International Journal of Informatics, Information System and Computer Engineering

AI-Based Chatbot System for Education and Recommendations on the Use of Native Papuan Herbal Plants using the Large Language Models method

Suhardi Aras, Maya Nurlati Kelian, Ilham S, Alfiyyah Faridah*

Program Studi Teknik Informatika, Universitas Muhammadiyah Sorong, Indonesia
*Corresponding Email: Suhardi.aras@um-sorong.ac.id

ABSTRACTS

This study develops an AI-based chatbot to provide education and recommendations for using Papuan herbal tea. The goal is to raise awareness and dispel traditional knowledge about Papuan herbal tea, which has historically been weak in local communities. The methods used include problem analysis, literature review, data collection, system testing using Large Language Models (LLM), and implementation using Google Generative AI (Gemini-Pro). This system is designed to analyze text queries and provide relevant answers using text-based query syntax. Pengujian is carried out using the BERTScore method to assess the system's adherence to the reference data. The study's results indicate very good performance, with a rata-rata Precision of 0.58357/58 %, a Recall of 0.66781/66%, and an F1-Score of 0.62205/62. %. More precise recall indicates that the system can capture a lot of pertinent information, even when there are performance differences between questions. This study demonstrates the potential of AI technology to improve understanding between local communities and the general public as well as to support local Papuan kearifan events.

ARTICLE INFO

Article History:
Received 23 July 2024
Revised 9 Sept 2024
Accepted 20 Nov 2024
Available online 14 Dec 2024
Publication date 01 June 2025

Keywords:
AI based chatbot,
Herbal plants native to Papua,
Large Language Models (LLM),
Google Generative AI (Gemini-Pro),
BERTScore.

DOI: https://doi.org/10.34010/injiiscom.v6i1.13575
p-ISSN 2810-0670 e-ISSN 2775-5584

1. INTRODUCTION

Native herbal remedies from Papua have significant potential for both traditional and modern medicine (Zakiyah et al., 2021). But knowledge about this phenomenon is frequently lacking in local communities and is not adequately documented. (Suardika et al., 2023) (Yuda et al., 2023). In this digital age, artificial intelligence (AI) technology opens up new opportunities for enhancing and effectively replacing traditional knowledge (Anggiratih et al., 2021). Artificial Intelligence-based chatbots have demonstrated significant potential in various fields, including education and health. (Hartati & Manullang, 2024). Using Large Language Models (LLM) in chatbot development enables more accurate comprehension of natural language and context understanding (Lubis, et al., 2024).

Chatbots are interactive systems that facilitate human-computer dialogue and enable two-way communication between users and cerdas. (Apriliyanto et al., 2024). After undergoing a revolution since 1945, chatbots now use AI to increase their capabilities. Created by OpenAI, ChatGPT is an example of a canggih chatbot AI designed to better understand and respond to users' needs in a timely manner (Vera et al., 2023)(Fatin et al., 2024). LLM is a branch of NLP that is an artificial intelligence model that produces human-like text (Afriani et al., 2024). Features include a large number of parameters, training on large amounts of data, and the ability to provide relevant textual and graphical content (Nasution et al., 2024; Saputra & Hadi, 2024).

The purpose of this study is to develop an AI-based chatbot that can provide education and recommendations for using Papuan herbal tea. It is hoped that this system will improve understanding between local communities and the general public as well as aid in local Papuan kearifan events.

2. METHOD

The results of the research that was conducted in this study can be seen in the graph below:

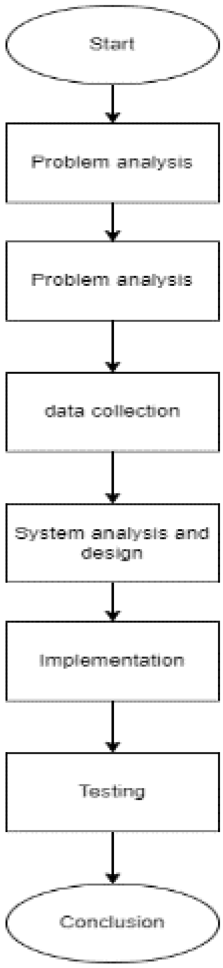

Figure 1. Research Flow

2.1. Problem analysis

Identifying a problem and understanding its limits is the first step in a research process. It is hoped that with this analysis, researchers will be able to understand the problems related to the research carried out both in terms of comprehensive functionality and help people deepen their knowledge which will help in better understanding (Pujiati & Rochmawati, 2022).

2.2. Study of literature

In this stage, the researcher collects all the necessary information, using various data collection methods according to the needs of the research being carried out, including previous journals and also e-books.

2.3. Data collection

The aim of this stage is to provide information that supports researchers in conducting this research, namely literature reviews and collecting information related to herbal plants in the form of related books. A literature review is carried out to study and understand various relevant materials such as books, journals or articles related to the research topic.

2.4. System analysis and design

The following is the process flow of a system that integrates voice recognition, natural language processing, and generative AI models.

DOI: https://doi.org/10.34010/injiiscom.v6i1.13575
p-ISSN 2810-0670 e-ISSN 2775-5584

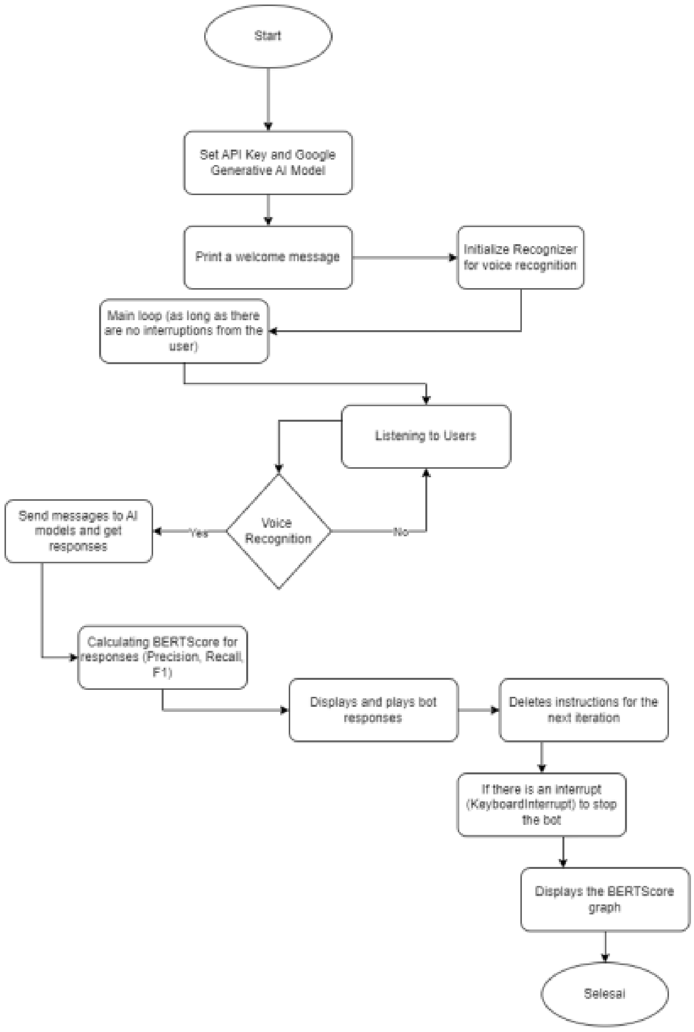

Fig. 2. System Flowchart

The following is an explanation of the process flow:

1. The system starts by setting the API Key and Google Generative AI model.
2. The system prints a welcome message and initializes the recognizer for voice recognition.
3. The system enters the main loop which runs as long as there are no interruptions from the user.
4. In the loop, the system listens for voice input from the user.
5. If voice recognition is successful:
   a. The system sends a message to the AI model and gets a response.
   b. b. The system calculates the BERTScore for the response (Precision, Recall, F1).
6. The system displays and plays the bot's response. The system deletes the instructions for the next iteration.
7. If speech recognition is unsuccessful, the system returns to the user listening stage.
8. The system checks for keyboard interrupts to stop the bot.
9. If there is an interrupt, the system displays the BERTScore graph and ends the process.

10. If there are no interrupts, the system returns to the beginning of the loop to listen for the next input.

This flow describes an interactive system that uses speech recognition to receive input, processes that input using a generative AI model, and provides a response while evaluating the quality of the response using BERTScore.

2.4.1 System analysis

This stage aims to meet application needs, which is the stage of investigation and determining essential needs. In this research, system requirements analysis uses the Large Language Model (LLM) to develop a Question Answering System (QAS) and evaluate the suitability of answers.

Large language model (LLM) is an artificial intelligence model in Natural Language Processing (NLP) that is capable of producing text with human-like quality. In this research, LLM plays a role in providing the highest weighted value from several answers given by the system, where the answer with the highest weight will be used as the answer to the user's question (Zahwa et al., 2023).

2.4.2 System design

This chatbot system is designed to provide information about the history of Jesus Christ through text-based interactions using NLP technology(Fatin et al., 2024). The system architecture involves Google Generative AI (Gemini-Pro) processing text queries and providing relevant answers. The API key is stored in the .env file to ensure secure access. Text answers are delivered via a text-based interface, such as a web form or chat application. The system can also use pyttsx3 to convert text to voice, although its main focus is on text output. Data security is maintained by not storing user information and protecting API keys. This system provides clear and easy to understand information with the potential to add voice features in the future (Kartika & Suharti, 2023).

2.5 Implemantation

System implementation is a step to implement and run a system that has been designed. This stage involves software installation, testing and ensuring that the system functions as required.

1. Development of a Voice-Based Interactive System

In developing this system, several technologies were used to create voice-based interactive applications. The system leverages several Python tools and libraries, including speech_recognition for speech recognition and gTTS for speech synthesis. The code also integrates a generative model to provide automatic answers based on voice questions. The use of this library allows the development of responsive and efficient applications without requiring in-depth knowledge of creating complex user interfaces. With this technology, developers can focus on core features, namely voice interaction and response

DOI: https://doi.org/10.34010/injiiscom.v6i1.13575
p-ISSN 2810-0670 e-ISSN 2775-5584

evaluation, while providing a pleasant user experience.

2. API Development and Use

To support system functions, this code uses several Application Programming Interfaces (APIs). An API is a collection of software methods that enable communication between applications through a set of rules that have been defined by a service provider. In this system, Google Generative AI API is used to access generative language models that provide answers based on text input. The API key is used to authenticate each request to the API server. In addition, Google Text-to-Speech API (gTTS) is used to convert text into voice, and Speech Recognition API is used to recognize speech in Indonesian. The generative model used in this system is Gemini, which is accessed via the OpenAI API.

2.6 Testing

In this research, testing was carried out by testing whether the answers produced by the system were in accordance with the existing reference data, namely the Bidirectional Encoder Representations From Transformers (BERT) method. BERTScore is a deep learning model that has provided significant advances in various NLP tasks. BERT was designed with the aim of understanding ambiguous sentences by using the context of the surrounding text, and building a more complete understanding through transformers. BERT adopts the encoder part of the transformer architecture as the basis of its pre-training model. This model can be applied to a variety of natural language processing tasks, including sentiment analysis, question-answer systems, and text summarization. In practical terms, BERT goes through two stages in the process, namely pre-training to understand the language and fine-tuning for specific tasks. The following is the equation for calculating BERTScore.

2.6.1 Precision

Precision is a measure of the extent to which the model predictions match the requested data.

$$P = \frac{1}{|C|}\sum_{c \in C} \max cosine(c, r)$$

Information:

P: Precision

|C|: Number of tokens in the candidate sentence

c: Token in the candidate's sentence

R: The set of tokens in the reference sentence

Cosine(c,r): Cosine similarity between tits (c) and (r)

2.6.2 Recall

Recall measure how effective the model is in predicting positive classes correctly.

$$R = \frac{1}{|R|}\sum_{r \in R} \max cosine(r, c) \tag{1}$$

P: Precision

|R|: Number of tokens in the candidate sentence.

r: Token in the candidate's sentence

(C): The set of tokens in the reference sentence

Cosine (r,c): Cosine similarity between tits (r) and (c)

$$\text{F1-Score} = 2 \times \frac{Precision\ x\ Recall}{Precision\ x\ Recall} \qquad (2$$

Information :

Precasion : the average cosine similarity of the tokens paired in the candidate sentence with the reference sentence.

Recall: the average cosine similarity of the tokens paired in the reference sentence with the candidate sentence.

3. RESULTS AND DISCUSSION

3.1. Test Results

The following are the BERTScore test results for 10 questions (P1-P10):

1. Precision average : 0.58357
2. Recall average : 0.66781
3. F1-Score average : 0.62205

- Recall is consistently higher than Precision
- P2 has the highest scores for all metrics
- P8 and P10 have the lowest scores
- The variation in scores between questions is quite significant.

Overall, the results show quite good performance with higher recall than precision, indicating the model tends to capture a lot of relevant information but may lack precision in some cases.

Table 1. Test result BERTScore

Question	Precision	Recall	F1-Score
P1	0.5838	0.6630	0.6209
P2	0.6845	0.6821	0.6833
P3	0.5850	0.6687	0.6241
P4	0.6072	0.6613	0.6331
P5	0.5842	0.6812	0.6290
P6	0.6171	0.6469	0.6317
P7	0.5429	0.6575	0.5947
P8	0.5175	0.6413	0.5728
P9	0.5959	0.7304	0.6563
P10	0.5176	0.6457	0.5746
Total	**5,8357**	**6,6781**	**6,2205**
Average	**0,58357**	**0,66781**	**0,62205**

DOI: https://doi.org/10.34010/injiiscom.v6i1.13575
p-ISSN 2810-0670 e-ISSN 2775-5584

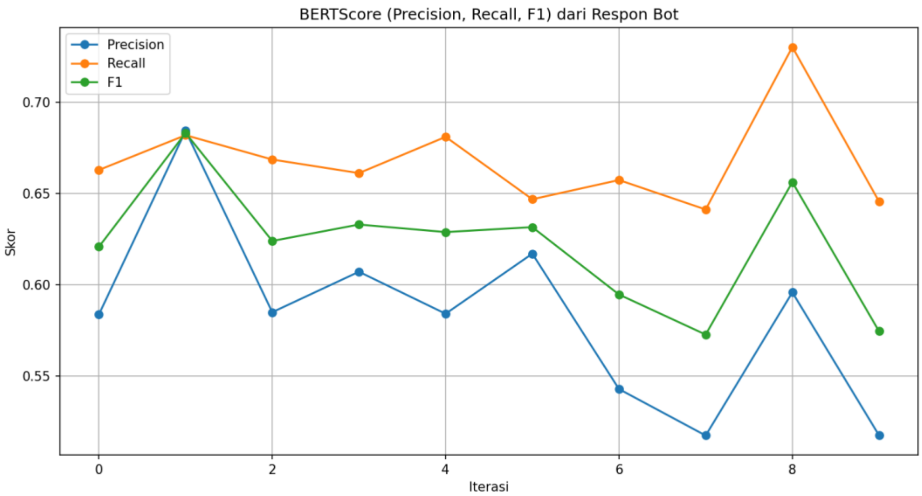

Fig. 3. Analysis diagram

The illustration in this diagram displays the BERTScore (Precision, Recall, F1) of the bot's response after a few iterations:

1. Four metrics that are reported:

- Precision (blue line)

- Recall (orange line)

- F1 score (green line)

2. The X-axis iteration varies from 0 to 8.

3. The Y-axis shows the score for each metric, ranging from 0.55 to 0.75.

4. General trend:

- Recalls tend to have the highest value among the three metrics.

- Precision tends to have the lowest value.

- The F1 score is somewhere in between, as the harmonic average of precision and recall.

- All metrics peak at the 1st iteration.

- There was a significant spike in the 8th iteration, especially for recall and F1.

- Precision shows the largest fluctuation between iterations.

5. Interpretation:

- Execute a bot with multiple iterations.

- High recall indicates bots tend to capture a lot of relevant information.

- Lower precision may indicate some irrelevant information is also included.

This diagram is useful for understanding bot performance in terms of accuracy and completeness of responses during development or testing iterations.

4. CONCLUSION

Based on research results, a Question Answering System (QAS) was developed using a Large Language Model (LLM) and data consisting of information about herbal tea in Papua New Guinea. The analysis performed on the system using the BERTScore method yielded a rata-rata Precision of 58%, a Recall of 66%, and an F1-Score of 62%. This suggests that the Question Answering System implementation carried out has a very high accuracy rate when it comes to answering inquiries about the herbal teas that are available in Papua.

This system's kinerja highlights its ability to provide accurate responses that meet user needs and simplify understanding of herbal tea in Papua and its usage. However, the study's findings identify the areas of system development in the past, such as improving performance and functionality through the integration of canggih AI and comprehensive data analysis. Subsequent research is advised to concentrate on improving data quality and output consistency in order to be more specific when answering user questions. As a result, it is hoped that this system would provide more accurate and useful information about herbal tea in Papua.

ACKNOWLEDGMENTS

The author would like to thank the lecturers in the Deep Learning course at Muhammadiyah University of Sorong who have helped so that this research can be completed.

REFERENCES

Afriani, E., Fikry, M., & Affandes, M. (2024). Aplikasi Tanya Jawab Tentang Fiqih Bersuci Berbasis Web. *ZONAsi: Jurnal Sistem Informasi*, *6*(2), 380-390.

Nasution, M. A., Fitri, A., Rizwinie, K. S., Silaban, V. S., & Khoirani, F. (2024). Implementasi NLP Dalam Pembuatan Chatbot Customer Service Publisher Jurnal Studi Kasus LARISMA. *Jurnal Sains, Teknologi & Komputer*, *1*(1), 13-17.

Anggiratih, E., Siswanti, S., Octaviani, S. K., & Sari, A. (2021). Klasifikasi Penyakit Tanaman Padi Menggunakan Model Deep Learning Efficientnet B3 dengan Transfer Learning. *Jurnal Ilmiah SINUS*, *19*(1), 75.

Apriliyanto, E., Putra, R. I., & Rahayu, Y. S. (2024). Peran AI Chatbots dalam Layanan Mahasiswa Ilmu Komunikasi Menggunakan Metode Natural Languange Processing (NLP) di Universitas Muhammadiyah Karanganyar. *Jurnal Ilmiah SINUS*, *22*(1), 59.

Fatin, D., Fajriyanti, D. T., Saputri, A. A., & Viratama, I. P. (2024). Dampak Dari Chat Gpt Bioteknologi. *Cendikia: Jurnal Pendidikan dan Pengajaran*, *2*(2), 13-20.

Hartati, R., & Manullang, E. B. (2024, March). Implementation of Telegram Chatbot AI with Natural Language Processing (NLP) in Learning Creative Entrepreneurship to Develop Students' Creative and Innovative Competence. In *Talenta Conference Series: Local Wisdom, Social, and Arts (LWSA)*, 7(2), 72-79.

DOI: https://doi.org/10.34010/injiiscom.v6i1.13575
p-ISSN 2810-0670 e-ISSN 2775-5584

Kartika, T., Suharti, B., & Sugiyanta, S. (2023). Jamu As Herbal Medicine: A Study Of Health Communication And Philosophy As Cultural Identity. *KOMUNIKA*, *6*(1), 18-27.

Lubis, A. T. U. B., Harahap, N. S., Agustian, S., Irsyad, M., & Afrianty, I. (2024). Question Answering System pada Chatbot Telegram Menggunakan Large Language Models (LLM) dan Langchain (Studi Kasus UU Kesehatan): Question Answering System on Telegram Chatbot Using Large Language Models (LLM) and Langchain (Case Study: Health Law). *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(3), 955-964.

Pujiati, R., & Rochmawati, N. (2022). Identifikasi citra daun tanaman herbal menggunakan metode Convolutional Neural Network (CNN). *Journal of Informatics and Computer Science (JINACS)*, *3*(03), 351-357.

Saputra, D. G., & Hadi, A. (2024). Sistem Integrasi Pembayaran Spp Di Sekolah Menggunakan Model Nlp Pada Toko Retail. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, *9*(2), 1073–1084.

Suardika, I. W. G., Dewi, N. M. W. A., & Megawati, F. (2023). ARTIKEL REVIEW: Penggunaan Obat Herbal Dalam Upaya Swamedikasi atau Pengobatan Sendiri Pada Penyakit Batuk Dan Flu. *Usadha*, *2*(2), 9-18.

Vera, M. C. S., & Palaoag, T. D. (2023). Implementation of a Smarter Herbal Medication Delivery System Employing an AI-Powered Chatbot. *International Journal of Advanced Computer Science and Applications*, *14*(3).

Yuda, A. K. S., & Ahmad, S. (2023). Implementasi Prediksi Tanaman Herbal Menggunakan Algoritma Convolutional Neural Network Berbasis Android. *Reputasi: Jurnal Rekayasa Perangkat Lunak*, *4*(2), 84-88.

Zahwa, A. F., Fiati, R., & Murti, A. C. (2023). Implementasi Chatbot untuk Customer Service menggunakan Metode Natural Language Processing (NLP) (Studi Kasus Website Theme62. com). *JIMP-Jurnal Informatika Merdeka Pasuruan*, *7*(2), 82-86.

Zakiyah, W., Agustin, A. E., Fauziah, A., Maharani, D., & Mukti, G. I. (2021). Metode Pemisahan Fitokimia Tanin pada Tanaman Herbal Indonesia. *PharmaCine: Journal of Pharmacy, Medical and Health Science*, *2*(1), 51-59.