or internal tool.