Journal of Informatics and Computer Engineering Research (JICER) ISSN: 3089-6320
Vol. 2, No. 2, Desember 2025, hlm. 14-20

WEB-BASED PUBLIC RELATIONS CHATBOT USING LARGE LANGUAGE MODELS AND THE RETRIEVAL-AUGMENTED GENERATION

Andi Muhammad Nuzul[1], Muhammad Nur Yasir Utomo[2], Tantri Indrabulan[3]
[1,2,3] Department of Informatics and Computer Engineering, Politeknik Negeri Ujung Pandang, Indonesia

Article Info

Article history:
Received, (20 November 2025)
Revised, (2 Desember 2025)
Accepted, (4 Desember 2025)

Keywords:
Web-based chatbot;
Large language models;
Retrieval-augmented generation;
Natual language processing;
Vectorscore

ABSTRACT

The Public Relations and Protocol Working Group (Pokja Humas) of Politeknik Negeri Ujung Pandang (PNUP) faces challenges in providing interactive and responsive information services. The official website functions only as a one-way medium, and the high volume of repeated questions causes delays in response time. This study developed a public relations chatbot based on Large Language Model (LLM) using the Retrieval-Augmented Generation (RAG) method to improve information services. The chatbot data were obtained through web scraping of the PNUP official website and internal PDF documents, which were processed through preprocessing, text splitting, and embedding using Hugging Face and stored in a FAISS vectorstore. The system was built using FastAPI as the backend and web-based interfaces for admin and user interactions. The results show that User Acceptance Test (UAT) involving 35 respondents achieved 91.93% acceptance (very good). The Retrieval-Augmented Generation Assessment (RAGAS) evaluation achieved average scores of 0.89 for Faithfulness, 0.91 for Answer Relevancy, 0.89 for Context Precision, and 0.89 for Context Recall, indicating that the chatbot produced relevant and contextually accurate answers.

ABSTRAK

Kelompok Kerja Hubungan Masyarakat dan Protokoler Politeknik Negeri Ujung Pandang (PNUP) menghadapi tantangan dalam memberikan layanan informasi interaktif. Website resmi yang dikelola masih bersifat satu arah dan volume pertanyaan calon mahasiswa yang tinggi sering menyebabkan keterlambatan respons. Penelitian ini bertujuan mengembangkan chatbot kehumasan berbasis Large Language Model (LLM) dengan metode Retrieval Augmented Generation (RAG) untuk mendukung layanan informasi PNUP. Data chatbot diperoleh dari hasil web scraping website resmi dan dokumen PDF internal Humas PNUP. Data tersebut melalui tahap preprocessing, splitting, dan embedding menggunakan Hugging Face kemudian disimpan dalam vectorstore berbasis FAISS. Sistem dikembangkan menggunakan FastAPI sebagai backend dengan halaman web untuk admin dan pengguna. Hasil pengujian menunjukkan User Acceptance Test (UAT) dengan 35 responden menghasilkan tingkat penerimaan sebesar 91,93% (kategori sangat baik). Evaluasi menggunakan Retrieval-Augmented Generation Assessment (RAGAS) menghasilkan skor rata-rata sebesar 0,89 untuk Faithfulness, 0,91 untuk Answer Relevancy, 0,89 untuk Context Precision, dan 0,89 untuk Context Recall, yang menunjukkan bahwa chatbot mampu menghasilkan jawaban yang relevan dan akurat secara kontekstual.

Penulis Korespondensi:
Andi Muhammad Nuzul
Department of Informatics and Computer Engineering, Politeknik Negeri Ujung Pandang, Jl. Perintis Kemerdekaan KM. 10, Makassar, Indonesia
Email: amuhnuzul@gmail.com

1. INTRODUCTION

Politeknik Negeri Ujung Pandang (PNUP) is a vocational higher education institution under the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia [1]. One of the key units responsible for information dissemination at PNUP is the Public Relations and Protocol Working Group (Pokja Humas) . This unit manages the official website, social media platforms, and external campus communications.

Pokja Humas faces several challenges, particularly limited service hours and the high volume of repetitive inquiries submitted by prospective students through social media and email. These issues often lead to delays in delivering information and may reduce the quality of public service. The demand for timely responses continues to increase, while the existing communication workflow still relies heavily on manual handling.

Previous research by Amrullah demonstrated the application of Natural Language Processing (NLP)-based chatbots to support academic services [2], and Rahmawati implemented a similar approach for new student admissions [3]. However, conventional NLP systems remain limited because they depend on predefined response patterns. Such systems are less suitable for dynamic institutional environments where information is updated frequently and user questions vary widely [4].

Recent advancements in Large Language Models (LLMs) such as GPT, LLaMA, and DeepSeek have significantly improved natural language understanding, enabling more flexible and context-aware responses [4]. Nevertheless, LLMs are prone to generating hallucinations or producing answers that are not grounded in actual source data. To address this issue, the Retrieval-Augmented Generation (RAG) method was developed as an approach that integrates the generative capabilities of LLMs with context retrieval from external knowledge sources [5]. This approach is particularly relevant for PNUP, where official information is distributed across web pages and internal documents, requiring a mechanism that can retrieve accurate content while maintaining natural language fluency.

Given these challenges and technological developments, this study focuses on developing a public relations chatbot for PNUP using a locally deployed LLM with the RAG architecture. The system is designed to provide users with fast, accurate, and contextually relevant responses by retrieving information directly from curated institutional data sources. This approach not only enhances service efficiency but also supports the institution's efforts to modernize its information delivery ecosystem.

To address the aforementioned challenges and limitations, this research provides the following key contributions:

- Development of a web-based public relations chatbot for PNUP using a locally deployed Retrieval-Augmented Generation (RAG) pipeline, enabling accurate and context-aware responses grounded in institutional information from both official web pages and internal documents.
- Construction of an institution-specific knowledge base through a structured preprocessing workflow, including web scraping, PDF parsing, text splitting, and vector-based embedding to support efficient and relevant information retrieval.
- Comprehensive evaluation of the proposed system using Blackbox Testing, User Acceptance Test (UAT), and RAGAS metrics, enabling a multidimensional assessment of system reliability, user satisfaction, and contextual answer quality.

This study presents a practical implementation of LLM and RAG technologies to enhance public information services in higher education. The following section describes the research methodology, including data collection, system design, implementation, and evaluation processes.

2. METHODOLOGY

The research was conducted within the Public Relations and Protocol Working Group (Pokja Humas) at PNUP. The research stages consisted of several parts, namely literature review and interviews, data collection, system design, system development, and system testing. The stages of the research are shown in Figure 1.

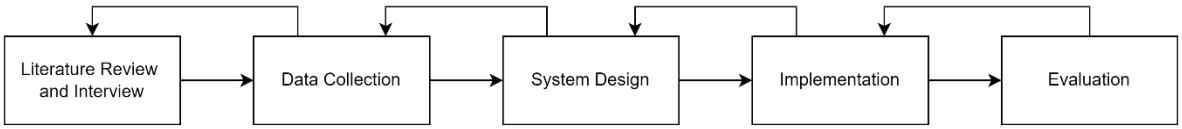

Figure 1. System Architecture

2.1. Literature Review and Interview

The initial stage involved studying theories related to chatbots, Large Language Models (LLMs), and the Retrieval-Augmented Generation (RAG) method. In addition, interviews were conducted with the PNUP Public Relations and Protocol Working Group (Pokja Humas) to identify the system requirements.

2.2. Data Collection

The data used in this study were obtained from two main sources:

- Web scraping, in which data were collected from the official PNUP website (https://poliupg.ac.id), including pages such as news, announcements, and departmental information. The scraped data were stored in JSON format to facilitate further processing.
- Internal documents, consisting of PDF files provided by the PNUP Public Relations and Protocol Working Group (Pokja Humas), containing official information related to the institution.

2.3. System Design

The system design was carried out using the Retrieval-Augmented Generation architecture, which consists of three main components: the retriever, the generator, and the vector store. The retriever component is responsible for retrieving relevant context from the data stored in the vector database, while the generator produces answer texts using the LLM model. The RAG architecture was implemented locally using the DeepSeek-R1 14B model executed through Ollama.

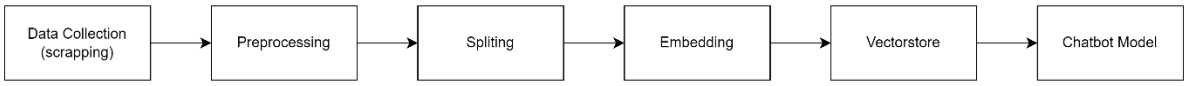

Figure 2. Chatbot Model Development

The system was developed using the Python programming language with the FastAPI framework as the backend. The RAG pipeline integration was implemented using the LangChain framework as shown in Figure 2 [6]. The implementation stages included:

- Preprocessing the data to remove irrelevant characters.
- Splitting the text into segments of 2,000 characters with an overlap of 100 characters.
- Embedding the documents using the intfloat/multilingual-e5-large model from Hugging Face [7].
- Storing the embedding results in a FAISS (Facebook AI Similarity Search) vector store [8].
- Integrating the DeepSeek-R1 model as the generator to produce context-based answers [4].

The system consists of two interfaces: an admin page for managing documents and building the vector store, and a user page for interacting directly with the chatbot.

2.4. Implementation

After the chatbot model created, the chatbot system delivered to user through API and web interface. The flow of implementing the model to web is shown in Figure 3.

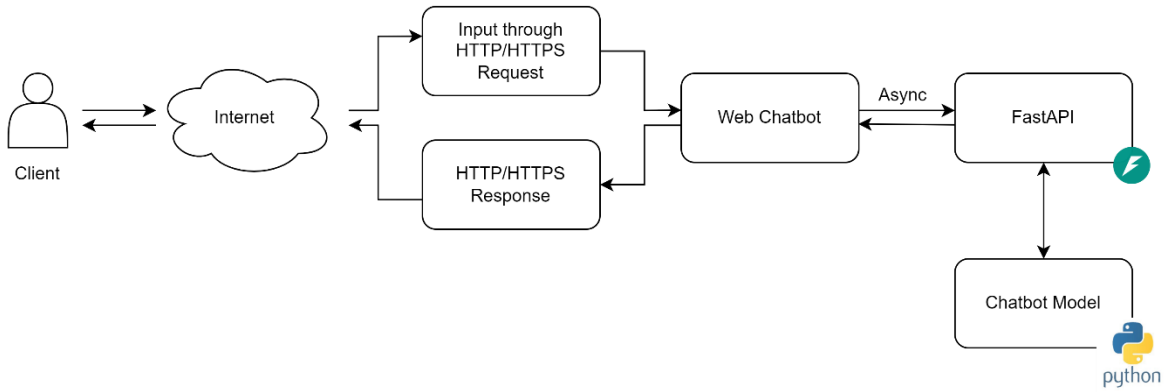

Figure 3. Chatbot Model Implementation on Web

The implementation of the system follows a client–server architecture in which users interact with the web-based chatbot through HTTP or HTTPS requests over the Internet. When a user submits a query via the chatbot interface, the request is forwarded to the backend service built using FastAPI, which serves as the main application layer responsible for managing communication between the frontend and the chatbot model. FastAPI processes the incoming request, executes the Retrieval-Augmented Generation workflow within the Python-based chatbot model, and retrieves the appropriate context and generated response. The resulting output is then returned to the user through an HTTP/HTTPS response, enabling seamless and real-time interaction between the user and the web chatbot.

2.4. Evaluation

Three testing methods were used to evaluate the system:

- User Acceptance Test (UAT), which was carried out with 35 respondents using a 1–5 Likert scale to assess usability, usefulness, and user satisfaction [9], [10].
- Retrieval-Augmented Generation Assessment (RAGAS), which was used to measure the quality of the chatbot's answers based on four key metrics: Faithfulness, Answer Relevancy, Context Precision, and Context Recall [11].

3. RESULTS AND DISCUSSION

This research resulted in a web-based chatbot designed to answer questions related to Politeknik Negeri Ujung Pandang (PNUP), built using Large Language Models (LLMs) with the Retrieval-Augmented Generation (RAG) method.

3.1. Implementation Result

The chatbot system was developed with two main interfaces. The admin page is used for uploading PDF documents, building the vector store, and monitoring user conversation history. Meanwhile, the user page, shown in Figure 4, provides an interactive web-based chat interface that allows users to submit questions and provide feedback.

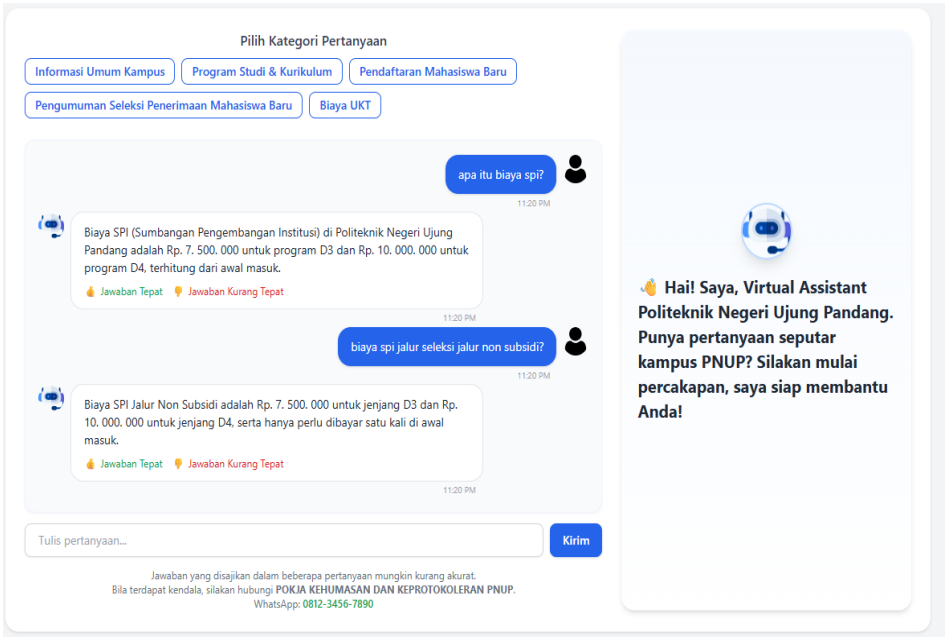

Figure 4. Chatbot Model Implementation on Web

The DeepSeek-R1 model was executed locally through Ollama and integrated with the LangChain pipeline to manage the retrieval and generation processes. This system integration enables the chatbot to produce responses that are aligned with the context of the source documents.

3.2. User Acceptance Test Result

The UAT results from 35 respondents, consisting of PNUP students from the 2025 cohort, showed an acceptance rate of 91.93%, which falls into the 'very good' category. Respondents indicated that the system was easy to use, had a simple interface, and was capable of providing the information they needed as shown in Table 1.

Tabel 1. User Acceptance Test Result

Question	Answer Options	Weight (N)	Frequency (F)	F x N
1.	Strongly Agree	5	24	120
	Agree	4	7	28
	Fairly Agree	3	4	12
	Disagree	2	0	0
	Strongly Disagree	1	0	0
	Total Score		35	160

Question	Answer Options	Weight (N)	Frequency (F)	F x N
	Keakuratan			91,42%
2.	Strongly Agree	5	21	105
	Agree	4	12	48
	Fairly Agree	3	2	6
	Disagree	2	0	0
	Strongly Disagree	1	0	0
	Total Score		35	159
	Accuracy			90,85%
3.	Strongly Agree	5	20	100
	Agree	4	13	52
	Fairly Agree	3	1	3
	Disagree	2	1	2
	Strongly Disagree	1	0	0
	Total Score		35	157
	Accuracy			89,71%
4.	Strongly Agree	5	25	125
	Agree	4	10	40
	Fairly Agree	3	0	0
	Disagree	2	0	0
	Strongly Disagree	1	0	0
	Total Score		35	165
	Accuracy			94,28%
5.	Strongly Agree	5	27	135
	Agree	4	5	20
	Fairly Agree	3	3	9
	Disagree	2	0	0
	Strongly Disagree	1	0	0
	Total Score		35	164
	Accuracy			93,71%
6.	Strongly Agree	5	24	120
	Agree	4	7	28
	Fairly Agree	3	3	9
	Disagree	2	1	2
	Strongly Disagree	1	0	0
	Total Score		35	159
	Accuracy			90,85%
7.	Strongly Agree	5	25	125
	Agree	4	7	28
	Fairly Agree	3	3	12
	Disagree	2	0	0
	Strongly Disagree	1	0	0
	Total Score		35	165
	Accuracy			94,28
8.	Strongly Agree	5	21	105
	Agree	4	10	40
	Fairly Agree	3	4	12
	Disagree	2	0	0
	Strongly Disagree	1	0	0
	Total Score		35	157
	Accuracy			89,71
9.	Strongly Agree	5	25	125
	Agree	4	7	28
	Fairly Agree	3	3	9
	Disagree	2	0	0
	Strongly Disagree	1	0	0
	Total Score		35	162
	Accuracy			92,57

Question	Answer Options	Weight (N)	Frequency (F)	F x N
10.	Strongly Agree	5	24	120
	Agree	4	8	32
	Fairly Agree	3	3	9
	Disagree	2	0	0
	Strongly Disagree	1	0	0
	Total Score		35	161
	Accuracy			92%
Average				**91,93%**

3.2. Retrieval Augmented Generation Assessment Result

The evaluation using RAGAS produced a Faithfulness score of 0.89, an Answer Relevancy score of 0.91, a Context Precision score of 0.89, and a Context Recall score of 0.89 as shown in Table 2. These results indicate that the system is capable of generating responses that are relevant, contextually consistent, and exhibit minimal hallucination.

Tabel 2. Retrieval Augmented Generation Assessment Result

Question Test	Faithfulness	Answer relevancy	Context precision	Context recall
1.	0,90	0,91	0,90	0,86
2.	0,88	0,94	0,88	0,86
3.	0,91	0,91	0,91	0,87
4.	0,92	0,86	0,92	0,93
5.	0,92	0,90	0,92	0,87
6.	0,87	0,90	0,87	0,91
7.	0,86	0,92	0,86	0,85
8.	0,90	0,86	0,90	0,94
9.	0,87	0,93	0,87	0,86
10.	0,85	0,92	0,85	0,85
11.	0,86	0,96	0,86	0,88
12.	0,93	0,93	0,93	0,93
13.	0,95	0,93	0,95	0,95
14.	0,90	0,90	0,90	0,87
15.	0,87	0,94	0,87	0,87
16.	0,89	0,90	0,89	0,88
17.	0,92	0,90	0,92	0,93
18.	0,90	0,91	0,90	0,93
19.	0,89	0,90	0,89	0,86
20.	0,90	0,94	0,90	0,90
21.	0,91	0,94	0,91	0,92
22.	0,93	0,95	0,93	0,91
23.	0,92	0,94	0,92	0,94
24.	0,91	0,94	0,91	0,86
25.	0,87	0,90	0,87	0,91
26.	0,86	0,92	0,86	0,87
27.	0,88	0,89	0,88	0,87
28.	0,90	0,90	0,90	0,92
29.	0,87	0,88	0,87	0,87
30.	0,86	0,89	0,86	0,85
Average	**0,89**	**0,91**	**0,89**	**0,89**

4. CONCLUSION

This research produced a public relations chatbot for PNUP based on a Large Language Model (LLM) using the Retrieval-Augmented Generation (RAG) method. The system has been demonstrated to improve information services at PNUP, achieving a user acceptance rate of 91.93%, while the RAGAS evaluation results indicate high answer quality. Future development is recommended to optimize response time and expand system integration to other platforms such as WhatsApp or Telegram to enable broader public use.

REFERENCES

[1] M. N. Y. Utomo, E. Tungadi, Muh. Ahyar, and Muh. Irsan, "Design of internal audit system to support accreditation process in higher vocational education institutions," presented at the The 2nd Makassar Conference of Applied Sciences (MCAS): Synergizing Research and Innovation for Mitigating Climate Change, Makassar, Indonesia, 2025, p. 040026. doi: 10.1063/5.0297547.

[2] A. Z. Amrullah, A. S. Anas, and G. Primajati, "Implementasi Chatbot sebagai Virtual Assistant Penerimaan Mahasiswa Baru pada Universitas Bumigora," *J. Bumigora Inf. Technol. BITe*, vol. 4, no. 1, pp. 17–26, June 2022, doi: 10.30812/bite.v4i1.1664.

[3] H. Rahmawati and A. Sudrajat, "IMPLEMENTASI CHATBOT PADA PENERIMAAN MAHASISWA BARU DI POLITEKNIK TEDC BANDUNG MENGGUNAKAN NATURAL LANGUAGE PROCESSING," *J. Inform. Dan Tek. Elektro Terap.*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5456.

[4] M. N. Y. Utomo, T. B. Adji, and I. Ardiyanto, "Geolocation Prediction in Social Media Data using Text Analysis: A Review," in *2018 International Conference on Information and Communications Technology (ICOIACT)*, Mar. 2018, pp. 84–89. doi: 10.1109/ICOIACT.2018.8350674.

[5] A. E. and M. L., "An Overview of Chatbot Technology," in *Advances in Information and Communication Technology*, Springer. Accessed: Nov. 17, 2025. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-49186-4_31

[6] 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, Apr. 2025, doi: 10.23960/jitet.v13i2.6361.

[7] M. D. A. Muhajir, N. Prastiti, and M. Koeshardianto, "IMPLEMENTASI CHATBOT MENGGUNAKAN FRAMEWORK LANGCHAIN BERBASIS LLM GPT," *JATI J. Mhs. Tek. Inform.*, vol. 9, no. 2, pp. 2151–2158, 2025.

[8] V. Jayalakshmi and M. Lakshmi, "Twitter Sentiment Analysis Tweets Using Hugging Face Harnessing NLP for Social Media Insights," in *Advancements in Smart Computing and Information Security*, S. Rajagopal, K. Popat, D. Meva, and S. Bajeja, Eds., Cham: Springer Nature Switzerland, 2024, pp. 378–389. doi: 10.1007/978-3-031-59097-9_28.

[9] J. Johnson, M. Douze, and H. Jégou, "Billion-Scale Similarity Search with GPUs," *IEEE Trans. Big Data*, vol. 7, no. 3, pp. 535–547, July 2021, doi: 10.1109/TBDATA.2019.2921572.

[10] W. Wulandari, N. Nofiyani, and H. Hasugian, "User Acceptance Testing (UAT) pada electronic data preprocessing guna mengetahui kualitas sistem," *J. Mhs. Ilmu Komput.*, vol. 4, no. 1, pp. 20–27, 2023.

[11] H. Yakub, B. Daniawan, A. Wijaya, and L. Damayanti, "Sistem Informasi E-Commerce Berbasis Website Dengan Metode Pengujian User Acceptance Testing," *J. Sist. Inf. Dan Teknol. Inf. Komput.*, vol. 2, no. 2, 2024, doi: https://doi.org/10.53624/jsitik.v2i2.362.

[12] S. Es, J. James, L. Espinosa Anke, and S. Schockaert, "RAGAs: Automated Evaluation of Retrieval Augmented Generation," in *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, N. Aletras and O. De Clercq, Eds., St. Julians, Malta: Association for Computational Linguistics, Mar. 2024, pp. 150–158. doi: 10.18653/v1/2024.eacl-demo.16.