JAIIT

Journal of Artificial Intelligence and Information Technology

E-ISSN : 3090-2428
P-ISSN : 3090-2436

*Research Article*

# Retrieved Augmented Generation (RAG)-Based Academic Chatbot for the Informatics Engineering Study Program, FTI UNISSULA

**Ikharista Ayu Nusrotun Afifah [1]*, Sam Farisa Chaerul Haviana [2]**

[1] Universitas Islam Sultan Agung, Indonesia
Email: ikharista4@gmail.com
[2] Universitas Islam Sultan Agung, Indonesia
Email: sam@unissula.ac.id

* Corresponding Author : ikharista4@gmail.com

**Abstract:** Fast, accurate, and reliable access to academic information remains a major challenge for Informatics Engineering students at the Faculty of Industrial Technology, Sultan Agung Islamic University (UNISSULA). Educational information found in various official documents is often difficult to find quickly, leading to confusion and dependence on external news that may be inaccurate. This research aims to create and implement an academic chatbot that uses Retrieval-Augmented Generation (RAG) technology to provide contextually relevant academic information based on official documents. The research method includes collecting and pre-processing official academic documents, dividing them into smaller parts, generating vectors using the Sentence-BERT model, and storing the vectors in the FAISS database to enable meaningful information retrieval. The RAG mechanism combines relevant document search results with the ability of a large language model (LLM) to generate accurate answers. Evaluation results show that the chatbot successfully answered all questions with a 100% success rate. Tests using the ROUGE-1 and BLEU-4 metrics showed excellent results in handling questions from the FAQ category, as well as demonstrating fairly good meaning relevance for non-FAQ questions, despite differences in answer inference. These findings indicate that the RAG approach successfully improves the accuracy, reliability, and context of chatbot answers in the academic field, as well as assisting in the development of intelligent academic information systems in higher education environments.

**Keywords:** Academic Chatbot, Academic Information System, FAISS, Informatics Engineering, Retrieval-Augmented Generation, Semantic Search, Sentence-BERT.

**Received:** January 16, 2026
**Revised:** January 30, 2026
**Accepted:** February 25, 2026
**Published:** February 28, 2026
**Curr. Ver.:** February 28, 2026

**Copyright:** © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (https://creativecommons.org/licenses/by-sa/4.0/)

## 1. Introduction

Access to accurate, fast, and reliable academic information is a key factor in supporting student success in higher education. However, in practice, students often experience difficulties in obtaining information related to academic activities, administration, and lectures. This phenomenon also occurs among students in the Informatics Engineering Study Program, Faculty of Industrial Technology (FTI), Sultan Agung Islamic University (UNISSULA), who often feel confused when searching for information such as the procedure for submitting a final project, the number of credits, filling out the student study plan (KRS), and other administrative requirements. Similar problems are not unique to this environment but are also a common issue at various universities, where academic information delivery systems are not yet fully integrated and easily accessible to students.

In practice, student questions tend to be repetitive and frequent. When official information is difficult to find or not presented concisely, students often rely on informal communication with upperclassmen. Unfortunately, the information obtained through this channel is often inaccurate because it is conveyed verbally without reference to official documents, potentially leading to misunderstandings and further confusion. This situation indicates a gap between the availability of academic information and student accessibility.

This gap is even more apparent in the context of researchers' role as aspirators (aspiration committee) in the Student Senate (SEMA) of the Faculty of Technology (FTI) at UNISSULA. Researchers frequently receive various questions from students, which must first be validated by authorized parties, such as the Head of Study Program, lecturers, or administrative staff. This process is quite time-consuming and often results in delayed responses, especially at crucial times that require quick responses, such as filling out the Study Plan Card (KRS) or fulfilling certain academic requirements. These findings align with research indicating that academic information systems in higher education institutions are still unable to meet students' real-time needs, both in terms of speed of access and clarity of information. Although most academic information is readily available through various media such as academic guidebooks, faculty social media accounts, and official WhatsApp groups, the presentation is still perceived as inefficient and unstructured. Students must manually search and read lengthy documents to find the information they need, making the search process inefficient and tedious. This fact demonstrates that the mere presence of information is not enough; a system is needed that can present information contextually and easily understood according to user needs.

In recent years, academic chatbots have begun to be developed as a solution to these problems. However, most of the chatbots used are still based on traditional methods such as rule-based systems, TF-IDF, or BM25. These methods have limitations in understanding the context and variations of the user's natural language, so the resulting answers are often rigid and less relevant. On the other hand, the use of large language models (LLM) purely also raises new problems, namely the tendency to produce information hallucinations, especially when handling administrative and procedural data that require a high level of precision.

The development of artificial intelligence technology in education has encouraged the emergence of more adaptive and reliable approaches, one of which is *Retrieval-Augmented Generation* (RAG). This approach combines the ability to search for documents from official sources (retrieval) with the natural language generation capabilities of the LLM model (generation). Thus, the system not only generates answers based on language patterns but also refers directly to valid and verified documents. Studies by have shown that RAG-based chatbots in academic environments are able to produce more relevant and reliable responses and reduce the risk of hallucinations. A similar implementation in the "Unimib Assistant" system also proves that the RAG architecture is effective for answering administrative and academic questions based on internal university documents. Although the RAG approach has been widely researched and used, its application specifically to meet academic information needs at the study program level, especially in the Informatics Engineering environment of the FTI UNISSULA, is still very limited. This indicates that there is still a gap in research regarding the development of RAG-based chatbots and the real needs of students for an integrated, contextually appropriate, and reliable academic information system. Therefore, this study states that the use of academic chatbots based on *Retrieval-Augmented Generation technology* , by utilizing official documents from study programs and faculties, is an appropriate and immediately necessary solution to address this inequality.

This research is significant because it contributes to the development of an academic AI assistant model that not only helps improve the speed and ease of information access for students but also simplifies administrative tasks performed by study program administrators. Furthermore, this research is expected to provide additional empirical knowledge regarding the application of RAG in higher education environments in Indonesia, particularly in Informatics Engineering study programs, and serve as a reference in building a smarter, more precise, and user-friendly academic information system.

Based on previous research, there has been no research specifically creating an academic chatbot using *the Retrieval-Augmented Generation* (RAG) method at the study program level, relying on official internal academic documents as the primary source of knowledge, especially in the Indonesian higher education environment. Therefore, this study aims to develop and evaluate an academic chatbot using RAG technology, designed to meet the official academic information needs of students in the Informatics Engineering Study Program, Faculty of Information Technology, UNISSULA.

The main contributions of this research are:

1. Building a RAG-based academic chatbot architecture sourced from official study program and faculty documents.
2. Help gain faster, more accurate, and more interactive access to academic information
3. Contribute to the development of RAG-based AI systems in local educational environments.

## 2. Previous Research or Literature Review

### Academic Chatbots in Higher Education Environments

The use of chatbots as academic assistants has been widely studied as a solution to improve the quality of information services in higher education. Academic chatbots are generally used to assist students in quickly and independently obtaining information related to lectures, administration, seminar requirements, final assignments, and other campus services. However, initial research indicates that most academic chatbots still use a rule-based or keyword-matching approach, which has limitations in understanding the context of user queries.

shows that the informal and unstructured academic information delivery system has the potential to cause misinformation and misunderstanding among students. This condition drives the need for a chatbot system that is able to provide accurate academic information based on official sources. Furthermore, developed an academic *helpdesk chatbot* aimed at reducing the workload of administrative staff. The results of their research show that the chatbot is able to provide fast and relevant responses, although it still faces challenges in system stability and adjustment to the academic context.

Furthermore, emphasized that academic chatbots play a crucial role in increasing the accessibility of campus information for students. The study demonstrated that the presence of chatbots can accelerate the information-seeking process and reduce students' reliance on manual confirmation from the administration. However, these studies are generally general in nature and have not specifically addressed the need for formal academic information at the study program level.

### Implementation of Retrieval-Augmented Generation (RAG) in Academic Chatbots

The use of chatbots as academic assistants has been widely researched as a solution to improve the quality of information services in higher education. Academic chatbots are generally used to help students obtain information related to lectures, administration, and campus services quickly and independently. However, initial research indicates that most academic chatbots still use a rule-based or keyword-matching approach, which has limitations in understanding the context of user questions

showed that the informal and unstructured academic information delivery system has the potential to cause misinformation and misunderstanding among students. This condition drives the need for a chatbot system that is able to provide accurate academic information based on official sources. Furthermore, developed an academic helpdesk chatbot aimed at reducing the workload of administrative staff. The results of their research showed that the chatbot was able to provide fast and relevant responses, although it still faces challenges in system stability and adjustment to the academic context.

Furthermore, emphasized that academic chatbots play a crucial role in increasing the accessibility of campus information for students. The study demonstrated that the presence of chatbots can accelerate the information-seeking process and reduce students' reliance on manual confirmation from the administration. However, these studies are generally general in nature and have not specifically addressed the need for formal academic information at the study program level.

### Implementation of Retrieval-Augmented Generation (RAG) in Academic Chatbots

*Retrieval-Augmented Generation* (RAG) is an approach that combines the retrieval process, or searching for documents from authoritative sources, with the generative capabilities of natural language models. This approach was developed to overcome the limitations of purely generative models, which are prone to *hallucinations* , especially in domains that require high information precision.

through his survey study concluded that RAG-based chatbots are able to produce more relevant and reliable answers because they directly refer to official data sources. This approach has proven effective in educational contexts, where information accuracy is a major factor. Research similar to research shows that RAG-based chatbots not only increase the speed of information access, but also improve student understanding and engagement in online courses.

Furthermore, in the context of scientific research chatbots proved that the application of RAG significantly reduced the risk *of hallucination* and increased the transparency of answers. This finding strengthens the argument that RAG is very relevant to be applied to academic chatbots that handle administrative and procedural information. Other studies also confirm that the success of RAG-based academic chatbots is highly dependent on the quality of the dataset and the suitability of the documents to the institutional context.

The study shows that the RAG approach can only be used in general learning contexts, online courses, or global support services, although it has been widely used and shown good results. There has been no research specifically developing a RAG-based academic chatbot to meet the formal information needs of students across study programs, particularly in the Informatics Engineering Program at FTI UNISSULA. Therefore, this study focuses on the design and implementation of the chatbot using official study program documents as the primary source of knowledge.

**Gap Analysis**

Although various previous studies have discussed the use of chatbots in educational and academic service contexts, several gaps remain, indicating the need for further research. To clarify the position of this research, a gap analysis is presented in the following table:

**Table 1** Gap Analysis.

| No | Researchers and Years | Research Title | Gap Analysis |
|---|---|---|---|
| **1.** | Swacha & Gracel (2025) | *Chatbot-Based Academic Assistance in Higher Education* *Education* | Research has not integrated *the Retrieval-Augmented Generation* (RAG) approach, so chatbot responses are still potentially less accurate. contextual and prone to *hallucinations* |
| **2.** | Husain, Wibisono & Anisyah (2025 | *Evaluating AI Chatbots for Student Support* *Services* | The research focus is more on evaluating chatbot performance, without exploring *retrieval- based approaches* from official documents to improve the accuracy and reliability of answers. |
| **3.** | Warto et al. (2020) | Implementation Chatbot for Information Services Academic Based NLP | The technology used is still basic and does not utilize transformer models or dynamic integration with official institutional data sources. |

| | | | |
|---|---|---|---|
| **4.** | Lang & Gurpinar (2025) | *Effectiveness of RAGbased Chatbots in Online Courses* | The research focused on the context of online courses, not specifically directed at institutional academic information services such as lecture schedules, theses, and student administration. |

Table 1 shows that most previous studies still face limitations in terms of methods and scope of application. The Retrieval Augmented Generation (RAG) approach has not been used by several studies. However, studies that use this method usually focus on the context of online learning or courses rather than formal, institutional academic information services. Furthermore, no study has specifically examined RAG-based academic chatbots at the study program level, which use official internal documents as the primary source of knowledge. Therefore, a RAG-based academic chatbot was developed to meet the information needs of students in the Informatics Engineering Study Program, FTI UNISSULA, quickly, accurately, and contextually.

## 3. Research Methodology

This study uses an applied research approach *with* the aim of developing and implementing a *Retrieval Augmented Generation* (RAG)-based academic chatbot system to assist students in obtaining academic information quickly and accurately. The research methodology focuses on system development, integration of artificial intelligence models, and evaluation of the chatbot's performance in answering academic questions **.**

### System Integration

The developed academic chatbot system architecture consists of several main components: data sources, *retrieval modules* , generative language models, and user interfaces. The data sources are derived from official documents of the Informatics Engineering Study Program, FTI UNISSULA, such as academic guidelines, KRS provisions, Final Project requirements, and other administrative information.

*retrieval* stage , user queries are processed to retrieve the most relevant documents from the knowledge base using vector-based search techniques. The selected documents are then combined with the user query and provided as context to a generative language model. This model then generates natural, contextual, and authoritative source-based answers. The RAG approach was chosen because it combines the advantages of structured document search with the flexibility of a generative language model. The generative language model used in this study is a Large Language Model (LLM) integrated into a Retrieval-Augmented Generation scheme. In this scheme, the answer generation process is constrained by the context of the documents obtained from the retrieval process, thereby reducing the risk of false or inaccurate answers.

### Research Stages

The following algorithm explains the workflow of *the Retrieval-Augmented Generation (RAG) -based academic chatbot* used in this study. This algorithm is structured based on the methodological steps applied in the research report and illustrates the system's process in answering students' academic questions using official documents from the Informatics Engineering Study Program, FTI UNISSULA.

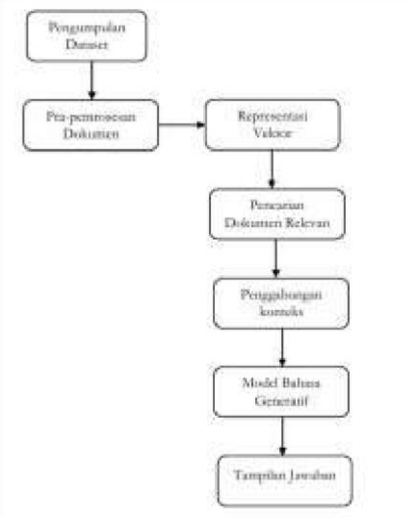


**Figure 1** Academic chatbot development workflow.

The workflow of *the Retrieval-Augmented Generation (RAG) based academic chatbot system* in this study is shown in the flow diagram and explained as follows:

1. **Dataset Collection**

   The initial stage begins with the collection of datasets in the form of official academic documents from the Informatics Engineering Study Program, FTI UNISSULA. These documents include academic guidelines, KRS (Student Study Plan) filling information, Final Project requirements, credit unit (SKS) requirements, and other administrative information that serve as the primary source of knowledge for the system.

2. **Document Pre-Processing**

   The collected documents then undergo a pre-processing stage. This stage aims to clean the data of unnecessary characters, standardize the text format, and divide the documents into text chunks to make them easier for the system to process.

3. **Vector Representation**

   The document preprocessing results are then converted into vector representations. This process allows the system to represent text numerically, allowing it to be used in semantic similarity-based search processes.

4. **Relevant Document Search**

   When a user asks a question, the system compares the question vector with the stored document vectors. Based on the similarity level of the vectors, the system retrieves the document or document fragment that is most relevant to the user's question.

5. **Context Merging**

   The acquired relevant documents are then combined with the user's query to form the input context. This context serves as the information foundation for the generative language model to generate appropriate answers.

6. **Generative Language Model**

   The generated context is processed by a generative language model. This model's task is to generate answers that are natural, contextual, and consistent with the information contained in official academic documents.

7. **Answer View**

   The final step is to display the system-generated answers to users via a chatbot interface. These answers are expected to provide fast, accurate, and easily understood academic information to students.

### System Evaluation

The system evaluation was conducted quantitatively using the ROUGE-1 and Bleu4 metrics, and testing was conducted qualitatively through an analysis of the suitability of answers to source documents related to KRS filling, Final Project requirements, and other academic administration procedures. Chatbot answers were compared with source documents to ensure the accuracy and suitability of the information. The evaluation results showed that the RAG implementation was able to produce contextual answers and reduce the risk of misinformation, although the quality of the responses still depended on the completeness and update of the academic documents used.

## 4. Results and Discussion

### Collected Dataset Results

The data collected in this study comes from 3 official academic documents of the faculty which are generally in PDF format consisting of: Final Project Guidebook, Academic Guide and Informatics Engineering PKK Guide as well as additional datasets in the form of FAQ Datasets which come from repetitive questions that are often asked by students through questionnaires.


**Figure 2** Academic Guide and FAQ dataset.

### Dataset Parsing and Preprocessing Results

**Table 2.** Preprocessing Results.

| No | Document Name | Document Contents |
|---|---|---|
| 1. | TA Guidebook | **Final Project Stages**<br>1. Stage of submitting the title/draft proposal for the Final Assignment and Supervisor Students who take the Final Assignment must first submit the title/draft proposal for the Final Assignment.<br>Procedures for Submitting Final Project Title/Draft Proposal and Supervisor:<br>a. The time for submitting the title/draft proposal can be at any time.<br>b. Final Project Courses have been entered into the Plan Card Study (KRS)<br>c. Register with the TA Admin Division by attaching the following requirements:<br>1) Photocopy of the KRS (Study Plan) that lists the Final Project Courses.<br>2) Print the Transcript of Grades, provided that the student has passed the MK with ≥130 credits and a cumulative GPA of ≥2.50.<br>3) TA proposal that has been approved by the TA coordinator.<br>4) fill in the proposal title/draft and supervisor submission form (in Adm. TA).<br>d. The form accompanied by the proposal is submitted to the TA coordinator. |

e. Prospective supervisors are determined by the TA coordinator and are adjusted to their respective scientific fields.

After determining the supervisor, students undertake a pre-proposal seminar guidance process to prepare for the proposal seminar.

| | | |
|---|---|---|
| **2.** | **Guide-Academic** | **Basic Understanding**<br>1. higher education under the Sultan Agung Waqf Foundation (YBWSA), which organizes academic, vocational and professional education programs in the field of science, including innovation, creation, application and development of science, technology and/or art.<br>2. Dharma Perguruan Tinggi and coordinate academic, vocational and/or professional education in one or a number of specific branches of science, technology and/or arts.<br>3. The curriculum is a set of plans and arrangements regarding graduate learning outcomes, graduate profiles, study materials, processes, and assessments used as guidelines for implementing study programs.<br>4. Graduate learning outcomes are a set of attitudes, knowledge, and skills that must be possessed, internalized, and mastered by students after studying a learning content, completing a program, or completing a particular educational unit.<br>5. Graduate competency standards are the minimum criteria for graduate ability qualifications which include attitudes, knowledge and skills stated in the formulation of graduate learning outcomes.<br>6. The intermediate semester is a unit of learning activities held between the even and odd semesters, equivalent to the regular semester according to the definition of semester credit units (sks).<br>7. Community Service Program (KKN) is a mandatory academic activity for undergraduate students which is carried out in a multidisciplinary manner in order to implement the knowledge they have to help solve problems in society. |

| | | |
|---|---|---|
| **3.** | **PKK Guide TIF** | **CHAPTER II PKK IMPLEMENTATION PROCEDURES**<br>**1. PKK Requirements**<br>Students who take the Group Work Project must meet the following requirements:<br>a) Students are registered as active students<br>b) Students have obtained ≥ 80 credits and GPA ≥ 2.50<br>c) Students have registered for the Group Work Project course on the Study Plan Card (KRS)<br>d) Students can show a list of grades/temporary transcripts that have been signed by the academic advisor.<br>e) Students pay the 4 credits load with a separate slip and submit it to the PKK Coordinator/Study Program Admin.<br>f) Students receive a Group Work Project Guidebook which can be obtained from the Group Work Project Administration.<br>g) Students fill out the Group Work Project application form and submit it to the Study Program admin. |
| **4.** | **FAQ_survey** | What are the requirements that must be met to register for the Proposal Seminar?<br>How to convert credits for those taking MBKM<br>What are the requirements to be able to take part in the Semhas?<br>What is the minimum credits required to take the Final Project?<br>Why do lecturers seem to care less about students?<br>What is the minimum number of credits to take the Final Project? What are the requirements for attending the proposal seminar? What are the requirements for registering for an internship?<br>What is the procedure for submitting a final project title and determining a supervisor?<br>Is it mandatory to take part in PKM?<br>What are the requirements for registering for the Proposal Seminar? What are the requirements for registering for KKN? Why can the number of credits sometimes change? Some credits can be 135, but it can become 133.<br>Is internship mandatory?<br>How many attitude points are needed to take the Final Assignment? |

The extraction phase is performed using the 'Pdflumber' library. After the text is successfully extracted, a *cleaning phase is performed* to remove unnecessary characters such as excess punctuation, symbols, irrelevant numbers, and double spaces. Next, case folding and text normalization are performed, converting all letters to lowercase to achieve a uniform text format. The text is then tokenized to break it down into smaller units called tokens, which can be words, phrases, or sentences, depending on the system's needs. All *preprocessing stages* that have been carried out produce *output* in the form of a file in the form of a .txt file.

**Chunking**

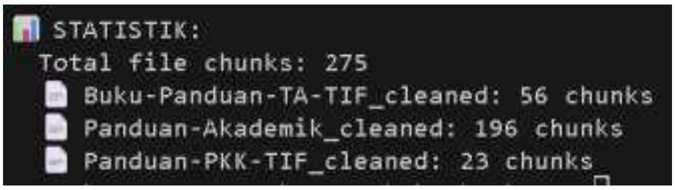


**Figure 3.** Dataset Chunking Results.

After *preprocessing* the data, the cleaned data will then be divided into small pieces called *chunks* with a size of around 300-500 words. This chunking process aims to facilitate further data processing, especially in the context of searching. Figures 4.3 and 4.4 show statistics on the number of chunks from three academic documents that have been processed, consisting of: Buku-Panduan- TATIF_cleaned: 56 chunks, Panduan-Akademik_cleaned: 196 chunks and Panduan-PKKTIF_cleaned : 23 chunks. The total chunks generated from all documents are 275 chunks. All chunks are stored in a folder named "chunks". This folder serves as a structured storage place for all chunked text pieces, ready to be used in the next stage.

**Embedding Process and Results**

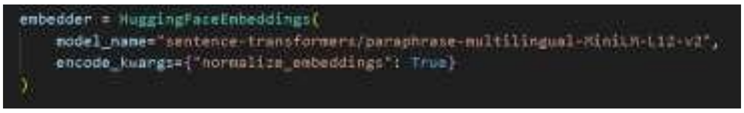


**Figure 4.** Embedding Creation.

Figure 4 above shows *the source code* for creating an academic document embedding. After the document has gone through parsing, preprocessing, tokenization, and chunking, each chunk of text has been converted into a numeric vector representation using the Sentence-BERT (SBERT) all-MiniLM-L12-v2 model. This model was chosen because of its optimal dimensions: 384 dimensions that are able to provide a balance between accuracy and speed, proven performance that has been validated in various *semantic search tasks, and support* for multilingualism which includes the ability to understand Indonesian context.


**Figure 5.** Embedding Results Folder.

Then the embedding results of academic documents are stored in the FAISS ( *Facebook AI Similarity Search* ) database, which functions as an indexed vector storage to facilitate semantic similarity-based searches . Inside the faiss_index folder are the index.faiss and index.pkl files of this database to help the system quickly identify and retrieve the most relevant pieces of information based on semantic similarity to the user's query.

**Grounded Dataset Results FAQ**

**Table 3.** Grounded FAQ and Document Results.

| | |
|---|---|
| **1** | **question,frequency,similarity_score,document_chunk,document_answer** |
| **2** | What are the requirements to be able to take part in the Semhas,37,0.526,-,No relevant references were found in official documents. |

| | |
|---|---|
| **3** | What are the requirements for registering for KKN,12,0.756,Academic-Guide_chunk_085.txt,"requirements for participating in KKN:<br>1. has reached 110 credits and<br>2. Obtain a Letter of Satisfaction (SP) for the Internship from the Internship Coordinator. The procedures and policies for implementing this KKN are determined by the KKN Coordinator at the LPPM level.<br>1. Final assignments can be taken by students who have participated in internships.<br>(KP)/ Group Work Practice (PKK)<br>2. have reached a minimum of 120 credits<br>3. with an IP ≥ 2.75 without an E grade for the Electrical Engineering and Informatics Engineering Study Programs, and a minimum of 130 credits with an IP ≥ 2.75<br>4. without an E grade for the Industrial Engineering Study Program. After completing each activity" |
| **4** | What are the requirements for registering for the Proposal Seminar, 10, 0.751, TA_chunk_126.txt Guidebook, Pre-proposal seminar guidance and TA Final Assignment must be done at least 4 times to meet the registration requirements |
| **5** | What is the minimum number of credits to take the Final Assignment,8,0.785,TA_chunk_008.txt Guidebook,"Minimum number of credits to take the Final Assignment: Students can take the TA (Final Assignment) course if they have completed at least 130 credits" |
| **6** | When will the KRS filling be opened,8,0.507,-,No relevant references were found in official documents. |
| **7** | How to convert credits for those who take MBKM,5,0.835,PanduanAkademik_chunk_072.txt,"how to convert credits for those who take MBKM:<br>Students who participate in the MBKM program can convert credits by fulfilling the conditions set by each faculty or study program.<br>1. taking MBKM courses<br>2. Attach proof of participating in the MBKM program to the Unissula SIM management" |
| **8** | How to apply for UKT payment dispensation,3,0.587,-,No relevant references were found in official documents. |

The FAQ grounding process is carried out by linking frequently asked questions to official documents as sources of answers. FAQ questions and official document snippets are converted into vector representations using a multilingual *SentenceTransformer model* , then the semantic similarity level is measured using *cosine similarity* . The document with the highest similarity score is selected as the answer reference if its value exceeds a specified threshold. If no sufficiently relevant documents are found, the system explicitly states that there are no suitable references. This approach ensures that the generated FAQ answers are factual, verified, and based on official documents.

**Query Processing and Document Retrieval**

The process of finding relevant documents is a core part of *the Retrieval Augmented Generation (RAG) architecture* because it determines the quality of the context used in answer generation. At this stage, the system does not directly generate answers but first searches for relevant information from official academic documents. The process of finding relevant documents based on a user's query is shown in Figure 5 below:

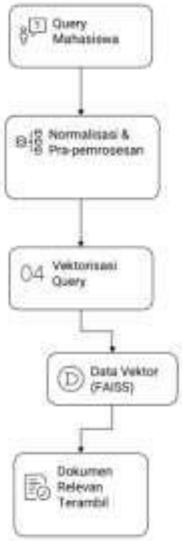


**Figure 5.** Document Retrieval Process Flow.

Figure 5 The flow begins when a student submits a question through the chatbot, which is then processed and converted into a vector representation using a *sentence embedding model* . The question vector is then compared with document vectors stored in a vector database to measure the level of semantic similarity. The document with the highest similarity value is selected as the relevant document and used as the main context in the *generation stage* . With this mechanism, the chatbot generates answers that are not only linguistically natural but also refer to official sources of information, thereby reducing the risk of misinformation and answer hallucinations.

**Answer Generation Process**

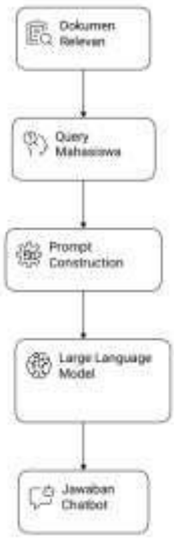


**Figure 6.** Answer Generation Process Flow in Chatbot.

Figure 6 shows the answer generation process carried out after the system obtains relevant documents from the *retrieval stage* . The answer generation process flow is shown at this stage, where the retrieved relevant documents are combined with student questions in the form of structured *prompts* . These prompts are then provided as input to *the Large Language Model* (LLM) to generate contextual answers that align with official documents. With this mechanism, the LLM does not generate answers freely, but is instead constrained by the context derived from valid academic information sources.

**Implementation of RAG-Based Academic Chatbot**


**Figure 7.** Academic Chatbot Interface.

Figure 7 displays the main interface of the academic chatbot system "TIF-AcadGuide" developed based on *Retrieval-Augmented Generation (RAG)* . This interface serves as a medium for interaction between students and the system in obtaining academic information quickly and in a structured manner. The left-side panel displays the system status indicating the readiness of the knowledge base and language model, indicating that official academic documents have been successfully loaded and are ready for use in the *retrieval* and *generation process* . The main part of the interface provides an academic question input field that students use to interact with the system. The entered questions will be processed through the stages of question processing, searching for relevant documents, and generating answers using a language model. The resulting answers are then displayed back to the user through this interface, so students can obtain academic information directly without having to manually browse documents or wait for confirmation from the administration.

**Results of User Interaction with Chatbot**


**Figure 8.** Interaction with chatbot.

Figure 8 shows a real interaction between a user and the TIF-Acadguide system, where a student named Ikharista asked the question "What are the requirements for a proposal seminar?" The system provided a complete answer that included four main requirements, namely having taken the Final Project course in the KRS, having completed at least 130 credits with a GPA ≥ 2.50, having a proposal approved by the supervisor, and completing the administrative files for the proposal seminar. The source of the answer came from the FAQ dataset clearly marked as "Academic FAQ", the information is from frequently asked questions and has been verified not generated by generative AI. This interaction proves the system's ability to understand specific academic questions and provide structured, accurate, and easily understood responses for students.

**Evaluation Results**

**Table 4.** Testing and evaluation results.

| No | Evaluation Aspects | FAQ | Non-FAQ | Information |
|---|---|---|---|---|
| 1 | Number of questions | 25 | 25 | Total 50 questions |
| 2 | Success Rate (100%) | 100% | 100% | System Not Failed to Answer |
| 3 | ROUGE-1 F1 | 0.957 | 0.057 | Similarity of answer unigrams |
| 4 | BLEU-4 Score | 0.877 | 0.024 | Precision n-gram |
| 5 | Average Length answer (character) | 124 | 128 | *Output* length stability |
| 6 | Source of answer dominant | FAQ Dataset | Document Academic | According to the priority system |
| 7 | Level of suitability of answers | Very high | Low-medium | Influenced by document editing |

The academic chatbot performance evaluation table shows a significant difference in performance between FAQ and Non-FAQ category questions. Of the total 50 questions tested, consisting of 25 questions in the FAQ category and 25 in the Non-FAQ category, all questions were successfully answered by the system with a 100% success rate. This shows that functionally, the system can respond to all input without experiencing process failures, both taken from the FAQ database and from academic documents, so that the stability of the system in providing answers can be considered very good.

However, the quality of the answer suitability measured using the ROUGE-1 and BLEU-4 metrics shows a significant difference. In FAQ questions, the ROUGE-1 value reaches 0.957 and BLEU-4 is 0.877, which indicates a very high level of lexical similarity and n-gram precision between the chatbot answers and the reference answers. This is due to the characteristics of FAQs which have short and structured answers, so that the system is able to produce answers that are almost identical to the reference. In contrast, in Non-FAQ questions, the ROUGE-1 and BLEU-4 values are relatively low because the answers are generated through a retrieval process and answer generation from long and narrative academic documents. Differences in editorial, paraphrasing, and the use of academic terms cause a decrease in the lexical similarity score, although semantically the answers remain relevant.

The difference in performance between FAQ and non-FAQ questions indicates that the characteristics of the information source significantly influence the evaluation using word comparison. Low ROUGE and BLEU scores on non-FAQ questions do not indicate system failure, but rather that academic documents are typically long and narrative, resulting in paraphrased answers. The answers remain relevant and closely match the original document, suggesting that the RAG approach is better evaluated using a combination of quantitative metrics and qualitative analysis.

## 5. Conclusion

This study successfully designed and implemented a *Retrieval-Augmented Generation* (RAG)-based academic chatbot for the Informatics Engineering Study Program, Faculty of Technology, University of Surabaya (FTI), utilizing official academic documents as the primary source of knowledge. The results showed that the integration of a semantic search-based *retrieval mechanism* with *a Large Language Model* was able to produce contextual, accurate, and verified answers, while reducing the risk of *hallucinations* commonly found in pure generative models. System evaluation proved that the chatbot had high stability with a success rate of answering questions reaching 100%, and excellent performance on FAQ questions.

The findings of this study support the research objective, which is to provide an efficient academic information access solution for students and reduce reliance on informal communication and the administrative burden of study programs. The application of RAG proved relevant in the context of institutional academic information services that demand high precision. The main contribution of this study lies in providing an implementation model for RAG-based academic chatbots at the study program level in Indonesian universities. However, this study has limitations in the scope of the documents used, so some questions do not have adequate references. Further research is recommended to expand and update the dataset regularly, integrate user *feedback mechanisms* , and conduct evaluations with additional human-based metrics to measure the quality of answers more comprehensively.

**Author Contributions:** Conceptualization: IANA and SFCH; Methodology: IANA; Software: IANA; Validation: IANA and SFCH; Formal analysis: IANA; Investigations: IANA; Resources: IANA; Data curation: IANA; Writing original draft preparation: IANA; Writing review and editing: IANA and SFCH; Visualization: IANA; Supervision: SFCH; Project administration: SFCH

**Funding:** This research received no external funding.

**Data Availability Statement:** The data used in this study consist of internal academic documents and questionnaire-based FAQ data from the Informatics Engineering Study Program, FTI UNISSULA. Due to institutional policy and privacy considerations, the datasets are not publicly available but can be accessed upon reasonable request to the corresponding.

**Acknowledgments:** The authors would like to thank the Informatics Engineering Study Program, Faculty of Industrial Technology, Sultan Agung Islamic University, for providing academic documents and administrative support during this research. The authors also acknowledge the use of AI-based tools to support system development and analysis.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

A. Kurniawan, A. Abdiansah, and A. Syahrini, "NL2SQL for Chatbot with Semantic Parsing Using Rule-Based Methods," vol. 5, no. 1, pp. 39-48, 2024. https://doi.org/10.36706/sjia.v5i1.66

C. Antico, S. Giordano, C. Koyuturk, and D. Ognibene, "Unimib Assistant: designing a student-friendly RAG-based chatbot for all their needs," vol. 7400, 2024.

D. Kristanto et al., "Pengembangan Chatbot Layanan Informasi Kampus Menggunakan TF-IDF," pp. 103-115, 2025, doi: 10.33364/algoritma/v.22-2.2350. https://doi.org/10.33364/algoritma/v.22-2.2350

D. M. Alfiansyah, L. Setiyani, and D. F. Wati, "Pengembangan Chatbot Berbasis Web untuk Layanan Informasi di Horizon," vol. 7, no. 3, 2025, doi: 10.32877/bt.v7i3.2318. https://doi.org/10.32877/bt.v7i3.2318

G. D. Albert and A. Voutama, "Pengembangan Chatbot Berbasis Pdf Menggunakan Local Retrieval-Augmented Generation (Rag) Dan Ollama," J. Inform. dan Tek. Elektro Terap., vol. 13, no. 2, 2025, doi: 10.23960/jitet.v13i2.6361. https://doi.org/10.23960/jitet.v13i2.6361

G. Lang and T. Gurpinar, "AI-Powered Learning Support: A Study of Retrieval-Augmented Generation (RAG) Chatbot Effectiveness in an Online Course," Inf. Syst. Educ. J., vol. 23, no. 2, pp. 4-13, 2025, doi: 10.62273/zklk5988. https://doi.org/10.62273/ZKLK5988

I. Ichsanudin Rachman Pratama and B. Sisephaputra, "Pengembangan Sistem Helpdesk Menggunakan Chatbot Dengan Metode Retrieval-Augmented Generation (Rag)," J. Informatics Comput. Sci., vol. 6, no. 03, pp. 696-710, 2024, doi: 10.26740/jinacs.v6n03.p696-710. https://doi.org/10.26740/jinacs.v6n03.p696-710

I. Ortiz-Garces, J. Govea, R. O. Andrade, and W. Villegas-Ch, "Optimizing Chatbot Effectiveness through Advanced Syntactic Analysis: A Comprehensive Study in Natural Language Processing," Appl. Sci., vol. 14, no. 5, 2024, doi: 10.3390/app14051737. https://doi.org/10.3390/app14051737

J. L. Perez, A. Gupta, and M. Chen, "Design and implementation of academic service chatbots using large language models," *Education and Information Technologies*, 2024, doi: 10.1007/s10639-024-12345-6.

J. Prayoga, F. R. S. Br Ginting, K. Siregar, N. Ramadani, and R. R. Al Hafiz, "Analisis Audit Sistem Informasi Absensi Pada Stmik Kaputama Menggunakan Framework Cobit-5," War. Dharmawangsa, vol. 19, no. 1, pp. 180-187, 2025, doi: 10.46576/wdw.v19i1.5823. https://doi.org/10.46576/wdw.v19i1.5823

J. Swacha and M. Gracel, "Retrieval-Augmented Generation (RAG) Chatbots for Education: A Survey of Applications," Appl. Sci., vol. 15, no. 8, 2025, doi: 10.3390/app15084234. https://doi.org/10.3390/app15084234

J. Tahsinia, A. A. Zulfa, T. Ibrahim, and O. Arifudin, "Peran sistem informasi akademik berbasis web," vol. 6, no. 1, pp. 115-134, 2025.

L. R. Hidayat, I. G. Pasek, S. Wijaya, and R. Dwiyansaputra, "Optimalisasi Layanan Sistem Informasi Mahasiswa dengan Integrasi Telegram: Chatbot Retrieval-Augmented-Generation berbasis Large Language Model," vol. 7, no. 1, pp. 121-131, 2025. https://doi.org/10.29303/jtika.v7i1.459

M. Amin, K. Nazik, and A. Salwa, "Interacting with educational chatbots: A systematic review," Springer US, 2023. doi: 10.1007/s10639-022-11177-3.

M. Izacard and E. Grave, "Leveraging passage retrieval with generative models for open domain question answering," *Proceedings of the 16th EACL*, pp. 874-880, 2021. https://doi.org/10.18653/v1/2021.eacl-main.74

M. L. Husain, Y. Wibisono, and A. Anisyah, "Development of an Academic Services Chatbot Based on Retrieval-Augmented Generation (RAG)," vol. 5, no. 2, pp. 727-735, 2025. https://doi.org/10.47709/brilliance.v5i2.6719

P. Lewis et al., "Retrieval-Augmented Generation for knowledge-intensive NLP tasks," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 9459-9474, 2020.

S. Elysia, "Technology Information and Data Analytic Chatbot Berbasis Retrieval Augmented Generation (RAG) untuk Peningkatan Layanan Informasi Sekolah," vol. 1, no. 2, pp. 52-58, 2024. https://doi.org/10.70491/tifda.v1i2.52

S. Gao, Y. Chen, K. Ding, and J. Han, "Retrieval-Augmented Generation for large language models: A survey," *ACM Computing Surveys*, 2024, doi: 10.1145/3638673.

X. Wang, J. Wei, and Y. Zhou, "Grounding large language models with retrieval-augmented generation: A review," *Information Processing & Management*, vol. 61, no. 1, 2024, doi: 10.1016/j.ipm.2023.103421. https://doi.org/10.1016/j.ipm.2023.103421