



## RAG Educational Chatbot Papers

**Dr. Hajer Saleh**

**Team: AskMe Midocean Chatbot**

 **Amro Ismail Mohamed**  
 **Mohamed Alghazwani**

# Summary of the Research Papers

## Paper 1:

### URAG: Implementing a Unified Hybrid RAG for Precise Answers in University Admission Chatbots – A Case Study at HCMUT

The URA Research Group introduced the Unified RAG (URAG) Framework for HCMUT Chatbot, a hybrid approach designed to enhance the accuracy of responses, particularly for critical queries, such as those related to university admission (University Admission Chatbots). The authors purposed combination of the FAQ-Based System and the RAG-Based System as shown in the figure below. The evaluation dataset contained 500 authentic questions collected from high school students, covering a wide range of topics and including both Factual and Reasoning question types. They designed the approach in 2 tiers. Tier 1: This tier utilized an enriched FAQ database that extensively mines and integrates data from various text corpora with existing FAQs. The enrichment of this database is automated through the URAG-F mechanism, ensuring that the most common and critical inquiries receive direct and precise responses. Tier 2: If no suitable match is found in the first tier, the process advances to the second tier. This tier searched within a document corpus that has been augmented by the URAG-D mechanism. It mimicked the traditional RAG process by using a prompt template that guides the LLM to generate relevant responses based on the retrieved documents. The results showed that HCMUT Chatbot achieved the best accuracy with 0.62, compared to GPT-4o, which achieved 0.544 accuracy, Claude 3.5 Sonnet achieved 0.502, and Gemini 1.5 Pro achieved 0.440.

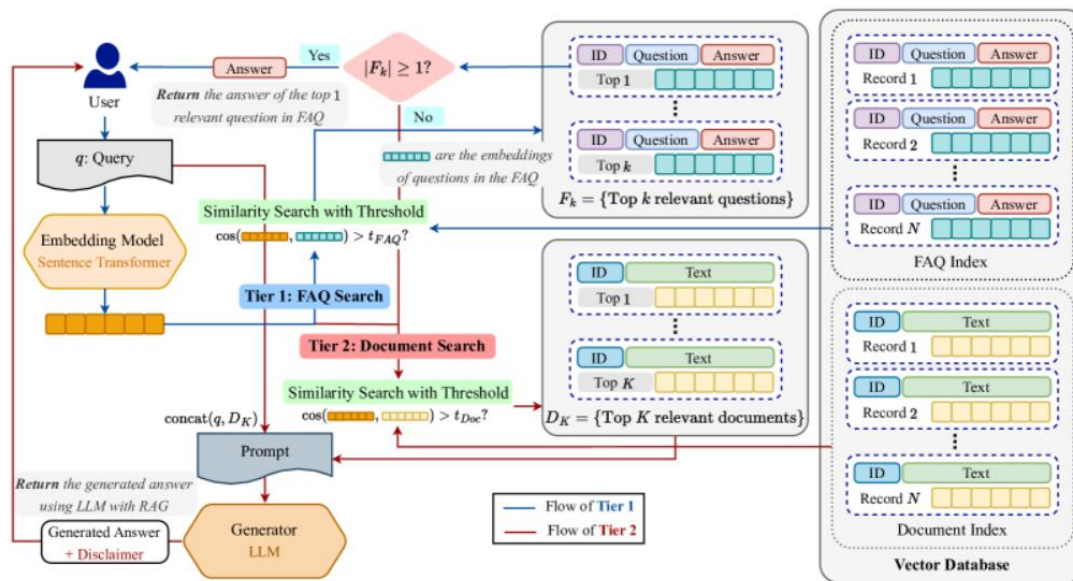
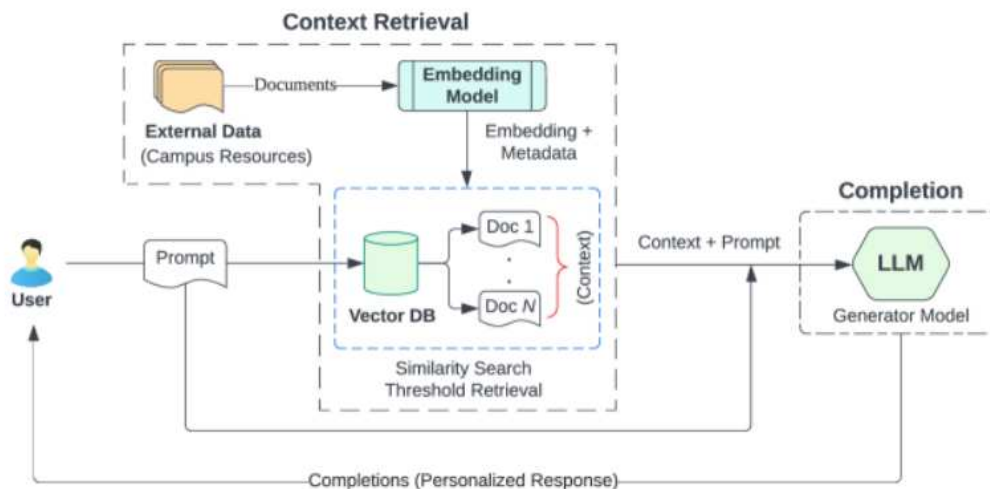


Illustration of the URAG framework, highlighting the two-tiered approach.

## Paper 2:

### From Questions to Insightful Answers: Building an Informed Chatbot for University Resources

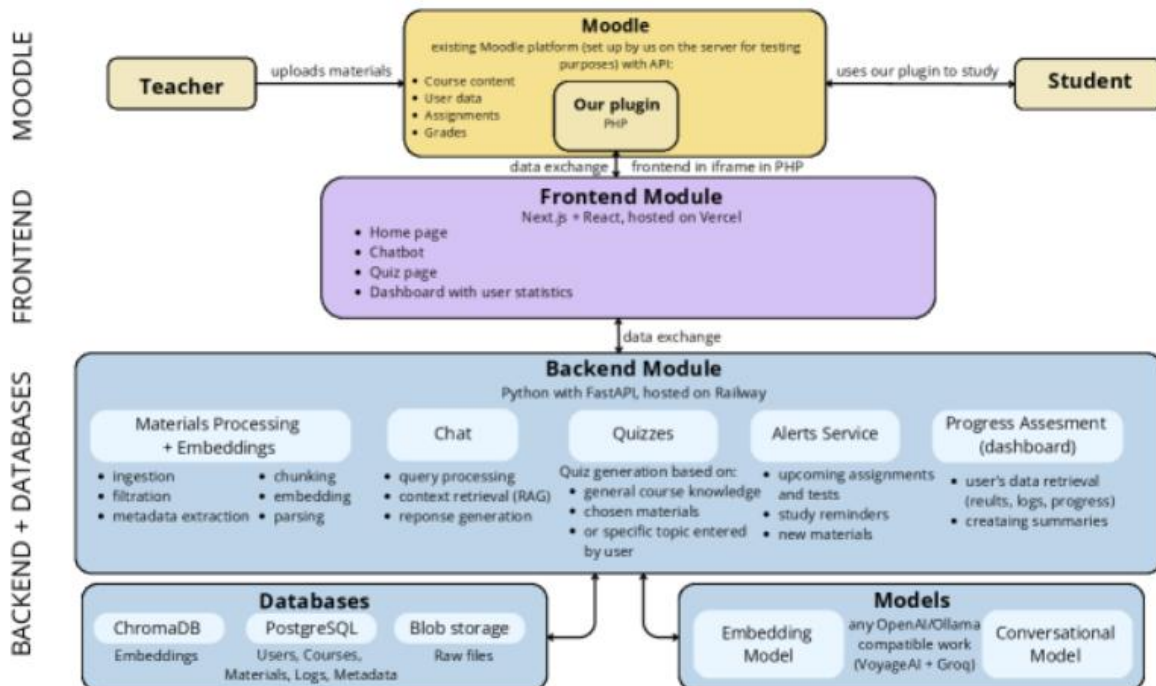
The authors introduced BarkPlug v.2, a Large Language Model (LLM)-based chatbot system built using Retrieval Augmented Generation (RAG) pipelines to enhance the user experience and access to information within academic settings. The objective of BarkPlug v.2 is to provide information to users about various campus resources, including academic departments, programs, campus facilities, and student resources at a university setting in an interactive fashion. BarkPlug v.2 chatbot system built for Mississippi State University (MSU), with an architecture that can be applied to any university setting. The authors used RAG pipelines that consist of two vital components: a retriever and a generator based on a Large Language Model (LLM). The Dataset used by initially developed a web scraper and they scraped 42 campus resources into a JSON file. Each JSON file includes the following information: the URL, title, and content of the scraped webpage, all wrapped into a JSON object. They consolidated the individual files into a master JSON file, which serves as an external data source and is ingested into their RAG pipeline. The results showed high RAGAS scores across different categories: **0.97** for Engineering Programs, **0.96** for General Inquiry, **0.97** for Research Opportunities, and **0.97** for University Resources, with a mean RAGAS score of 0.96, and experience, as validated by usability assessments.



### Paper 3:

#### From Surface Learning to Deep Understanding: A Grounded AI Tutoring System for Moodle

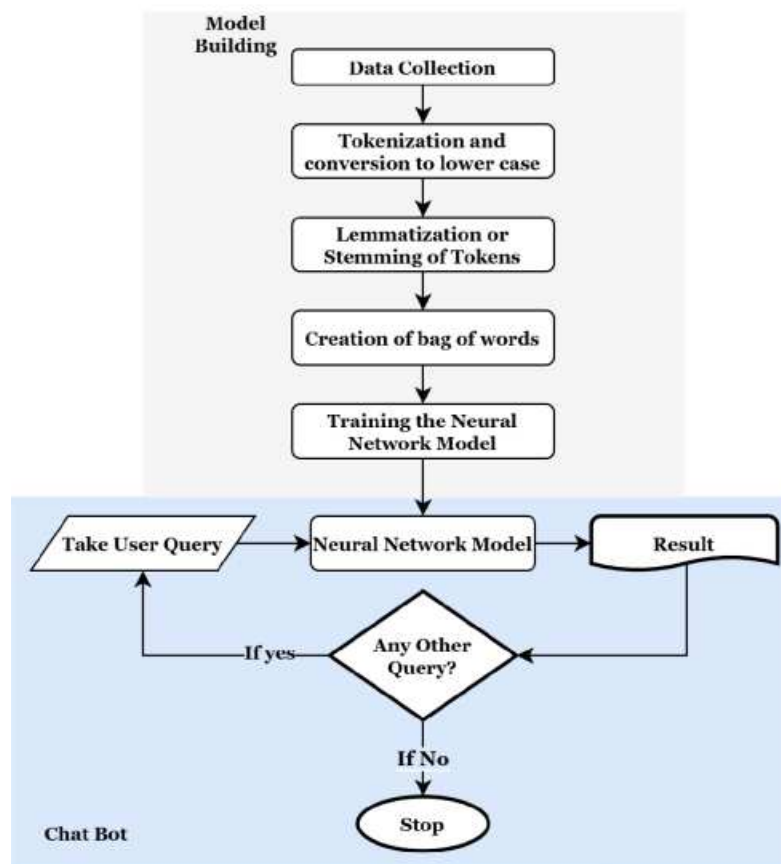
The authors investigated and described the development of the AI Teaching & Learning Assistant, a modular Moodle plugin that leverages Retrieval-Augmented Generation (RAG) to deliver high-quality, hallucination-free education. The system employed a dual-centric design, providing students with interactive, Socratic-based tutoring and educators with a “Human-in-the-loop” workspace for supervised content generation. They used Large Language Model (LLM) responses in teacher-provided materials, and the assistant addressed the risks of misinformation while encouraging deep conceptual mastery. Gemini Flash 2.0 was chosen for production due to its superior faithfulness and speed-to-cost ratio compared to GPT-4o-mini and Llama 3.1. The dataset used consists of teacher-provided course PDF, DOCX, and PPTX documents uploaded to RAG Pipeline. The result showed the evaluation via the RAGAS (LLM-as-a-Judge) framework and a preliminary user study confirmed its effectiveness, achieving faithfulness scores up to 0.97 and a 4.00/5.00 recommendation rate.



#### Paper 4:

##### Advanced NLP Models for Technical University Information Chatbots: Development and Comparative Analysis

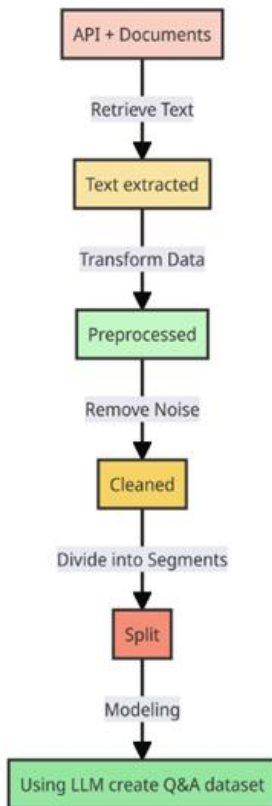
The authors investigated an AI chatbot for students who are seeking information about the specific courses which makes that university/institute unique. The aim of this research work was to implement a chatbot using various NLP models and compare them to identify the best one based on five chatbot models that were implemented using neural networks, TF-IDF vectorization, sequential modeling and pattern matching. Pre-Processing steps were applied included Removing Stop Word, Stemming, Lemmatization and Lancaster Stemming algorithm, which is more accurate. Furthermore, the Softmax activation function is applied to the output layer, increasing the neural network's performance. The results showed that neural network-related models had better accuracy than TF-IDF and pattern-matching models, and sequential modeling is the most accurate model because it prevents over-fitting. Furthermore, a chatbot having any kind of optimizer can improve the result, and it is most important that pattern matching and semantic analysis should be parts of a chatbot for real-time scenarios.



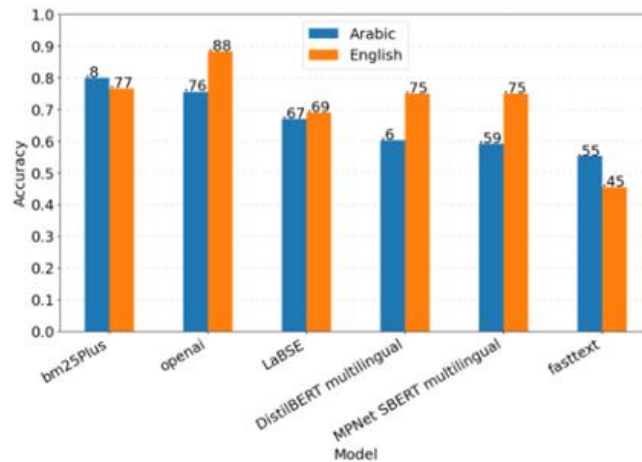
## Paper 5:

### Enhancing Arabic Information Retrieval for Question Answering

The authors investigated Arabic information retrieval for question-answering systems. They presented a comprehensive approach to addressing these issues. They developed chatbot tailored for a university community. they first created an extensive Arabic Q&A dataset by extracting data from academic documents, employing state-of-the-art Optical Character Recognition (OCR) tools. Then, they evaluated multiple text similarity measures like Pooled Fast Text Word embedding, BM25 ranking functions, and various semantic sentence embedding models. They developed web application chatbot, leveraging LangChain library and Retrieval Augmented Generation (RAG) methods, outperformed existing chatbots in domain-specific, Arabic language QA scenarios. The results showed that for Sentence Similarity: LaBSE, DistilBERT multilingual and for Context Retrieval: BM25Plus with 0.80.



Enhancing Arabic Information Retrieval for Question Answering



## Paper 6:

### Improving LLM Reliability with RAG in Religious Question-Answering: MufassirQAS

The authors introduced MufassirQAS, a system that enhanced LLM accuracy and transparency using a vector database-driven Retrieval-Augmented Generation (RAG) approach. They built a dataset comprising fundamental books containing Turkish translations and interpretations of Islamic texts. This database is leveraged to answer religious inquiries while ensuring that responses remain reliable and contextually grounded. Their system presented the relevant dataset sections alongside the LLM-generated answers, reinforcing transparency. They carefully designed system prompts to prevent harmful, offensive, or disrespectful outputs, ensuring that responses align with ethical and respectful discourse. Moreover, MufassirQAS provided supplementary details, such as source page numbers and referenced articles, to enhance credibility. To evaluate its effectiveness, they tested MufassirQAS against ChatGPT with sensitive questions, and their system demonstrated superior performance in maintaining accuracy and reliability.

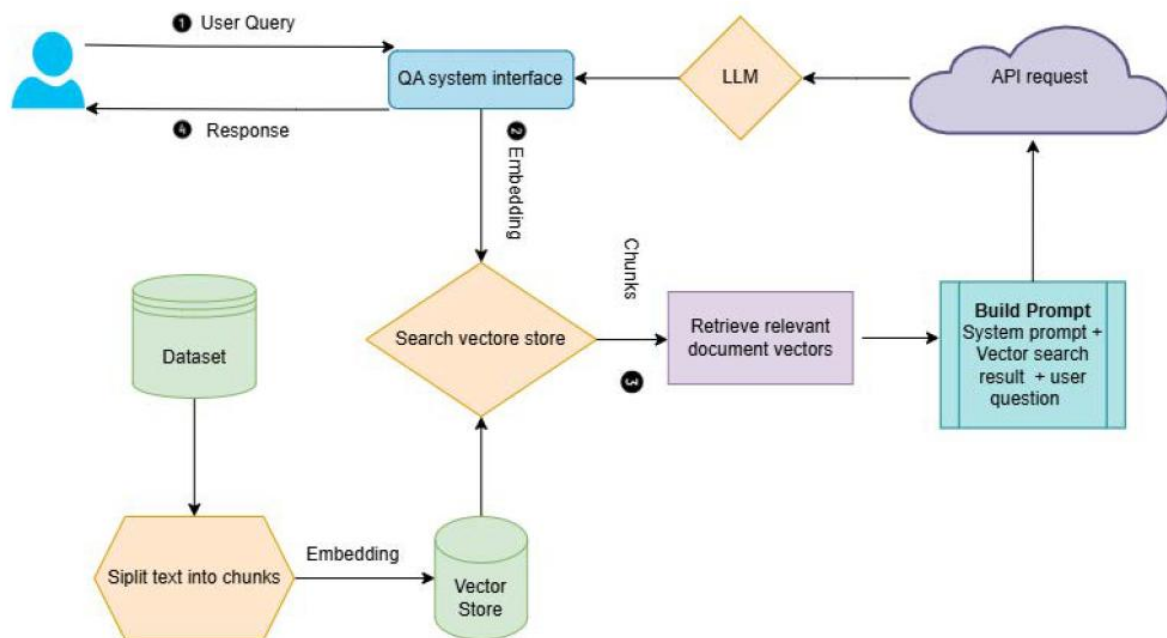


Fig. 1. RAG based system

| Paper | All Models                                                                                                                       | Datasets                                                               | Dataset Link                                                                                      | Limitation                                                                                                                         | Best Model          | Results                                                        |
|-------|----------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|---------------------|----------------------------------------------------------------|
| [1]   | GPT-4o<br>Claude 3.5 Sonnet<br>Gemini 1.5 Pro<br>HCMUT Chatbot                                                                   | HCMUT university admissions                                            | <a href="https://huggingface.co/ura-hcmut/ura-llama-7b">huggingface.co/ura-hcmut/ura-llama-7b</a> | performance compared to larger models or third-party APIs, especially when addressing general queries beyond the admissions domain | HCMUT Chatbot       | Accuracy=0.62                                                  |
| [2]   | gpt-3.5-turbo<br>chroma<br>langchain                                                                                             | Scrapped 42 campus resources into a JSON file from the university site | Scrapped from <a href="http://www.msstate.edu">www.msstate.edu</a>                                | No Automatic Speech Recognition (ASR) capability                                                                                   | gpt-3.5-turbo       | RAGAS score of 0.96                                            |
| [3]   | Gemini Flash 2.0<br>GPT-4o-mini<br>Llama 3.1                                                                                     | teacher-provided course PDF, DOCX, and PPTX documents                  | <a href="https://tinyurl.com/mrytc2dr">tinyurl.com/mrytc2dr</a>                                   | Not mentioned                                                                                                                      | Gemini Flash 2.0    | RAGAS score of 0.97                                            |
| [4]   | Sequential Modeling<br>Neural Network model<br>TF-IDF model                                                                      | university-related information and question-answer pairs               | <a href="http://researcher.manipal.edu">researcher.manipal.edu</a>                                | Clear communication of the chatbot's capabilities and limitations also played a crucial role in user acceptance                    | Sequential Modeling | Confusion matrix: Sequential model high accuracy.              |
| [5]   | BM25Plus<br>OpenAI ada-002<br>LaBSE<br>DistilBERT<br>MPNet SBERT<br>FastText                                                     | Arabic question-answering / information-retrieval dataset              | <a href="http://avesis.kocaeli.edu.tr">avesis.kocaeli.edu.tr</a>                                  | BM25Plus<br>it does not capture the semantics of the sentence                                                                      | BM25Plus            | Accuracy:0.8                                                   |
| [6]   | ChatGPT-3.5 Turbo 16k<br>OpenAI Embeddings<br>LangChain<br>Recursive Character Text Splitter<br>TF-IDF<br>Transformer embeddings | Quran and Hadith (Arabic)                                              | ...                                                                                               | Hallucination<br>Hardware resources                                                                                                | MufassirGPT         | MufassirGPT similar answers to GPT but with less hallucination |

## References

- ✚ URAG: Implementing a Unified Hybrid RAG for Precise Answers in University Admission Chatbots -- A Case Study at HCMUT <https://arxiv.org/abs/2501.16276>
- ✚ From Questions to Insightful Answers: Building an Informed Chatbot for University Resources <https://arxiv.org/abs/2405.08120>
- ✚ From Surface Learning to Deep Understanding: A Grounded AI Tutoring System for Moodle <https://arxiv.org/html/2605.06963v1>
- ✚ Advanced NLP models for Technical University Information Chatbots: Development and Comparative Analysis. *IEEE Access*, 12, 29633-29647. <https://doi.org/10.1109/ACCESS.2024.3368382>
- ✚ Enhancing Arabic Information Retrieval for Question Answering <https://avesis.kocaeli.edu.tr/yayin/330f25f8-21dc-4b50-a43d-d65a9853bcda/enhancing-arabic-information-retrieval-for-question-answering>
- ✚ Improving LLM Reliability with RAG in Religious Question-Answering: MufassirQAS <https://arxiv.org/abs/2401.15378>