Intelligent Web-Based Virtual Assistant for Klabat University Using Retrieval-Augmented Generation (RAG) and Text-to-Speech (TTS) Integrated with Large Language Models

Semmy Wellem Taju[1] , Green Arther Sandag [2], Ogi Wemy Corinta[3], Rahayu Mokodompit[4] , Mercy Yuliana Sakka[5]

Informatics, Faculty of Computer Science, Klabat University, Airmadidi, Indonesia
e-mail: [1]semmy@unklab.ac.id , [2]greensandag@unklab.ac.id
[3]s22110060@student.unklab.ac.id, [4]s22110053@student.unklab.ac.id,
[5]s22110311@student.unklab.ac.id

Abstract

Abstract - The development of information technology has brought significant changes in the world of education, including at Klabat University. To increase accessibility and ease of obtaining information, a web-based virtual assistant system was developed that utilizes Retrieval-Augmented Generation (RAG) and Text-to-Speech (TTS) technologies. This research aims to implement RAG and TTS technology in virtual assistant system to to provide faster and more interactive answers to users. With this system, users can get information faster. Extreme Programming (XP) is used with an iterative approach in software development. The development process includes planning, design, coding, testing, and continuous feature enhancement. The results showed that the implementation of RAG and TTS improved efficiency in information delivery. Users can ask questions and receive answers in the form of text and voice, making information search more interactive. This system is expected to be an innovative solution for information services at Klabat University.

Keywords— Asisten Virtual, Sistem Informasi, Large Language Model, Retrieval-Augmented Generation, Speech-to-Text, Text-to-Speech.

I. INTRODUCTION

The development of increasingly sophisticated information technology has had a major impact in various aspects of life (Kanivia et al., 2024) Technologies such as artificial intelligence, and machine learning have opened up new opportunities to increase the effectiveness of information technology in various fields. Such as business, health, government, environment and also education. In the field of education for example, information technology contributes effectively by providing support in terms of communication and enhancing the experience for the community (Anjelina, 2025). These changes have transformed the world of education around the world, affecting the way we obtain, deliver and access information (Inayah et al., 2024). To get, convey and access information we can use various platforms, such as search engines, news portals, social media and also websites. Speaking of websites, Klabat University (UNKLAB) as an

educational institution in North Sulawesi provides campus information services through a website. The website provides various information such as campus history, vision and mission, profiles of campus officials, lecturers and staff, faculties, study programs, and facilities. However, access to this information is still conventional, where users must search and read web pages one by one to find the information they need. To provide an adaptive solution, a virtual assistant that utilizes Large Language Model (LLM), Speech To Text (STT) & Text To Speech (TTS) technology and 3D animation was developed. With this system, users can ask questions directly and receive specific answers easily, either in text or voice. This virtual assistant is designed to improve the user experience for students, prospective students, lecturers, and staff in finding information related to UNKLAB with an intuitive approach.

In this study, researchers try to implement RAG and TTS on virtual assistants at Klabat University. RAG is an innovation in artificial intelligence that combines generative modeling capabilities with real-time access to external knowledge sources. Unlike traditional generative AI models, which rely solely on static, pre-trained data, RAG dynamically accesses and integrates relevant information, ensuring accurate, contextually relevant and up-to-date responses (Fitriawati et al., 2020). In this research, RAG is used to provide responses to users according to the available information. In addition, Text-to-Speech (TTS) is a system that converts text into sound similar to the human voice. TTS can be applied to web applications using responsive voice (Kreativitas & Informatika, 2023). Modern TTS architecture usually goes through three steps: first, the text is converted into basic sounds (phonemes) using a special tool (phonemizer); second, these basic sounds are converted into sound images (mel-spectrogram); and finally, the sound images are used for the original voice (Pratama et al., 2025). In this research, TTS is used to convert the text output from the LLM model into a voice like a human voice.

II.   LITERATURE REVIEW

2.1 Information System

System is a combination of a collection of elements, components or variables that are interconnected with each other in order to achieve a certain goal (Alifianda & Djutalov, 2024). And information is data processed from one or more sources that provide value and benefits to the recipient (Susilawati et al., 2020). From the above understanding, it can be concluded that the information system is a unit consisting of elements that are interconnected to process data from various sources, so as to produce information that has value and benefits for its users in achieving certain goals. A good and quality information system is the main thing that is very concerned and becomes a goal that must be achieved in order to provide satisfaction and convenience in obtaining the information needed for companies or agencies to support decisions (Andriyan et al., 2020). In addition, a well-designed information system can improve operational efficiency and support faster and more accurate decision making.

2.2 Virtual Assistant

Virtual assistants are artificial intelligence (AI)-based computer programs or applications designed to carry out specific tasks through interactions that resemble human conversations, both in text and voice (Arifiyanti et al., 2023). These software programs

use AI technology to mimic human abilities in performing various tasks, such as answering questions, providing information, and completing certain jobs (Azis & Pramono, 2024). With their growing capabilities, virtual assistants have the opportunity to change the way we interact with technology and positively impact our lives (Arifiyanti et al., 2023). Currently, virtual assistants have been implemented in various fields of business, government and are starting to be adopted in the field of education (Soetiyono et al., 2024).

2.3 Large Language Model

LLM is a machine learning model trained with a lot of text data, which allows it to learn, analyze, and create text similar to natural human language (Lu, 2024). LLM works by learning vector representations for words, sentences, or documents, and uses multiple parameters to encode semantic relationships among words present in the text (Gao & others, 2023). LLM can capture context and nuances in language, thus being able to provide precise and consistent responses. LLM can be used in various fields, such as language translation, content creation, and chatbot development. Its training process involves complex deep learning algorithms, which allow the model to learn from patterns present in the data. Currently, there are three main types of LLM architectures namely Encoder-only, Decoder-only, and Encoder-decoder. All three use unsupervised learning methods to learn the data and improve model performance. Each of these architecture types is suitable for different NLP (Natural Language Processing) tasks, thus creating various LLM models with different approaches (Leng et al., 2024). With the continuous advancement of technology, LLMs are evolving and becoming more adept at generating text that sounds natural and resembles human language.

2.4 Retrieval-Augmented Generation

Retrieval-Augmented Generation (RAG) is a method that integrates LLM with external document search systems to improve the accuracy and reliability of text generation. This technique works by searching for information from external sources based on semantic similarity, thus helping the language model to minimize factual errors (hallucination), access current information, and increase transparency and traceability in its reasoning process (Lewis & others, 2020). RAG utilizes both retrieval-based and generation-based approaches in natural language processing, and has shown notable improvements in the quality of question answering systems in many fields and tasks (Youvan, 2025). By accessing external information before generating text, RAG ensures answers are more accurate, transparent, and easy to update without the need for retraining (Thombre et al., 2023). RAG can also improve user experience by generating answers that are more contextual and relatable, thus increasing user satisfaction in interacting with the system. In addition, this approach allows developers to integrate various data sources, such as knowledge sources in the form of articles and documents, which can enrich the information available to the model. RAG not only serves as a tool to improve accuracy, but also as a bridge to connect LLM with broader information systems.

2.5 Speech-to-Text

Speech-to-Text (STT) is a technology known as Automatic Speech Recognition (ASR) and is used in various applications such as virtual assistants to recognize and convert human speech into text. The STT work process consists of sound wave analysis, speech pattern recognition, and translation of the pattern into text that can be read by

humans or processed by computer systems (Millett, 2021) (Rodríguez & Cox, 2023). The technology is capable of recognizing multiple languages and accents, making it suitable for use in multilingual environments. STT also plays a role in improving accessibility, especially for people with hearing impairments.

2.6 Text-to-Speech

Text-to-Speech (TTS) is a technology that converts text into clear, natural-sounding synthetic speech. The TTS process involves three main steps: first, the text is converted into phonemes using a phoneme converter; second, the phonemes are translated into mel-spectrograms; and third, the mel-spectrograms are converted into sound waves through a vocoder model. The main goal of TTS is to produce sounds that are not only easy to understand, but also have intonation and rhythm like human speech (Pratama et al., 2025). The utilization of TTS allows the creation of applications, such as websites, that serve to ease human tasks and support those with special needs (Kreativitas & Informatika, 2023). In addition, this technology can be applied in various fields, such as education, entertainment, and customer service, further expanding access to information for all. With advances in artificial intelligence and machine learning, the quality of the sounds produced by TTS can get closer to the human voice, making it a very useful tool in modern communication.

2.7 Conceptual Framework

Figure 1 below is an overview of the conceptual framework of the virtual assistant information system.
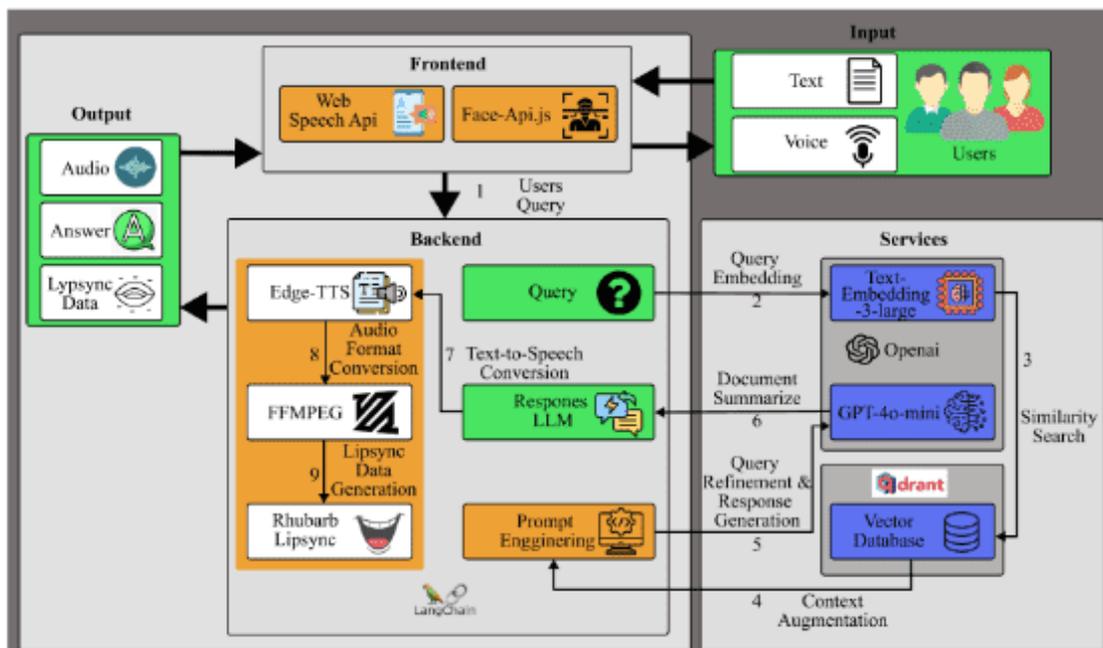

Figure 1. Conceptual Framework

The following is an explanation of the conceptual framework of this system:

1. Input: This part is the starting point of interaction between users in this case students, prospective students, staff, and lecturers. Questions that are input can be in the form of text and also through voice.

2. Frontend: in this frontend section the user input data will be processed, if the input is in the form of voice then, it will be converted first into text using the Web Speech API, and there is also Face-API.js which functions to detect the user's gender in real time. Then the user input will be sent to the backend for further processing.

3. Services: in this system using supporting services for the RAG process, there are two services used, namely Openai and Qdrant. The user query is converted into a vector representation using text-embedding-3-large, which is then used to search for similarity in the vector database. The GPT-4o-mini model will generate answers that match the context and user query.

4. Backend: in this backend process, Prompt Engineering will compose the right prompt based on the user's query and the context similarity generated by the vector database to provide specific instructions to the LLM model, then the LLM model will return a response in the form of an answer according to the context and user query. The LLM answer results are then converted into audio using text to speech, then the audio format is adjusted to produce lipsync data.

5. Output: The output generated from the backend includes three components, namely audio files, LLM answers and lipsync data, which will be sent and processed by the frontend, to provide appropriate information for users.

III. RESEARCH METHODOLOGY

3.1 Research Method

The process of developing a system can be applied through a number of development methods, and the selection of this development method depends on several factors, such as the needs of a project, the number and capacity of the research team, and so on. In this study, researchers chose to use the Extreme Programming (XP) method in the process of developing this system. Extreme Programming is an Agile software development methodology that focuses on providing high-quality software through continuous feedback, collaboration, and adaptation. This method tends to use an object-oriented approach and the target of this method is a team formed on a small to medium scale. This method is also suitable when faced with unclear requirements or very rapid changes in needs (Nurlailah & Wardani, 2023). The following is a brief explanation of the 4 stages in the extreme programming method that will be carried out by researchers:

1) Planning, the researcher will take the first steps in system development where there are a number of planning activities carried out, namely identifying problems, analyzing needs to determine the schedule for implementing system development.

2) Design, where at this stage a modeling process is carried out which includes modeling the system, architecture, and database. System and architecture modeling is described using Unified Modeling Language (UML) diagrams while database modelling is structured using flowcharts.

3) Coding, at this stage the previously designed modeling begins to be implemented into the form of a user interface by utilizing a programming language. The programming language used is Python while database management uses Qdrant.

4) Testing, after the coding stage is complete, the next stage is system testing. This test aims to find out what errors arise when the application is run, as well as ensuring that the system developed is in accordance with user needs.

IV. DISCUSSION AND IMPLEMENTATION

4.1 *Use Case Diagram*

For this system development, researchers used UML to describe the work of the virtual assistant information system. Use Case Diagrams are created by researchers to show how the interaction between the system and the user. In addition, this diagram is also used to show the features that can be accessed by admins and users. The following is a use case diagram that describes the virtual assistant information system that will be developed by researchers.
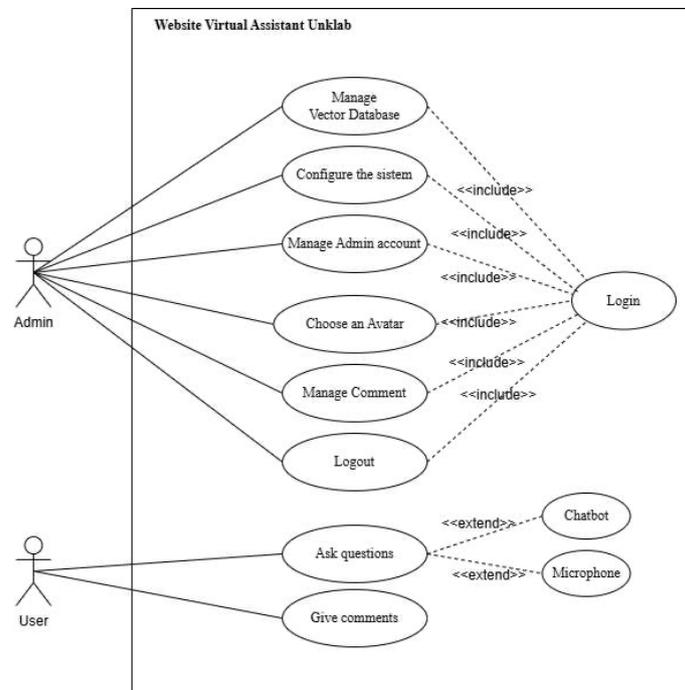

Figure 2. Use Case Diagram

Figure 2 use case diagram, there are two actors, namely admin and user, with different roles and access. Admin chooses full access to the system, starting with the login process to enter the admin menu page. Admins can manage the vector database, configure the system, manage admin accounts, choose avatars, manage comments and logout. Meanwhile, users choose limited access, namely users can ask questions via chatbot and microphone and can provide comments about the system.
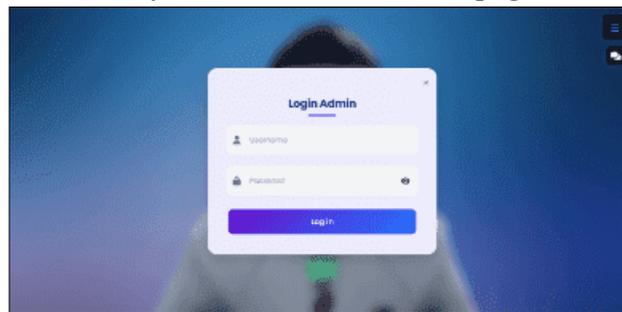
4.2 *Implementation*

These are some examples of system interface implementations developed by the researcher as part of this study. Each implementation is designed to meet the needs of the system, with the aim in ease of use and overall system performance in accordance with the research objectives.
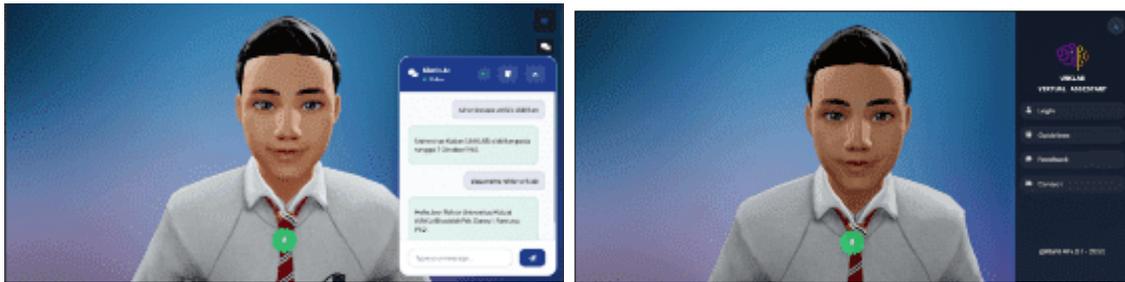
Figure 3. Virtual Assistant & Sidebar Menu

Figure 4 is the implementation of the virtual assistant interface based on the design made earlier. This page contains 3D animation, a microphone button to receive sound and the system can provide answers after the user asks questions, as well as a camera display that is used to detect user gender. This implementation of the chatbot interface is designed for users to interact with the system. Users can ask questions through the chat field provided. After typing the question, the user can send the question by pressing the send button, then the answer to the question will be displayed. the implementation of the sidebar page interface which serves as the main navigation for users. This sidebar is designed so that users can easily access the features and pages available in the system.

Gambar 4. Login Admin

Figure 4 shows the implementation of the admin login interface, which is used to securely access the system. Admins must enter their registered username and password to log into the system and manage the available features. The implementation of the database vector management interface can be performed online. The admin can manage data collections in the system, the admin can also create new collections by specifying the name and type of collection, and perform actions such as viewing, updating, and deleting available collections. Figure 5 shows the implementation of the system settings interface, which allows the admin to customize the system settings as needed. This page provides options to select the LLM model used, set the model temperature, structure the prompt, select the TTS voice, and update the active Qdrant collection.
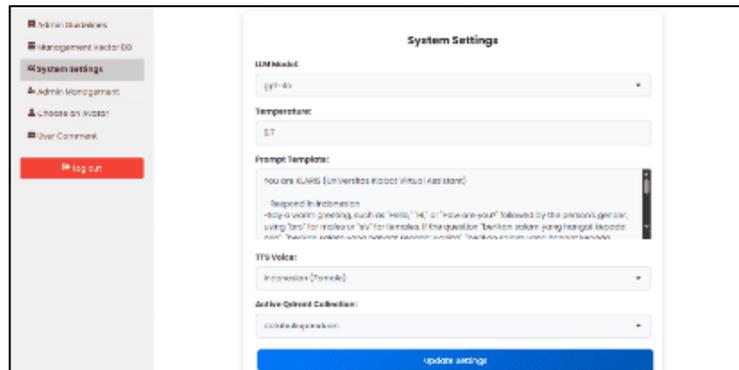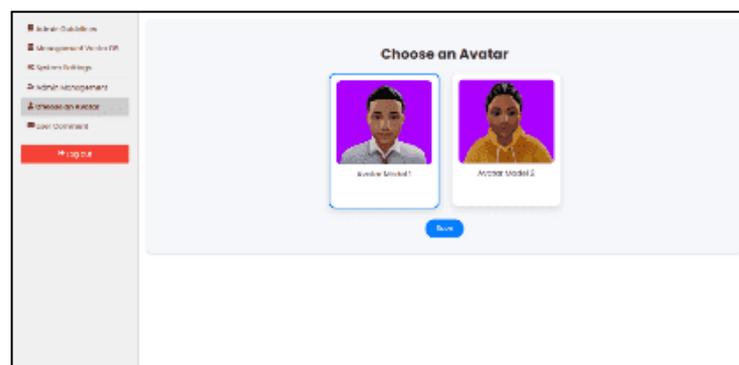
Figure 5. System Settings

Figure 6. Choose an Avatar

Figure 6 displays the implementation of the choose an avatar interface. Admins can choose the gender of the avatar to be used in the virtual assistant display.

4.3 Testing

The test used in the virtual assistant information system is black box testing (Table 1). Software testing method was proposed using black box testing which focus on testing system functions from the user's side, without paying attention to the internal structure or program code used. This test focuses on the input provided by the user and the output produced by the system. This is done by providing various input scenarios, then it will be evaluated whether the resulting output is as expected based on the functional criteria of the virtual assisstant information system.

Tabel 1. Black-box Testing

Input	Process	Expected Output	Status
Virtual Assistant	- Open the Virtual Assistant feature - User speaks through the microphone or types in the chatbox	Virtual Assistant responds to the user's question	Success
Object Detection	- Access the camera - Detect the user's face - Run gender classification algorithm	Gender is successfully detected	Success
Chatbot	- User types a question - System processes and generates a response	Chatbot responds to the user's question	Success

Sidebar	- Click on the sidebar menu navigation buttons	Sidebar menu appears and works properly	Success
Guideline	- Click the user guide menu	Display the system usage guide	Success
Add Comment	- User fills and submits the comment form	Comment is successfully saved into the system	Success
Contact	- Click the contact menu	Displays contact information	Success
Login Admin	- Admin enters username and password, then clicks login	Admin successfully logs into the system	Success
Admin Guidelines	- Admin open the admin guide menu	Displays the admin usage guide	Success
Management Vector Database	- Admin opens the vector database menu - Create, view, update, or delete vector collections	Vector database collections are successfully managed	Success
System Settings	- Admin configures the LLM model, temperature, prompt template, and TTS voice	System settings are updated successfully	Success
Admin Management	- Admin open the admin management menu - Add or edit admin account data	Admin accounts are successfully added or edited	Success
Choose an Avatar	- Admin selects an avatar from the available list	Avatar is successfully selected	Success
User Comment	- Admin opens the user comments list - Deletes specific comments	User comments are successfully viewed ordeleted	Success

The performance test results of the main features also has been reviewed in the virtual assistant system. Testing is done with two types of questions and inputs used. Short/Single Question is a question that only focuses on one information, so that the answer output can be faster. While Complex/Multiple Question is a type of question that consists of two or more questions in one sentence, this question asks for answers to several information at once, so it takes longer for the answer output. It can be concluded that this virtual assistant system successfully provides appropriate answers to questions related to campus information and also simple math. However, the system cannot answer questions outside of campus information, because the system has given restrictions in answering questions. The resulting response time also varies depending on the input used, how complex the question is, the internet connection, and the hardware used.

V. CONCLUSION AND FUTURE WORK

5.1 Conclusion

Based on the results of the research that has been conducted, it can be concluded that the virtual assistant information system can provide answers that are relevant and in

accordance with the questions asked by the user. The system utilizes Retrieval Augmented Generation (RAG) technology to help improve response quality by combining natural language processing as well as access to external information sources. In addition, there is a Speech to Text (STT) feature that allows users to ask questions using voice and not just through the chatbox. The system also supports Text to Speech (TTS) technology, which allows the system to provide voice output. All features that have been integrated into the system, including academic information and general information related to Klabat University, can be run properly and can be accessed by users according to the function and purpose of each feature.

5.2 *Future Work*

Although this research was successful in implementing RAG and TTS in the virtual assistant information system, there are several suggestions that researchers can provide as consideration for future researchers in connection with the development of research on virtual assistant information systems in the future. bu tmterialfor thr (i) Expanding the ability of many languages other than Indonesian, especially specifically in the TTS section. (ii) It is also necessary to optimize the performance of the model used, such as selecting a model that is more efficient and in accordance with system needs to increase response speed. (iii) Development of a better server infrastructure, so that the system can handle larger data loads and increase faster and more stable response times. (iv) It is expected that the virtual assistant can be equipped with a history feature that allows users to look back at previous questions and answers. (v) The virtual assistant is also expected to be integrated with the campus webpage to expand access to the information system. (vi) Researchers also suggest testing and evaluating the user experience to find out which parts need to be improved over time and the development of user needs.

REFERENCES

Alifianda, S., & Djutalov, R. (2024). Perancangan Sistem Informasi E-Commerce Berbasis Web Untuk PT Sentra Meta Fiber Dengan Metode Agile. *Buletin Ilmiah Ilmu Komputer Dan Multimedia*, *1*(5), 683–693.

Andriyan, W., Septiawan, S. S., & Aulya, A. (2020). Perancangan Website sebagai Media Informasi dan Peningkatan Citra Pada SMK Dewi Sartika Tangerang. *Jurnal Teknologi Terpadu*, *6*(2), 79–88. https://doi.org/10.54914/jtt.v6i2.289

Anjelina, K. (2025). *Pengembangan Produk Virtual Reality Dengan Kemajuan Teknologi*.

Arifiyanti, A. A., Permatasari, R., Rezha, A., & Najaf, E. (2023). KLIK: Kajian Ilmiah Informatika dan Komputer Asisten Virtual sebagai Pusat Layanan Informasi pada Perguruan Tinggi di Indonesia - A Systematic Literature Review. *KLIK (Media Online)*, *4*(2), 740–745. https://doi.org/10.30865/klik.v4i2.1222

Azis, A., & Pramono, P. (2024). Optimalisasi Potensi Digital Menuju Era Industri 5.0 dengan Asisten Virtual. *Jurnal Kajian Dan Penalaran Ilmu Manajemen*, *2*(2), 46–53. https://doi.org/10.59031/jkpim.v2i2.400

Fitriawati, L. S., Bijaksana, A., Negara, P., & Nyoto, R. D. (2020). Implementasi Text To Speech Pada Website Menggunakan Metode Shallow Parsing. *InfoTekJar: Jurnal Nasional Informatika Dan Teknologi Jaringan*, *5*(1), 61–69.

Gao, Y., & others. (2023). *Retrieval-Augmented Generation for Large Language Models: A Survey*.

Inayah, M. R. N., Nurhayadi, & Idris, M. (2024). Pengaruh Pemanfaatan Teknologi Informasi terhadap Pemahaman Konsep Matematis Siswa SMP Negeri 9 Kota Palu Kelas VIII. *Jurnal Y*, *7*, 707–717.

Kanivia, S. G. B. A., Puspitarini, D. A., Dewi, D. K., Akbari, S., & Chandra, A. K. (2024). Implementasi Teknologi Informasi terhadap Kualitas Audit Internal. *Jurnal X*, *14*, 170–178.

Kreativitas, J., & Informatika, M. (2023). *Artificial Intelligence Vision Dalam* (pp. 169–172).

Leng, Q., Portes, J., Havens, S., Zaharia, M., & Carbin, M. (2024). *Long Context RAG Performance of Large Language Models*.

Lewis, P., & others. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, *2020-December*.

Lu, Z. (2024). Large Language Models: Development in Model Scale and Challenges. *Journal Z*, *0*, 154–161. https://doi.org/10.54254/2755-2721/114/2024.18276

Millett, P. (2021). Accuracy of Speech-to-Text Captioning for Students Who are Deaf or Hard of Hearing. *Journal of Educational and Pediatric Habilitative Audiology*, *25*, 1–13.

Nurlailah, E., & Wardani, K. R. N. (2023). Perancangan Website Sebagai Media Informasi Dan Promosi Oleh-Oleh Khas Kota Pagaralam. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, *8*(4), 1175–1185. https://doi.org/10.29100/jipi.v8i4.4006

Pratama, A. R., Wabula, F., Ilmandry, H., Isabela, M. L., & Sianipar, R. (2025). *Literature Review The Impact of Machine Learning in Modern Industries*.

Rodríguez, L. M., & Cox, C. (2023). Speech-to-text recognition for multilingual spoken data in language documentation. *Comput. 2023 - 6th Workshop on Use of Computational Methods for the Study of Endangered Languages Proceedings*, 118–123.

Soetiyono, A., Kurnia, Y., & Kurnia, R. (2024). Pengaruh Penggunaan Chatbot dan Asisten Virtual terhadap Peningkatkan Kepuasan Pelanggan serta Dampaknya terhadap Pengambilan Keputusan Pembelian. *ECo-Buss*, *6*(3), 1367–1381. https://doi.org/10.32877/eb.v6i3.1169

Susilawati, T., Yuliansyah, F., Romzi, M., & Aryani, R. (2020). Membangun Website Toko Online Pempek Nthree Menggunakan PHP dan MySQL. *Jurnal Teknik Informatika Mahakarya*, *3*(1), 35–44.

Thombre, S., Bhattacharyya, P., & Jyothi, P. (2023). *Survey: Text-to-Speech Synthesis*.

Youvan, D. C. (2025). *Retrieval-Augmented Generation (RAG): Advancing AI with Dynamic Knowledge Integration*. https://doi.org/10.13140/RG.2.2.30888.89606