2023 5th International Conference on Cybernetics and Intelligent System (ICORIS) | 979-8-3503-6948-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICORIS60118.2023.10352259

# AI-powered Chatbot for Information Service at Klabat University by Integrating OpenAI GPT-3 with Intent Recognition and Semantic Search

Semmy Wellem Taju
*Department of Computer Science*
*Klabat University*
Manado, Indonesia
semmy@unklab.ac.id

Green Ferry Mandias
*Department of Computer Science*
*Klabat University*
Manado, Indonesia
green@unklab.ac.id

Jimmy Herawan Moedjahedy
*Department of Computer Science*
*Klabat University*
Manado, Indonesia
jimmy@unklab.ac.id

Andria Kusuma Wahyudi
*Department of Computer Science*
*Klabat University*
Manado, Indonesia
andriawahyudi@unklab.ac.id

Reymon Rotikan
*Department of Computer Science*
*Klabat University*
Manado, Indonesia
reymonr@unklab.ac.id

Edson Yahuda Putra
*Department of Computer Science*
*Klabat University*
Manado, Indonesia
eyahuda@unklab.ac.id

***Abstract*— The rapid advancement of Artificial Intelligence (AI) in recent years has transformed numerous sectors, including information services. This study introduces Unklabot 1.0, an innovative AI-powered chatbot specially designed and proposed for information services at Klabat university. Unklabot 1.0 was developed to take advantage of an advanced large language model (LLM), namely OpenAI GPT-3 (*text-davinci-003*), and integrate it with intent recognition and knowledge-based context understanding using semantic search technique to provide accurate and efficient question answering capabilities about Klabat University. To enhance Unklabot's performance, intent recognition is integrated into the Unklabot system. Moreover, this intent recognition helps identify the purpose behind user queries, allowing the chatbot to understand user intentions more accurately. Furthermore, semantic search technique is employed to enhance the capabilities of the Unklabot by understanding the meaning behind user queries and retrieve information in external knowledge based on the closest context or chunk. Utilizing OpenAI GPT-3 API, Unklabot can processes natural language inputs, providing comprehensive and contextually relevant answers to user inquiries. This research successfully implemented the OpenAI GPT-3 API in information services, offering a valuable tool to increase the engagement of AI usage in educational institutions such as Klabat University.**

***Keywords—chatbot, artificial intelligence, unklabot, machine learning, large language model, openai gpt, intent recognition, semantic search, chroma database.***

## I. INTRODUCTION

In recent years, the development of Artificial Intelligence (AI) has opened up tremendous opportunities in various fields, including the development of intelligent chatbots. Researchers have observed significant success and impact from a number of large language models, such as OpenAI ChatGPT (GPT-3 and GPT-4) [1], Meta LLaMA 2 [2], Anthropic Claude 2, Baidu ERNIE 3.0 [3], Google Bard [4] and other generative AI language models released recently. This current research aims to explore the potential of utilizing the OpenAI GPT-3 API, one of the current state-of-the-art generative AI language models (*GPT-3 or third-generation Generative Pre-trained Transformer*), to create an artificially intelligent chatbot focused on information service at Klabat University (UNKLAB). However, generative AI language models do not have the ability to provide the most up-to-date information and cannot search for actual data independently. This problem often occurs with AI-generated text (*we tried with language model API*) where it sometimes generates responses based on its own assumptions, which do not match the actual situation. Therefore, we attempt to stop actions that produce inappropriate output by predicting the intent of the user's question using intent recognition approach.

The proposed AI-powered chatbot (called Unklabot 1.0) utilizes the capabilities of the OpenAI GPT-3 model (*text-davinci-003*), an advanced Large Language Model (LLM) introduced by OpenAI, to generate human-like responses to user queries. By utilizing this GPT-3 model, the proposed AI-chatbot can understand and process natural language input, ensuring a more interactive and intuitive user experience. In addition, OpenAI GPT-3's extensive knowledge base enables the chatbot to provide comprehensive and contextually relevant answers to a wide range of questions. To enhance the performance of the Unklabot 1.0, intent recognition was integrated into the Unklabot system. Intent recognition helps identify the intent (*purpose*) behind a user query (*request/question*), thus allowing the Unklabot to understand the user's intent more accurately at the outset. We hope that this integration will enable Unklabot to respond with *relevant information* and take *appropriate actions* based on the user's query. We also apply semantic search technique to improve the ability of the proposed Unklabot to provide responses according to the closest context data from the external knowledge database (*memory-like capability*). This external knowledge stored in the Chroma vector embedding database refers to the utilization of additional knowledge or information from external sources that are separate from the OpenAI GPT-3 model training data. Hence, semantic search technique allows for retrieving this additional information based on context rather than relying solely on keyword matching. This approach not only provides accurate responses according to context, but also a deeper understanding of the user's question.

Based on our observation, we found that AI language models (*one of the popular generative AI models*) sometimes give inappropriate responses in some situations where AI language model does not have special restrictions or rules like traditional chatbots. This can lead to undesirable responses in certain contexts (*inappropriate*). The following are some of the limitations of generative AI language models including: (1) Lack of deep context understanding from AI language models which often cannot maintain a deep understanding of the context of the ongoing conversation. (2) Lack of intent understanding which does not automatically extract the "*intent*" of a given sentence even though it can provide responses to specific questions. (3) If the training data contains certain biases, it may produce responses that reinforce those biases. (4) The AI language model does not have up-to-date knowledge after the cut-off date. (5) Limited training data can cause AI language models to struggle to provide appropriate responses and produce unsatisfying outcomes. (6) AI language models sometimes experience the problem of "*generating*" or "*producing*" incorrect information about a specific domain (*certain area*) due to limitations in the training data and the way the model fills in information gaps with incorrect assumptions.

TABLE I. DATA DETAILS FOR AI-POWERED CHATBOT TO OBTAIN GENERAL INFORMATION ABOUT KLABAT UNIVERSITY

| Unklab Data | Text Chunks (2000 Characters) | Data Source |
|---|---|---|
| Informasi Umum/General (*general information*) | 8 segments | Unklab web, Wikipedia |
| Biaya Kuliah Pemondokan (*accommodation tuition fees*) | 4 segments | Unklab web |
| Pendaftaran Mahasiswa (*student registration*) | 7 segments | Unklab web |
| Grading System S1 | 4 segments | PDF |
| Sejarah Unklab (*history*) | 2 segments | Wikipedia |
| Visi Misi Tujuan (*vision mission purpose*) | 2 segments | Unklab web |
| Beasiswa Unklab (*scholarship*) | 5 segments | Unklab web |
| ASMI (*Akademi Sekretari Manajemen Indonesia*) | 2 segments | Unklab web |
| Fakultas Ekonomi Bisnis (*faculty of business economics*) | 8 segments | Unklab web |
| Fakultas Filsafat (*faculty of philosophy*) | 3 segments | Unklab web |
| Fakultas Ilmu Komputer (*faculty of computer science*) | 8 segments | Unklab web |
| Fakultas Pendidikan/Keguruan (*faculty of education*) | 5 segments | Unklab web |
| Fakultas Pertanian (*faculty of agriculture*) | 4 segments | Unklab web |
| Fakultas Keperawatan (*nursing faculty*) | 3 segments | Unklab web |

By demonstrating the possibility of combining OpenAI GPT-3 API with *intent recognition* and *semantic search technique* to create an advanced AI-powered chatbot for question-answering system, this research paper makes a contribution to the field of artificial intelligence. The results of this study can be used to improve information service at educational institutions like Klabat University and provide the foundation for future developments in smart chatbot technology. Additionally, by creating intelligent chatbot systems for academic institutions, this research advances the area of artificial intelligence while highlighting the possibilities of using cutting-edge AI technology such as Large Language Model for information services.

## II. RELATED WORK

One of the main benefits of the question answering system is the ability to understand natural language questions and accurately provide relevant answers to the user's questions that are widely accessible. Currently, AI-powered chatbots are also more widely used in the business groups to reduce customer service cost and handles multiple users at a time. These types of intelligent chatbots are widely utilized across many organizational domains where it can replace humans. Ranoliya *et al.* [5] developed a chatbot for university's Frequently Asked Questions (FAQs) to provide an efficient and accurate answer using Artificial Intelligence Markup Language (AIML) and Latent Semantic Analysis (LSA). Several studies have also tried to integrate the latest techniques such as machine learning and natural language processing (NLP) to improve the performance of chatbots in interacting with users. In our research, Rampengan *et al.* [6] designed a chatbot by implementing NLP technique with machine learning approach to develop a chatbot widget as online virtual assistant for dormitory room information and reservation service at Klabat University.

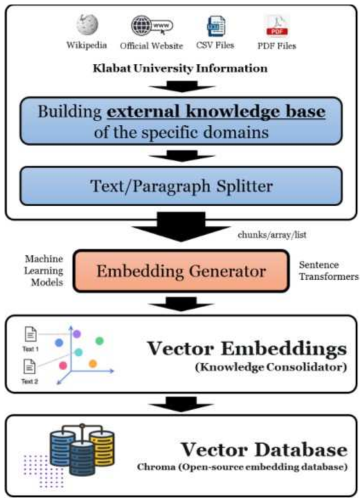


Fig. 1. Vector Embedding Processes for Building External Knowledge of Unklabot 1.0.

There have been many attempts to integrate artificial intelligence (AI) with chatbot systems to provide more sophisticated and responsive information services. Several studies have explored the utilization of advanced language models such as OpenAI GPT-3. Chintagunta, Bharath, *et al.* [7] presented an algorithm to create synthetic training data with an explicit focus on capturing medically relevant information by utilizing OpenAI GPT-3 to produce high quality training data that can further be combined with human labelled data to get summaries. Sarol *et al.* [8] applied the OpenAI GPT-3 language model to a chatbot system in the aviation sector to improve communication, enhance natural language interactions and the usability response from selected airlines passengers' feedback. This AI chatbot should actually assist airline customers in acquiring more accurate related information such as flight booking, schedules, and updates.

This research highlights the crucial role of intent recognition in understanding user queries/requests and emphasizes the significance of semantic search to retrieve appropriate and relevant information (*closest context or chunk*) from a vector embedding database as an external knowledge source. Unfortunately, there is limited research focus on the implementation of intelligent chatbots, OpenAI GPT-3 API, intent recognition, and semantic search in the university environment, especially at Klabat University. Thus, this research offers an in-depth look to fill the knowledge gap in these areas, especially in a university context.

## III. METHODOLOGY

### A. Data Collection

In generally chatbot development, some challenges are addressed such as data freshness, knowledge of a particular domain, or accessing internal information from the target domain. The information available on the official Klabat University website that has been shared publicly can be utilized to find the latest news. In this development, we first start by building the knowledge base of Unklabot. Various data sources were used to build a strong knowledge base for our proposed Unklabot system. The data are generated from various sources, including the official Klabat University website (*https://www.unklab.ac.id/*), Wikipedia, CSV and PDF files.

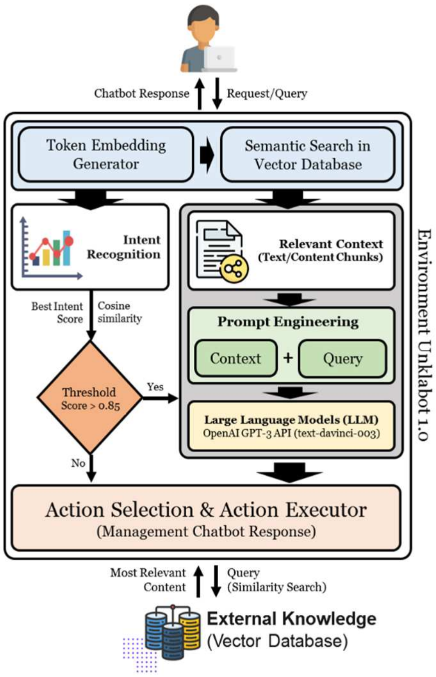


Fig. 2. AI-powered Chatbot System Architecture for Information Service at Klabat University

Table 1 presents a description of the data as external knowledge used in the development of Unklabot for information services at Klabat University. These data are then split into text chunks (*data parts*) where each text chunk has a length of less than 2000 characters. In total, 65 text chunks have been retrieved from various sources and used it as a knowledge base (*knowledge relevant to human language for language processing and context understanding*) to support the performance of Unblabot which integrated with OpenAI GPT-3 model API, intent recognition and semantic search capabilities. The use of various data sources is expected to increase the knowledge, accuracy and relevance of information services provided by Unklabot in the Klabat University.

This current study includes a comprehensive list of target intents, which is illustrated in Table 2. A total of 22 intents, comprising 854 carefully curated samples, have been painstakingly generated during the design of this AI-based chatbot. The goal is to accurately enable Unklabot to understand the user's intent. This endeavor aspires to enhance Unklabot's capacity to comprehend and respond effectively to users' needs, fostering a more meaningful and engaging interaction between the system and its users. Integration with the OpenAI GPT-3 API gives the chatbot the ability to understand human language more simply according to the given context with the appropriate target intent after the intent prediction process, therefore it can provide better and relevant responses to user questions about Klabat University.

TABLE II. VARIOUS INTENTS AND TOTAL SAMPLES FOR BUILKING INTENT RECOGNITION

| Intent | Brief Descriptions | Total Data |
|---|---|---|
| GREETING | Intent for greetings or salutations. | 15 samples |
| TIME | Intent for inquiring or informing about time. | 8 samples |
| DATE | Intent for inquiring or informing about dates. | 14 samples |
| NAME | Intent for inquiring or providing names. | 19 samples |
| HOBBY | Intent for inquiring or discussing hobbies. | 16 samples |
| HUMAN | Intent for identifying or discussing human-related topics. | 17 samples |
| THANKYOU | Intent for expressing gratitude or appreciation. | 10 samples |
| GOOD | Intent for conveying positive sentiments or responses. | 19 samples |
| GENERAL | Intent for general inquiries or statements about Unklab. | 227 samples |
| BIAYA | Intent for inquiring or discussing costs or fees. | 33 samples |
| PENDAFTARAN | Intent for inquiries or discussions related to registration. | 47 samples |
| GRADING | Intent for inquiries or discussions related to grading or assessment. | 25 samples |
| SEJARAH | Intent for inquiries or discussions related to history. | 39 samples |
| VISIMISI | Intent for inquiries or discussions related to vision and mission. | 26 samples |
| SCHOLARSHIP | Intent for inquiries or discussions related to scholarships. | 28 samples |
| ASMI | Intent for informing about ASMI. | 41 samples |
| EKONOMI | Intent for informing about faculty of business economics. | 46 samples |
| FILSAFAT | Intent for informing about faculty of philosophy. | 37 samples |
| FILKOM | Intent for informing about faculty of computer science. | 67 samples |
| FKIP | Intent for informing about faculty of teaching. | 41 samples |
| FAKPER | Intent for informing about faculty of agriculture. | 36 samples |
| FKEP | Intent for informing about nursing faculty. | 43 samples |

*B. Building External Knowledge Base*

As can be seen in Figure 1, the development of Unklabot for Information Services at Klabat University by integrating OpenAI GPT-3 API with external knowledge data which is presented in the form of vector embedding to facilitate the search for relevant and contextually appropriate information easily. We manually collect data from different sources and by adopting the concept of knowledge consolidator to perform mechanisms such as collecting, curating, and integrating information from various external knowledge sources. With this concept, we explored trusted sources such as the official website of Klabat University, Wikipedia, and also relevant current news sources about Klabat University that are publicly accessible. The information is processed into vector embeddings (*embedding generator*) so that it can be accessed and utilized by Unklabot to provide more accurate and informative answers to users according to the closest context in the embedding database of the given query.

*C. Vector Embedding Database*

In this research, Chroma (*https://docs.trychroma.com/*) vector database is utilized as an open-source vector embedding database to give the proposed Unklabot a memory-like capability. This database plays an important role in building an external knowledge base for Unklabot. Chroma Database serves as a repository for vector representations of various data sources in the information service as a target domain of Klabat University. Any information that has been processed by the Knowledge Consolidator will be stored in the Chroma database, which later become an external knowledge source that can be accessed by Unklabot quickly and efficiently. Basically, the cosine function [9] and K-Nearest Neighbor (KNN) algorithm [10] are utilized in the chroma database search query to measure the similarity (*distance*) between two sets of documents/texts in high-dimensional space. By integrating OpenAI GPT-3, intent recognition, and semantic search techniques with vector embeddings representation of the Chroma database, Unklabot at Klabat

University becomes a potent and trustworthy information tool, assisting users in searching and obtaining current, relevant information about Klabat University.

### D. *Semantic Search Technique*

Search with meaning is known as semantic search [11]. In order to match a user query to the appropriate content, semantic search [12] utilizes the user's intent, context, and conceptual meanings. Semantic search uses vector search to return the closest context or chunk, even when there are no word matches. The Semantic Search technique is one of the key elements that is essential to improve Unklabot's ability to interpret the questions asked by users and provide most relevant answers (*closest context*). Therefore, the implementation of semantic search involves the representation of data into the vector embedding form, where all sentences containing information about Klabat University are converted into numerical representations in vector space. By utilizing vector embedding, Unklabot can identify and understand the context of a question. By utilizing vector embedding, Unklabot can recognize and understand the context of meaning of diverse questions, so that it can search for the most appropriate and precise information in the Chroma database.

### E. *Intent Recogntion Approach*

The main objective of intent recognition is to recognize and understand the intentions hidden behind the questions asked by users. With this ability, Unklabot can determine the true purpose of the question, direct action selection to the right place and respond with more relevant answers. The use of intent recognition is particularly important as users often ask questions in a variety of different forms and sentences. Only by accurately understanding the user's intention, Unklabot can provide more effective and satisfying information services. We can actually use machine learning algorithms such as Gaussian Naive Bayes (GNB) [13], Random Forest Classifier (RFC) [14], *k*-Nearest Neighbors algorithm (KNN) [15], Support Vector Machine (SVM) [16] or Single Layer Neural Network [17] to perform intent recognition. However, we prefer to use the Chroma database and cosine similarity algorithm to demonstrate this intent recognition.

In the *intent recognition approach*, Figure 3 shows the representation of vector embedding data in a high-dimensional space and it actually plays a key role. Questions posed by users are first converted into numerical representations in the vector embeddings in order to compute the Cosine(θ) of two vector embeddings. By utilizing vector embedding, the Unklabot can identify patterns of meaning hidden in questions to find the most relevant intent and its intent label. The collection of intents was first stored in Chroma database and it uses to search the 10 top best scores that are strong related to the user query. In the second step, we applied again the cosine similarity algorithm (*cosine distance*) to find the best intent score. This cosine distance is defined as Similarity(A, B), ranging from 0 to 1. A cosine distance of 1 indicates high relevant or identical intent, while a value of 0 signifies complete dissimilarity intent. This process allows Unklabot to access and utilize the knowledge stored in Chroma database as to give the proposed Unklabot a memory-like capability.

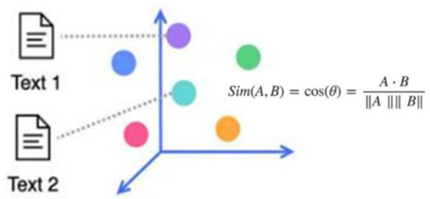


Fig. 3. The Representation Vector Embeddings of Texts in a High-dimensional Space

### F. *Prompt Engineering*

To achieve optimal performance, the use of appropriate, clear and directed prompts will help the generative AI language model to extract/summarize relevant information from the context data to be provided more efficiently. Prompt Engineering focuses on designing the input or question used to initiate interaction with the OpenAI GPT-3 model. With this prompt engineering, Unklabot can accurately recognize the user's intention, so that the process of extracting relevant information can run smoothly and can provide answers that match the questions or needs of the user.

### G. *OpenAI GPT-3 API*

One of the Large Language Model (LLM) is utilized to improve the performance of the proposed AI chatbot for information services at Klabat University. LLM is a type of generative AI model that uses transformer-based neural networks with millions to billions of parameters to generate natural text. The AI language model used in this research is *text-davinci-003*, which is one of the models from the GPT-3 family. This model is trained with large datasets, we utilise this unique capability to be able to understand the context of the data to be fed as well as answer questions according to the context of the data. In forming responses from Unklabot, the *text-davinci-003* model uses two important elements, namely content and query. Context is additional information or external knowledge, while query is a question from the user. By combining context and query, the model can understand the question more thoroughly and provide a more appropriate response according to the context of the given data.

### H. *Chatbot Performance Evaluation*

It is important to measure the proposed Unklabot performance. Therefore, we conducted a chatbot performance evaluation by including human evaluation factors. This evaluation approach aims to measure the extent to which Unklabot is able to provide effective and satisfying information services for users. In this evaluation, we interact with Unklabot in various question and answer scenarios that cover a variety of contexts and levels of question complexity. Furthermore, the evaluation process is carried out by collecting a number of questions and each conversation is labelled based on the intent expressed by the user and the intent predicted by Unklabot. But the important thing is that each response from Unklabot will be assessed/evaluated based on criteria such as relevance and accuracy of the response to the question asked. We calculated the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), which allowed us to calculate the Accuracy (Acc) as an evaluation metric.

$$Acc = \left(\frac{TP+TN}{TP+FP+TN+FN}\right) = \left(\frac{15+2}{15+1+2+2}\right) \times 100\% = 85\% \quad (1)$$

With this method, we can systematically evaluate the strengths and limitations of Unklabot's ability to accurately response to user questions. Additionally, to understand the error patterns made by the Unklabot. We used the TP, to represent responses where the Unklabot accurately identified a user's intent or delivered correct answers. TN represents responses where the Unklabot correctly identified that it couldn't answer a user's query or didn't understand the intent. FP represents responses where the Unklabot generated a response that appeared correct to the Unklabot but was actually incorrect or irrelevant. FN represents responses where the Unklabot failed to recognize a user's intent or provide a correct response when it should have. By identifying *FP* and *FN*, we can identify which type questions or situations that often lead to errors in responses.

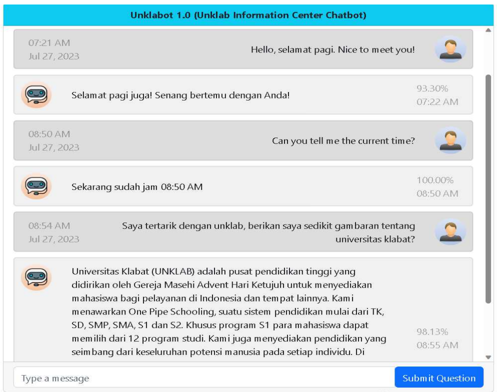


Fig. 4. Unklabot response snapshot 1

## IV. Results

### A. Unklabot System Development Environment

To support the implementation of our AI-powered chatbot, several important Python packages have been installed, including openai (0.27.8), chromadb (0.3.2), and sentence-transformers (2.2.2). These packages play an important role in natural language processing and converting text into a 768-dimensional vector representation using the Sentence Transformer model with the '*distilbert-base-nli-mean-tokens*' method. In addition, the use of Chroma database is key in organising and storing relevant information in two collections, namely *unklab_collection* and *intent_collection*. To run Chroma, the local API was directly used and successfully loaded in total of 920 embeddings and 2 collections. The use of the OpenAI GPT-3 model API with version "*text-davinci-003*" contributed greatly to understanding the questions and generating appropriate answers, with a conversation quality threshold set at 0.85 based on the predicted intent score. we utilized Google Colaboratory (Colab), a platform for generating the chromadb embedding collections. For frontend web development, we employed Bootstrap v5.0 to design the graphical user interface (GUI), employing PHP, HTML, CSS, and JavaScript (JQuery 3.6.0). Python programming along with various packages were used for backend development to implement the Unklabot functionalities.

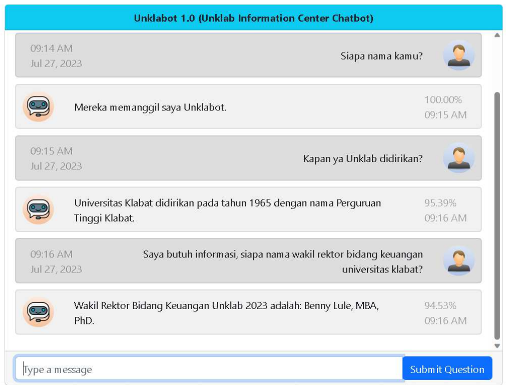


Fig. 5. Unklabot response snapshot 2

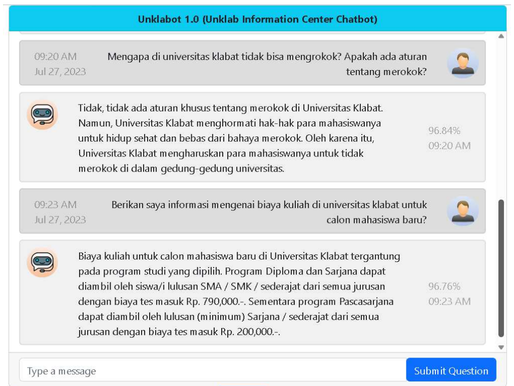


Fig. 6. Unklabot response snapshot 3

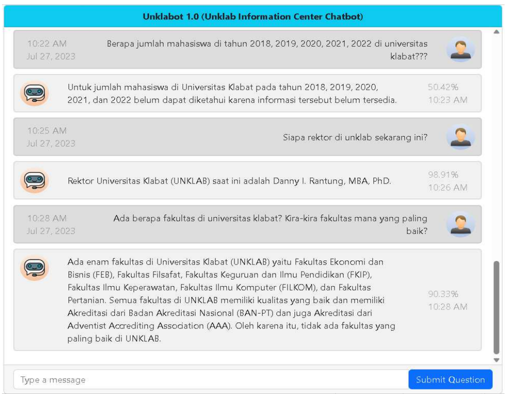


Fig. 7. Unklabot response snapshot 4

### B. Unklabot Performance Evaluation

To evaluate the performance of Unklabot, about 20 questions were carefully generated and researchers carefully observed each response from Unklabot. After the responses of Unklabot were assessed, we found the following measurement values TP = 15, TN = 2, FP = 1, and FN = 2. Thus, Unklabot has an Accuracy of about 85%. The Accuracy rate of 85% indicates the Unklabot has a good performance in recognizing questions that can be answered and identifying questions that cannot be answered correctly. In addition, the 85% Accuracy rate shows Unklabot's good performance in recognizing questions that can be answered and identifying questions that cannot be answered correctly.

### C. Web Based Graphical User Interface (GUI)

In this research, we have developed an interactive Unklabot for Klabat University related to question answering system. Unklabot web with GUI serves as the user's main interaction website, allowing good communication to take place. Through the GUI, users can input their questions and receive responses. As reported in Figure 4, Figure 5, Figure 6, Figure 7 and Figure 8, it presents the web user interface of the proposed AI-powered chatbot system.

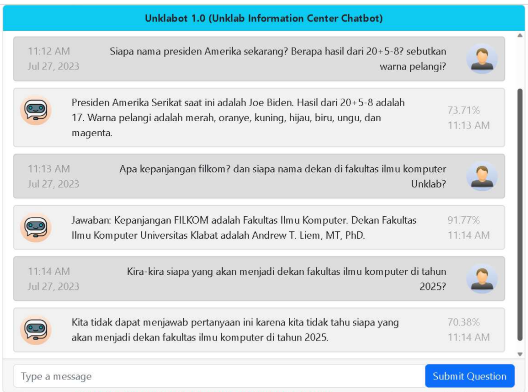

Fig. 8. Unklabot response snapshot 5

### D. Source Code Available on Github

The source code for an artificial intelligence (AI)-based chatbot developed for information services at Klabat University, which integrates OpenAI GPT-3 API with intention recognition approach and semantic search technique, has been made openly available on Github at *https://github.com/semmytaju/unklabot-1.0*. By providing access to the source code, it is hoped that researchers, developers and other interested parties can build upon this work, encouraging continuous improvement and innovation in the field of AI-based chatbots for educational institutions.

## V. Conclusion and Future Work

### A. Conclusion

Generative AI language models, chatbots, AI conversational agents and AI personal assistants are becoming increasingly popular for mobile and web services in various domains, such as entertainment, commercial systems and even academics. The use of an AI-powered chatbot based on the OpenAI GPT-3 API model that has been integrated with intent recognition approach and semantic search capabilities in information service at Klabat University promises tremendous potential in improving the quality and efficiency of services to prospective students. With this advanced technology, it is expected that users can easily get the information they need quickly and accurately. In addition, this integration can also reduce the workload of information services staff, allowing them to focus on more complex and strategic tasks.

### B. Future Work

For further development (*future work*), several aspects can be improved in this research, including (*i*) the long response time due to the process of loading the vector embedding database, predicting intent and requesting to the OpenAI language model can be overcome by optimising the algorithm and infrastructure. (*ii*) to improve the performance, more data can be added to the intent training data so that the system can better recognize user intent. (*iii*) selection of a better version of the GPT model, such as GPT-4, can improve the capabilities and response quality of the chatbot. (*iv*) integration with more external knowledges with Unklabot is essential to enrich the knowledge of the AI-powered chatbot. (*v*) special attention to security needs to be prioritised to protect user data and privacy. (*vi*) implementing this chatbot as a mobile application can expand user reach and improve accessibility. (*vii*) fine tuning the AI model can also increase the knowledge of the proposed chatbot and improve its performance. By taking these steps, this research can produce an AI chatbot that is more efficient, innovative and beneficial to more information services at Klabat University.

## Acknowledgment

This work is supported with good condition and funded by the Computer Science Faculty at Klabat University.

## References

[1] R. J. M. Ventayen, "OpenAI ChatGPT generated results: Similarity index of artificial intelligence-based contents," *Available at SSRN 4332664*, 2023.

[2] H. Touvron *et al.*, "Llama: Open and efficient foundation language models," *arXiv preprint arXiv:2302.13971*, 2023.

[3] Y. Sun *et al.*, "Ernie 3.0: Large-scale knowledge enhanced pre-training for language understanding and generation," *arXiv preprint arXiv:2107.02137*, 2021.

[4] B. Ram and P. V. Pratima Verma, "Artificial intelligence AI-based Chatbot study of ChatGPT, Google AI Bard and Baidu AI," *World Journal of Advanced Engineering Technology and Sciences*, vol. 8, no. 01, pp. 258–261, 2023.

[5] B. R. Ranoliya, N. Raghuwanshi, and S. Singh, "Chatbot for university related FAQs," presented at the 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, 2017, pp. 1525–1530.

[6] E. G. C. R. Rampengan, S. W. Taju, and V. Tewu, "Chatbot Widget as Online Virtual Assistant for Dormitory Room Information and Reservation Service at Klabat University," *CogITo Smart Journal*, vol. 9, no. 1, pp. 181–192, 2023.

[7] B. Chintagunta, N. Katariya, X. Amatriain, and A. Kannan, "Medically aware GPT-3 as a data generator for medical dialogue summarization," presented at the Machine Learning for Healthcare Conference, PMLR, 2021, pp. 354–372.

[8] S. D. Sarol, M. F. Mohammad, and N. A. A. Rahman, "Mobile Technology Application in Aviation: Chatbot for Airline Customer Experience," in *Technology Application in Aviation, Tourism and Hospitality: Recent Developments and Emerging Issues*, Springer, 2022, pp. 59–72.

[9] C. VASISHTA and B. CHRISSIE, "Netflix Movie Recommendation Using Cosine Similarity".

[10] N. Ritha *et al.*, "Sentiment Analysis of Health Protocol Policy Using K-Nearest Neighbor and Cosine Similarity," presented at the ICSEDTI 2022: Proceedings of the 1st International Conference on Sustainable Engineering Development and Technological Innovation, ICSEDTI 2022, 11-13 October 2022, Tanjungpinang, Indonesia, European Alliance for Innovation, 2023, p. 195.

[11] E. Mäkelä, *Survey of Semantic Search Research*. 2008.

[12] N. Charbel, C. Sallaberry, S. Laborie, and R. Chbeir, "FEED2SEARCH: a framework for hybrid-molecule based semantic search," *International Journal of General Systems*, vol. 52, no. 3, pp. 343–383, 2023.

[13] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine learning*, vol. 29, no. 2, pp. 131–163, 1997.

[14] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.

[15] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy k-nearest neighbor algorithm," *IEEE transactions on systems, man, and cybernetics*, no. 4, pp. 580–585, 1985.

[16] S. Huang, N. Cai, P. P. Pacheco, S. Narrandes, Y. Wang, and W. Xu, "Applications of support vector machine (SVM) learning in cancer genomics," *Cancer genomics & proteomics*, vol. 15, no. 1, pp. 41–51, 2018.

[17] J. J. Shynk, "Performance surfaces of a single-layer perceptron," *IEEE Transactions on Neural Networks*, vol. 1, no. 3, pp. 268–274, 1990.