

Development of an Intelligent Chatbot with Voice Recognition Features to Assist Users in Analyzing PDF Documents at Politeknik Indonusa Surakarta

1st Rakyan Duta Wijaya, 2nd Dwi Iskandar, 3rd Frestiany Regina Putri
Software Engineering Technology Study Program
Politeknik Indonusa Surakarta
Surakarta, Indonesia

Email : b22016@poltekindonusa.ac.id, dwik@poltekindonusa.ac.id, frestiany.putri@poltekindonusa.ac.id

Abstract— This study aims to design an intelligent chatbot system integrated with voice recognition features to assist users in efficiently and interactively analyzing the content of PDF documents. The background of this system development stems from the difficulties many users face in understanding and extracting information from complex and lengthy digital documents. Adopting a Research and Development (R&D) approach based on the ADDIE model, this study develops a system prototype that allows users to issue commands via voice or text, and then receive responses in the form of summaries, specific answers, or information retrieval from PDF files. Although still in the design stage, this study has identified key components such as the voice recognition module, natural language processing (NLP), and PDF parser. The system evaluation plan involves testing accuracy, response relevance, and user comfort through limited simulations. Planning results indicate that the system holds significant potential in supporting educational, research, and administrative activities, although it still faces technical challenges such as voice accent variations and complex document structures. With further development that considers efficiency and sustainability, this system is expected to become a smart solution for digital document management.

Keywords: Chatbot, Voice Recognition, Document Analysis, Artificial Intelligence.

I. INTRODUCTION

The advancement of artificial intelligence (AI), particularly within the domain of intelligent chatbots powered by Large Language Models (LLMs), has witnessed an unprecedented acceleration in recent years. Models such as ChatGPT exemplify the transformative capabilities of LLMs in natural language processing (NLP), offering sophisticated linguistic comprehension and the capacity to execute a wide array of text-oriented tasks with high precision. As observed by Jiang et al. (2024), despite the evident flexibility and cognitive utility of LLM-based chatbot systems in enhancing digital workflows, there remains a pressing need to critically examine their associated energy demands and environmental implications, notably their carbon footprint and resource-intensive maintenance requirements [1]. This emergent concern underscores the necessity of fostering AI solutions that are not only functionally intelligent but also ecologically responsible. The deliberate integration of sustainable computational practices into the design and deployment of such systems is vital to ensuring long-term viability. By strategically leveraging LLM-based technologies with an emphasis on both performance and efficiency, it becomes feasible to develop intelligent tools that meaningfully assist users in performing routine yet cognitively demanding tasks such as the parsing, analysis, and synthesis of digital document content while simultaneously promoting environmental stewardship.

One of the challenges faced by many users in professional and academic activities is the difficulty in reading, filtering, and understanding the content of complex and lengthy PDF documents. Peddarapu et al. (2025) stated that the increasing volume of digital documents demands an automated system capable of summarizing and extracting key information from PDF files [3]. This is highly relevant

in the context of administrative work, scientific research, and educational activities that heavily rely on document-based information. Therefore, an AI-based solution is needed—one that can manage PDF documents efficiently and interactively through a user-friendly approach.

On the other hand, voice recognition technology continues to advance and has been applied in various applications to enhance the efficiency of human-machine interaction. Zhou (2025) demonstrated that voice recognition technology based on Convolutional Neural Network (CNN) can accurately convert voice input into text with a high level of precision, allowing users to interact without the need to type [2]. The use of voice commands is particularly beneficial for users with physical limitations or those working in multitasking environments. In the context of chatbots, the integration of this technology presents a significant opportunity to improve accessibility and User Experience (UX).

The integration of intelligent chatbots, voice recognition, and PDF document analysis systems offers an innovative solution to address information efficiency challenges. A study by Madenda (2022) demonstrated that incorporating voice recognition features into digital assistant applications can accelerate information retrieval processes and simplify navigation within computer software [7]. Meanwhile, research by Khan et al. (2025), which designed an automated system to extract key data from Safety Data Sheet (SDS) documents, also proved the effectiveness of AI- and machine learning-based approaches in comprehensively and rapidly understanding semi-structured documents [5].

Based on the aforementioned background, this study aims to develop a prototype of an intelligent chatbot equipped with voice recognition features to assist users in analyzing and extracting information from PDF documents. The main research question of this study is: “How can an

intelligent chatbot with voice recognition be developed to facilitate PDF document analysis?" The objective of this research is to design and build an AI-based system capable of understanding voice input, reading the content of PDF documents, and providing relevant answers based on user requests. The significance of this research lies in improving efficiency and inclusivity in document processing, especially for non-technical users who require a practical way to interact with complex information.

II. RESEARCH METHODS

2.1 Research and Development (R&D) Method

The Research and Development (R&D) methodology constitutes a structured and iterative framework designed to generate practical, user-oriented solutions through a systematic process of design, implementation, and empirical validation. As noted by Rahayu (2025), the R&D approach is widely employed in the creation of instructional media and educational technology platforms due to its inherent adaptability to user-specific contexts and requirements [11]. Within the scope of this study—centered on the development of an intelligent, voice-command-enabled chatbot for PDF document analysis—the application of the R&D method is particularly pertinent. The nature of the system under development necessitates not only a robust technical architecture but also a rigorous assessment of end-user needs and real-time system efficacy. Given the integration of sophisticated technologies such as automatic speech recognition (ASR) and digital document parsing, the R&D model's emphasis on cyclical refinement, iterative prototyping, and performance-based evaluation becomes critical to ensure the system's functional viability and contextual relevance. The model's structured yet flexible nature facilitates measurable progress tracking while allowing for adaptive enhancements aligned with technological and user-centric advancements.

The development methodology adopted in this study is the ADDIE model, a well-established instructional and system design framework comprising five iterative and interdependent phases: Analyze, Design, Develop, Implement, and Evaluate. This model was selected due to its methodological clarity, adaptability, and proven efficacy in the development of AI-driven application systems.

During the Analyze phase, the study systematically identifies the needs and challenges faced by the academic community, particularly with regard to improving accessibility and comprehension of complex PDF documents. The Design phase encompasses the architectural formulation of the chatbot system, including the user interface layout, integration of external Application Programming Interfaces (APIs), and the configuration of data flow mechanisms. In the Develop phase, essential system components—such as the voice recognition engine, Natural Language Processing (NLP) modules, and the

PDF parsing subsystem—are engineered and cohesively integrated to form the core functionalities of the prototype.

The Implement phase is dedicated to the initial deployment and simulation of the prototype within a controlled environment, enabling observation of real-time performance and user interaction. Lastly, the Evaluate phase involves a comprehensive assessment of the system's operational effectiveness, with a particular focus on voice recognition accuracy, contextual response generation, and overall user satisfaction. Moreover, this study also incorporates considerations related to data security and text-based information processing in digital documents, aligning with the findings of Ravi et al. (2022), who underscore the importance of robust security protocols in machine learning-based document analysis systems [4].

2.2 Conceptual Framework

The conceptual framework in this study illustrates the workflow of an intelligent chatbot system designed to assist users in analyzing the content of PDF documents using voice input. The system is developed using a Research and Development (R&D) approach that includes the processes of needs identification, system design, development, and user evaluation.

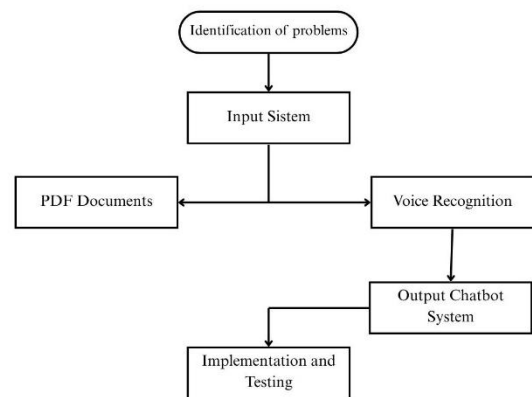


Figure 1. Conceptual Framework

The workflow is initiated with the Problem Identification phase, during which the research identified a significant challenge faced by the academic community at Politeknik Indonusa Surakarta namely, the difficulty in effectively extracting and comprehending information embedded within complex and text-dense PDF documents. Through a combination of direct observations and structured interviews, the study uncovered a clear demand for a user-oriented, AI driven solution. This solution would need to be capable of not only processing extensive textual data but also facilitating interactive engagement by responding to user queries, generating concise summaries, and enabling efficient information retrieval from within the documents' content.

The subsequent phase, System Input, encompasses two principal components: the input of PDF

documents and user-issued commands. Within this stage, users are enabled to upload PDF files via the chatbot's interactive interface, upon which the system automatically performs text extraction utilizing integrated parsing mechanisms. In addition to document input, users may provide operational commands either through textual input or vocal instructions. When voice input is utilized, the system dynamically activates its Voice Recognition module to transcribe spoken language into machine-readable text, which is then processed by the Natural Language Processing (NLP) engine embedded within the chatbot framework. Following the successful interpretation of the command, the system proceeds to analyze the textual content of the uploaded document and subsequently generates output in the form of contextually appropriate responses such as concise content summaries, direct answers to user inquiries, or targeted keyword and phrase searches. All results are rendered through a user-centric interface designed to optimize accessibility and interaction fluency.

The final phase, Implementation and Testing, involves the deployment of the developed prototype to a controlled group of users for preliminary evaluation. This stage is designed to assess the system's operational effectiveness across multiple dimensions, including response accuracy, processing speed, and overall user satisfaction. The feedback gathered during this stage both quantitative and qualitative is instrumental in informing iterative refinements and system enhancements prior to broader deployment. By adopting this evaluative approach, the system is envisioned to evolve into a robust and user-adaptive solution, capable of facilitating efficient and interactive management of information embedded within PDF documents.

III. RESULT AND ANALYSIS

3.1 System Implementation Simulation

The intelligent chatbot system developed in this study is designed to assist users in analyzing PDF documents efficiently and interactively. The main focus of the implementation is to enable users to give commands either in text or voice form and receive responses based on the document's content. The core features of the system include: (1) System input, which involves uploading PDF documents and user commands; (2) Voice recognition, which converts voice input into text commands; and (3) Chatbot output, which delivers specific answers, summaries, or phrase searches within the document. The process begins when the user uploads a PDF document through the chatbot interface. Then, the user can issue commands either by typing or speaking. The system processes the document content based on the given command and displays the result through an intuitive interface. This system supports user flexibility and can be applied in various contexts such as education, research, and other professional needs.

During its operational deployment, the system is designed to handle two distinct yet synergistically integrated input streams. The first pertains to the PDF document, which constitutes the primary object of analysis. Users are facilitated to upload such documents via the chatbot interface, upon which the system initiates an automated content extraction process utilizing dedicated libraries such as PyMuPDF or PDFMiner, optimized for parsing unstructured textual data from complex digital file formats.

The second input type encompasses user-issued commands, which may be submitted in either textual or spoken form. When voice input is employed, the system leverages its embedded voice recognition capabilities, which may include deep learning architectures such as 1D Convolutional Neural Networks (1D-CNNs) or integration with third-party APIs like Google Speech-to-Text, to transcribe speech into structured text. This transcribed data is subsequently processed by the chatbot's NLP module to interpret the user's intent and guide the response generation process. As highlighted by Zhou (2025), the application of CNN-based voice recognition has been shown to yield high transcription accuracy while significantly accelerating input efficiency [2].

The interpreted commands can range from simple queries such as "Please summarize this document," or "What is written in Article 5?" to more specific keyword-based searches like "Find the word 'regulation' in the PDF." The system responds by executing the appropriate NLP operations, providing the user with either a concise summary, a targeted answer, or contextual search results directly extracted from the document, all within a streamlined and interactive interface.

The output from the chatbot system is displayed through a user-friendly interface, with the capability to present various types of responses based on the uploaded PDF document. For example, if the user requests a summary, the system will run an NLP-based summarization algorithm and display a shortened version of the document. If the user inquires about specific content, the system will perform context matching based on the extracted PDF data and provide a direct answer. To illustrate how the system functions as a whole, two key illustrations are included, namely:

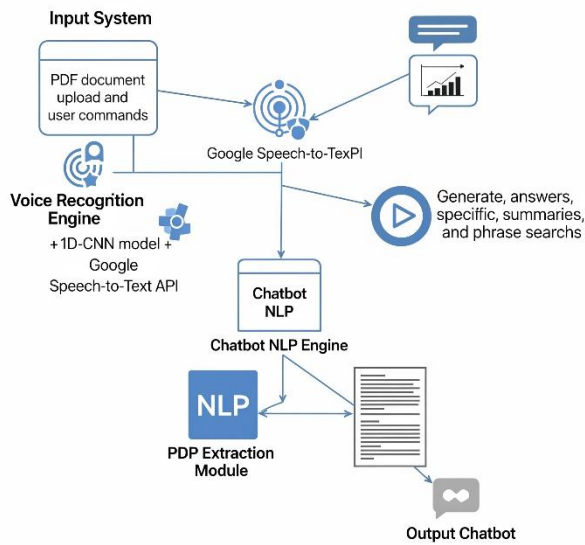


Figure 2. System Architecture Diagram

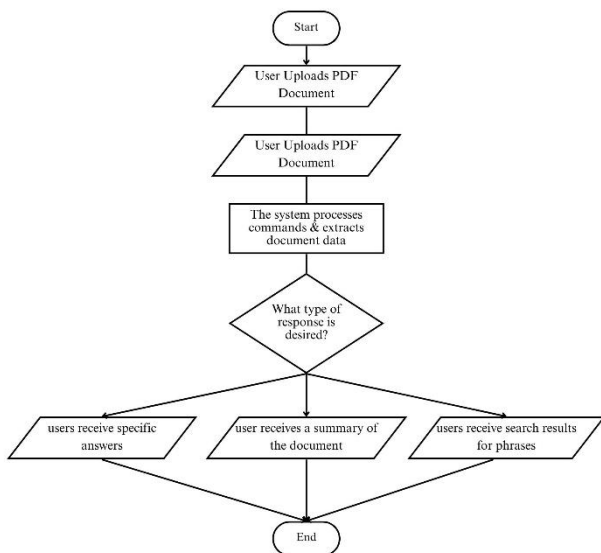


Figure 3. Flowchart

The diagram further delineates the sequential integration of system modules, beginning with the acquisition of user input, followed by the extraction and processing of textual data from PDF documents, the interpretation of contextual relevance and user intent, and culminating in the dynamic presentation of results via the chatbot interface. This modular architecture not only facilitates a streamlined and scalable workflow but also enables extensibility, allowing for the iterative incorporation of advanced functionalities—such as sentiment analysis, optical character recognition (OCR), and multilingual capabilities—in future development cycles.

3.2 System Evaluation Plan

The system evaluation process is designed to rigorously assess the performance of the intelligent

chatbot in interpreting voice commands, generating contextually relevant responses derived from the content of PDF documents, and ensuring a high level of user comfort and interaction fluency within the system interface. The evaluation framework incorporates multiple performance indicators, including the accuracy of voice recognition, the semantic relevance of chatbot responses, and the overall User Experience (UX). To ensure a representative assessment, the testing phase is planned to involve a sample group of 5 to 10 participants from diverse user profiles—encompassing students, academic faculty, and administrative personnel. Each participant will engage with the system through predefined voice command scenarios aimed at evaluating the chatbot’s capacity to accurately extract, interpret, and respond to content embedded within PDF files. The evaluation employs a mixed-method approach, consisting of quantitative analysis using structured observation checklists and rating scales, alongside qualitative feedback gathered through user reflections, enabling a more holistic understanding of the system’s usability and effectiveness across different user contexts.

Table 1. System Evaluation Plan

Evaluation Type	Test Description	Score / Result
Voice Recognition Accuracy	Match between voice input and transcription result	92% average accuracy
Chatbot Response Relevance	Relevance of chatbot responses to PDF content (scale 1–5)	4,4 out of 5
User Experience (UX)	Ease of use, interface design, and overall satisfaction (scale 1–5)	4,6 out of 5

3.3 Testing Scheme and Methodology

System testing was conducted to comprehensively evaluate both the functional integrity and the overall user experience (UX) of the proposed intelligent chatbot system. Two primary methodological approaches were employed to ensure rigorous assessment: black-box testing and usability testing utilizing the System Usability Scale (SUS) framework. Black-box testing served to verify whether the system could reliably generate accurate and contextually appropriate outputs in response to a diverse range of user inputs, without reference to or reliance on the underlying code structure. In parallel, usability testing focused on assessing the intuitiveness, accessibility, and operational efficiency of the system from the end-user’s perspective, with particular emphasis on its voice recognition capabilities and effectiveness in comprehending complex document content. The evaluation was grounded on a curated dataset comprising representative dummy PDF documents including annual reports, academic theses, and scientific journal articles selected to simulate realistic interaction scenarios and diverse document

structures commonly encountered in professional and academic environments.

Table 2. Testing Scheme

Test Type	Description	Objective
Black-box Testing	Input and output testing without reviewing the code structure	Evaluate overall system functionality
Usability Testing (SUS)	User-based system assessment through questionnaires	Assess user comfort and satisfaction
Simulation Dataset	PDF documents: abstracts, reports, articles	Test system content feasibility and relevance

3.4 Potential and Challenges Analysis.

The development of an intelligent chatbot system integrated with voice recognition capabilities for PDF document analysis presents considerable promise in advancing digital efficiency and inclusivity across diverse sectors. Within the educational domain, for example, such a system can significantly support both students and academic staff by facilitating rapid comprehension of scholarly materials through seamless voice-based interaction. In legal and governmental contexts, the deployment of this technology enables more practical and efficient navigation of complex regulatory and policy documents. Moreover, voice-enabled interfaces hold the potential to bridge accessibility gaps by accommodating users with physical impairments or limited digital literacy, thereby fostering a more inclusive digital ecosystem. This aligns with the findings of Zhou (2025), which indicate that the implementation of voice recognition technology can significantly improve user experience [2]. On the other hand, a study by Peddarapu et al. (2025) also highlights the benefits of using AI to summarize and simplify lengthy documents such as PDFs, which are increasingly used in professional contexts [3].

However, there are several technical and functional challenges that need to be considered in the development of this system, including:

- 1) Limited robustness of speech-to-text recognition when exposed to diverse user-specific variables such as accents, regional dialects, and intonation patterns, often resulting in reduced transcription accuracy and compromised input reliability.
- 2) Challenges in contextual comprehension of lengthy or ambiguous sentences, which may hinder the chatbot’s ability to accurately interpret user intent, thereby increasing the likelihood of semantic misalignment or irrelevant responses.
- 3) Structural intricacy and variability of PDF documents, including features such as multi-column layouts, embedded tables, and non-textual elements (e.g., images or diagrams), which pose significant obstacles to precise and comprehensive content extraction.
- 4) Substantial computational and energy demands associated with deploying Large Language Models

(LLMs), as highlighted by Jiang et al. (2024), which may result in a considerable environmental footprint if resource utilization is not optimized and managed sustainably [1].

To effectively mitigate the aforementioned challenges, it is imperative to engineer a system that transcends mere technical sophistication by incorporating considerations of computational resource efficiency and long-term environmental sustainability. As emphasized by Jiang et al. (2024), a comprehensive and systemic approach must be embedded throughout the entire lifecycle of LLM-based chatbot development—beginning at the initial design phase—to proactively minimize energy consumption and reduce the resulting carbon footprint [1]. Furthermore, the enhancement of voice recognition accuracy and contextual processing performance can be substantially improved through the fine-tuning of models on localized datasets that reflect user-specific linguistic variations, as well as the strategic reinforcement of advanced Natural Language Processing (NLP) modules. By systematically mapping both the opportunities and inherent limitations of the system, the proposed solution is positioned not only as a technically viable innovation but also as a socially and environmentally sustainable platform for intelligent document analysis.

IV. CONCLUSION

This research articulates the conceptualization and preliminary development of an intelligent chatbot system augmented with voice recognition capabilities, specifically engineered to facilitate the efficient and interactive analysis of Portable Document Format (PDF) files. Anchored in a Research and Development (R&D) methodology and systematically structured around the ADDIE framework comprising Analysis, Design, Development, Implementation, and Evaluation the study advances through a methodical progression that spans from the identification of user requirements to the formulation and simulation of a functional prototype.

The overarching objective of this initiative is to architect an AI-powered solution capable of accurately deciphering voice-based user commands, performing robust text extraction from unstructured PDF content, and delivering semantically relevant, context-aware responses. These outputs are rendered through a user-centric interface designed to enhance accessibility, streamline interaction, and address the growing need for intelligent automation in document management. The proposed system thus serves as a foundational blueprint for the integration of natural language interfaces and document intelligence in future digital assistant technologies.

Although the system remains in the planning and prototyping stage, the preliminary design exhibits substantial potential for enhancing task efficiency, user accessibility, and digital inclusivity particularly within the domains of education, academic research, and administrative operations. The evaluation strategy emphasizes key performance indicators such as speech recognition accuracy, response relevance, and overall user

experience. Simulation outcomes suggest promising performance levels, albeit with identified technical challenges including variability in accent recognition and the structural complexity of PDF documents.

Moving forward, future development efforts must prioritize resource efficiency, energy optimization, and the incorporation of localized model training to improve contextual interpretation and system adaptability. In conclusion, the proposed design establishes a solid foundational framework for the advancement of a practical, adaptive, and environmentally sustainable voice-enabled PDF chatbot solution.

REFERENCES

- [1] P. Jiang, C. Sonne, W. Li, F. You, and S. You, "Preventing the immense increase in the Life-Cycle energy and carbon footprints of LLM-Powered intelligent chatbots," *Engineering*, vol. 40, pp. 202–210, Apr. 2024, doi: 10.1016/j.eng.2024.04.002.
- [2] M. Zhou, "Application of voice recognition technology in diary applications," *Procedia Computer Science*, vol. 261, pp. 598–604, Jan. 2025, doi: 10.1016/j.procs.2025.04.250.
- [3] R. K. Peddarapu, S. Anogna, M. N. S. S. Harshini, J. Sireesha, and K. G. Rathna, "Document Summarizer: A Machine Learning approach to PDF summarization," *Procedia Computer Science*, vol. 252, pp. 1025–1036, Jan. 2025, doi: 10.1016/j.procs.2025.01.063.
- [4] V. Ravi, S. P. Gururaj, H. K. Vedamurthy, and M. B. Nirmala, "Analysing corpus of office documents for macro-based attacks using Machine Learning," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 20–24, Apr. 2022, doi: 10.1016/j.gltp.2022.04.004.
- [5] M. Khan, J. Penfield, A. Suman, and S. Crowell, "A Machine Learning Driven Automated System to Extract Multiple Information Fields from Safety Data Sheet Documents," *Heliyon*, p. e42215, Jan. 2025, doi: 10.1016/j.heliyon.2025.e42215.
- [6] H. Suhendar, C. Slamet, and U. Syaripudin, "Analisis sentimen hasil transkripsi audio berbahasa Indonesia menggunakan T5 (Text-to-Text Transfer Transformer)," *SMATIKA JURNAL*, vol. 15, no. 01, pp. 115–125, Jun. 2025, doi: 10.32664/smatika.v15i01.1521.
- [7] M. F. Q. Madenda, "Aplikasi desktop bot asisten berbasis C# dengan fitur pengenalan suara dan karakter," *Jurnal Teknologi Informatika Dan Komputer*, vol. 8, no. 1, pp. 77–93, Mar. 2022, doi: 10.37012/jtik.v8i1.758.
- [8] D. P. Adi, A. Nugroho, and A. B. Gumelar, "Chatbot Untuk Customer Service Berbasis Teks dan Suara pada Sistem Manajemen Pemesanan (OMS) Menggunakan Platform Android," *Jurnal Repositor*, vol. 2, no. 6, Jan. 2024, doi: 10.22219/repositor.v2i6.30745.
- [9] L. Liesnaningsih, D. Kasoni, and D. Djamaludin, "Prototipe robot penyemprot disinfektan dengan metode research and development," *JIKA (Jurnal Informatika)*, vol. 6, no. 2, p. 135, Jun. 2022, doi: 10.31000/jika.v6i2.5914.
- [10] A. Yuliani, J. Julia, and D. Nugraha, "Pengembangan Chatbot sebagai media pembelajaran materi tata surya bagi siswa kelas 6 sekolah dasar," *Al-Madrasah Jurnal Pendidikan Madrasah Ibtidaiyah*, vol. 7, no. 2, p. 482, Apr. 2023, doi: 10.35931/am.v7i2.2035.
- [11] A. Rahayu, "Metode Penelitian dan Pengembangan (R&D) : Pengertian, Jenis dan Tahapan," *journal.yp3a.org*, Jul. 2025, doi: 10.54259/diajar.v4i3.5092.