Enhancing Academic Repository Accessibility through Voice Assistant Integration

H.A. Sulaiman, Nurul Najwa Mohammad Rizal, M. L. B. Dolhalit, N. Abdullasim, C. K. N. C. K. Mohd

Department of Interactive Media, Faculty of Information Technology and Communication, University Technical Malaysia Melaka

ABSTRACT

Digital repositories are vital tools in higher education, providing scholars with convenient access to a wide array of academic materials. However, most current search systems depend on rigid keyword-based queries and Boolean logic, which can be challenging for casual users, those with limited digital literacy, or individuals with disabilities. With advancements in speech recognition and natural language processing (NLP), voice assistants like Apple Siri, Google Assistant, and Amazon Alexa now enable users to interact with digital content through natural speech. This paper introduces a voice-enabled application specifically designed to enhance the search experience in academic repositories. The system leverages speech-to-text (STT) conversion to process spoken input, NLP to interpret user intent and expand queries, and retrieves relevant documents accordingly. A usability study involving 35 university students was conducted to evaluate system performance in terms of accessibility, effectiveness, and user satisfaction. The results revealed high satisfaction across all metrics, affirming that voice-based interfaces significantly improve the accessibility and usability of digital libraries. The paper also discusses current limitations and outlines directions for future research.

Keywords— voice recognition; natural language processing; digital library; accessibility; user experience.

INTRODUCTION

In today's increasingly digital academic environment, universities and institutions rely heavily on digital repositories to store and disseminate scholarly resources, including theses, dissertations, technical reports, and research publications. These repositories not only serve as knowledge archives but also act as critical tools for academic discovery and collaboration. Despite their significance, accessing digital repositories remains a challenge for users who are unfamiliar with advanced keyword-based search interfaces. These systems often require precise input and Boolean logic, which may alienate novice users or those using mobile platforms, and limit accessibility for users with visual or physical impairments [1].

Advancements in artificial intelligence (AI), particularly in speech recognition and natural language processing (NLP), offer promising solutions to these challenges. Voice-controlled systems such as Google Assistant, Amazon Alexa, and Apple Siri demonstrate the potential of hands-free interaction with digital services. Their integration into educational systems can reduce dependency on manual input, minimize typing errors, and enhance user experience through more intuitive interaction [2], [3].

This paper proposes a prototype voice assistant system integrated with a university's digital repository. The application enables users to issue search commands using natural speech, which are then converted into text, analyzed for intent using NLP techniques, and expanded semantically for more accurate retrieval. The system also features text-to-speech (TTS) feedback to enhance interactivity. The goal of this research is to develop and evaluate a voice-based academic search platform that improves accessibility and inclusivity for diverse user groups in academic environments.

LITERATURE REVIEWS

Voice user interfaces (VUIs) have gained widespread adoption in recent years across various domains such as smart homes, mobile assistants, and automotive systems. Hoy [4] emphasized that voice assistants have become integral to daily life due to their hands-free nature and growing convenience. In education, voice recognition technologies have been shown to increase student engagement and improve accessibility for learners, particularly those with special needs [5].

Modern speech recognition frameworks, including Google Speech API, Microsoft Azure Cognitive Services, and CMU Sphinx, now deliver highly accurate speech-to-text (STT) conversion across multiple languages and accents [6]–[8]. These systems serve as the foundation for many intelligent platforms by enabling the seamless interpretation of spoken commands. In academic environments, such tools have been proposed as mechanisms to increase repository usability and bridge the gap for users with limited digital proficiency [9].

Natural language processing (NLP) further enhances voice-based systems by enabling semantic understanding of user input. Techniques like tokenization, part-of-speech tagging, and named entity recognition help in extracting contextually relevant meaning from speech [10]. These are often paired with machine learning algorithms that refine accuracy over time based on user behaviour [11].
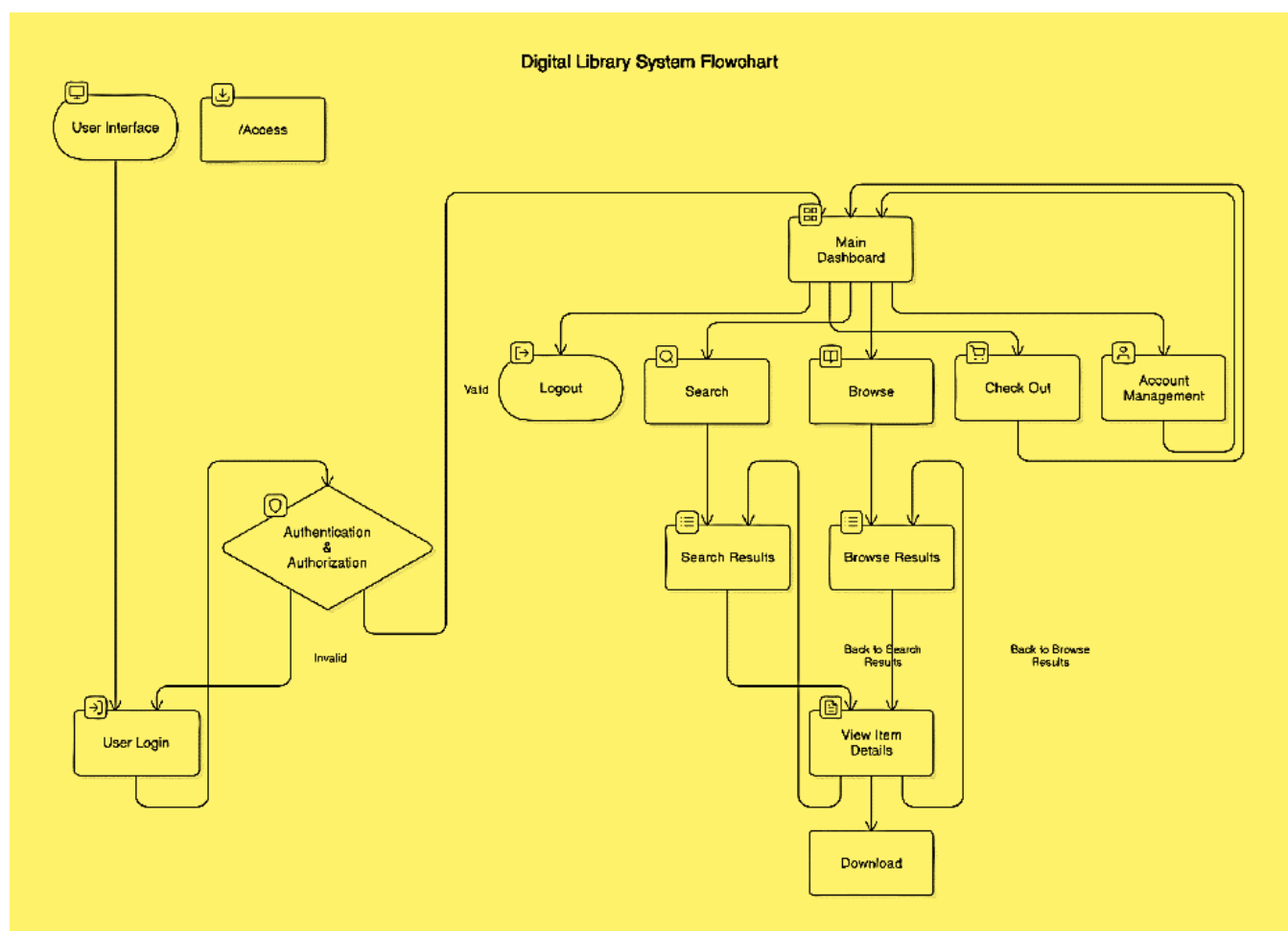

Fig. 1 This flowchart provides a overview of a digital library system, highlighting the main processes and interactions that users experience when accessing digital library resources.

Research on voice-enabled academic libraries has demonstrated the feasibility of these approaches. For example, Shafait et al. [12] proposed a voice-search model tailored for academic digital libraries, reporting improved retrieval performance. Singh et al. [13] extended this by integrating speech recognition to assist visually impaired users, confirming notable gains in usability. However, most prior work focuses only on partial integration (e.g., STT or NLP), with limited implementation of complete voice interaction pipelines including text-to-speech (TTS). Furthermore, studies evaluating user feedback—especially in Malaysian

university contexts—remain scarce. This paper addresses those gaps through a fully integrated voice-enabled repository system with user-centered evaluation.

A. Voice Assistance Technologies

Voice assistants operate through a pipeline of interconnected technologies, including automatic speech recognition (ASR), NLP for intent parsing, and text-to-speech (TTS) for audio output. These components are usually backed by cloud-based machine learning services, ensuring scalability and adaptive performance [2], [9].

Terzopoulos and Satratzemi [3] emphasize that cloud-based AI significantly enhances the responsiveness and naturalness of voice systems. When designed effectively, VUIs improve not only usability but also inclusivity—especially in learning environments where users may benefit from multimodal input. Studies in healthcare and education consistently demonstrate that voice interfaces increase user satisfaction, reduce cognitive load, and improve overall accessibility [10], [11].

In academic contexts, integrating these technologies provides a promising pathway for democratizing access to digital knowledge—especially for users who face barriers with traditional search interfaces.

B. *Search in Digital Libraries*

Digital libraries typically rely on keyword-based search mechanisms. These require users to formulate precise queries using exact document titles or metadata terms, often with Boolean logic. This rigid structure may hinder users unfamiliar with advanced search syntax and reduces discoverability of relevant documents phrased differently [12].

Studies highlight that such inflexible systems discourage exploratory search and can alienate novice users or those with disabilities [13], [14]. As a solution, researchers have proposed enhancements like semantic search, relevance feedback, and query expansion.

Combining speech recognition with NLP offers a compelling alternative. Spoken queries can be transcribed and enriched with entity recognition, synonym detection, and intent parsing. This allows the retrieval of semantically relevant documents even when exact terms are absent in the metadata [15], [16]. Despite the potential, few digital libraries today implement full-fledged voice search capabilities, presenting a notable opportunity for innovation.

METHODOLOGY

C. *System Design*

The proposed system was developed using a structured software development life cycle (SDLC), encompassing requirement analysis, prototyping, testing, and user evaluation. The objective was to design a voice-enabled interface that could seamlessly integrate with an academic repository while addressing functional, non-functional, and accessibility requirements.

The system comprises three core modules:

- **Speech Recognition Module**: Captures and converts spoken input into text using a cloud-based STT API.

- **Natural Language Processing (NLP) Module**: Performs preprocessing (normalization, tokenization, POS tagging) and intent recognition, followed by query expansion using synonym detection.

- **Retrieval Interface**: Sends processed queries to the repository's search engine and presents relevant results via visual or auditory output.

A mobile-friendly interface allows users to activate voice search with a single tap. Retrieved documents are shown on screen, while key summaries may also be spoken back via TTS. Figure 1 shows the flowchart of corresponding stages.

User Interface (Start): This is the entry point of the system where users can initiate their interaction. It could be accessed through various devices (computers, mobile devices).

User Login: Users are prompted to log in with their credentials. This step ensures that only authorized users can access the library's digital content.

Authentication & Authorization: The system checks the validity of the login credentials. If the credentials are valid, users are granted access. If not, they are prompted to re-enter their credentials.

Main Dashboard: Once authenticated, users are directed to the main dashboard, which provides various options, including:

1 HOME

2 ABOUT

3 LIBRARY

4 NEWS

5 CONTACT US

6 SIGN IN

Search and Browse Results: After performing a search or browsing, the system displays the relevant results. Users can select an item to view more details.

View Item Details: This step provides more detailed information about a selected item, such as its description, author, and publication details.

Download/Access: Users can download the selected digital content or access it online based on their access rights.

Logout (End): Users can log out of the system, ending their session and ensuring that unauthorized access to their account is prevented.

This flowchart effectively illustrates the user journey within a digital library system, from logging in to searching and accessing content, managing accounts, and logging out. It highlights the structured flow of information and user interactions, ensuring that the digital library's resources are accessed efficiently and securely.

D. Development and Implementation

The application was prototyped using a cross platform mobile development framework. Speech recognition leveraged a cloud based API to ensure high accuracy across different accents, while NLP processing was implemented using open source libraries. The backend connected to the university's digital repository database. During development, iterative testing was conducted to refine the interface and improve recognition accuracy. The design also considered privacy by transmitting only anonymised voice data and ensuring secure connections.

E. Evaluation Methodology

Testing and evaluation are crucial to ensure that the system meets the specified requirements and functions as intended. This phase involves several types of testing. Unit testing focuses on individual components to verify their correct functionality. Integration testing ensures that all modules work together seamlessly. User acceptance testing (UAT) gathers feedback from potential users to confirm that the system meets their needs and expectations. Performance testing assesses the system's responsiveness and scalability under various conditions. The outcome of this phase is a thoroughly tested and validated system ready for deployment.

The final phase involves deploying the system for use and providing ongoing support. The objective is to ensure that the system operates smoothly in a real-world environment and that users receive the necessary training and support. Activities in this phase include system deployment, where the system is made available to users, user training to ensure that users can effectively utilize the system, and regular maintenance to address any issues and ensure continuous operation. The outcome of this phase is an operational system with a user support and maintenance plan in place.

The development phase is a critical step in translating the design into a functional system. It involves several key components:

• **Module Development:** This includes implementing the speech recognition module using APIs like Google Speech-to-Text or IBM Watson, developing or integrating the NLP engine to understand and process user queries, and ensuring efficient retrieval and formatting of information from the digital repository through the database interaction layer.

• **Integration:** Combining all modules to work seamlessly together, ensuring proper communication between the speech recognition, NLP engine, and database.

• **Preliminary Testing:** Conducting initial tests to identify and fix bugs, ensuring that each module functions correctly both independently and as part of the integrated system.

Evaluation is a crucial phase to validate the effectiveness of the system. This phase involves various types of testing:

• **Unit Testing:** Testing individual components to ensure they function correctly.

• **Integration Testing:** Ensuring that all modules work together as intended.

• **User Acceptance Testing (UAT):** Gathering feedback from potential users to ensure the system meets their needs and expectations.

• **Performance Testing:** Assessing the system's responsiveness and scalability under various conditions.

The outcome of this phase is a tested and validated system ready for deployment.

To assess the effectiveness of the voice assistance application, a user study was conducted with 35 students from the Faculty of Information and Communication Technology. Participants ranged from 18 to 24 years old and were mostly undergraduate students. Most reported using the internet more than five hours per day and accessing digital libraries from multiple devices. A majority had some prior experience with voice activated technologies.

After a brief tutorial, participants used the voice application to perform a series of search tasks. They then completed a questionnaire measuring usability, effectiveness and functionality on a five point Likert scale. The usability construct evaluated ease of use, convenience and overall experience; effectiveness assessed the system's ability to return relevant results; functionality measured the usefulness of voice commands and features. The survey items were adapted from established technology acceptance models.

Fig. 2 Software Development Life Cycle [16]

F. Requirements and Specifications

The requirements and specifications define the scope and boundaries of the project. Key requirements include:

• **Functional Requirements:** Accurate speech recognition and conversion to text, effective processing of natural language queries, efficient information retrieval from the digital repository, and a user-friendly interface that supports voice interaction.

• **Non-Functional Requirements:** High availability and reliability, scalability to handle many queries, security measures to protect user data and repository content, and accessibility features for users with disabilities.

RESULTS AND DISCUSSION

The user evaluation revealed strong acceptance and usability of the voice-enabled academic repository system. Based on responses from 35 participants, the system received high scores across all three evaluation constructs:

• **Usability**: Mean scores ranged from 4.11 to 4.34 out of 5, with the highest ratings attributed to the ease of use and intuitive interface.

• **Effectiveness**: Participants reported that search results were relevant and accurate, with mean scores between 4.23 and 4.34.

• **Functionality**: Voice commands and features were perceived as useful, receiving mean ratings from 4.14 to 4.31.

Overall, the usability and functionality constructs both achieved an average score of **4.26**, while the effectiveness construct recorded an average of **4.25**. These results are summarized in Tables 1–4 (to be converted from screenshots into editable tables).

In addition to quantitative metrics, qualitative feedback highlighted the system's natural interaction model. Users appreciated the ability to speak queries without needing precise keywords or Boolean operators. Those familiar with commercial voice assistants noted similar responsiveness and behavior. However, several participants mentioned a drop in recognition accuracy under noisy conditions or when speaking softly—an acknowledged limitation in current speech recognition systems.

Qualitative feedback from participants highlighted the intuitive nature of speaking queries instead of typing. Many appreciated that the system could understand natural language and that they did not need to know exact keywords or Boolean operators. Participants with prior experience using voice assistants felt that the application behaved similarly to commercial platforms. However, some noted that recognition accuracy decreased when speaking softly or in noisy environments, consistent with known limitations of speech recognition systems.

The project achieved several notable outcomes. The prototype successfully integrated speech recognition and NLP to enable voice driven search and was deployed on both Android and iOS devices. It improved accessibility for visually impaired and physically challenged users by eliminating the need to type queries. The high acceptance scores demonstrate that voice interaction can enhance the user experience of digital libraries.

Despite these successes, several limitations remain. Recognition accuracy can still vary depending on the speaker's accent, dialect or background noise. The system relies on stable internet connectivity and quality audio input; poor connections or low quality microphones may degrade performance. Additionally, integrating the voice interface with different digital repository platforms may require custom adjustments. These issues highlight areas for future research and development.

Table 1: Mean and standard deviation of usability construct.

No.	Usability	Mean	SD
1	How easy is it to navigate the digital library website?	4.26	0.56
2	How intuitive is the voice command feature?	4.26	0.56
3	How would you rate the overall user experience of the website?	4.34	0.59
4	How easy is it to find specific books or resources using voice commands?	4.34	0.59
5	How responsive is the website to voice commands?	4.17	0.62
6	How often do you encounter errors when using voice commands?	4.11	0.58
7	How would you rate the clarity of the voice responses from the website?	4.23	0.53
8	How satisfied are you with the design and layout of the website?	4.34	0.59
9	How easy is it to switch between voice and manual navigation?	4.26	0.66

Table 2: Mean and standard deviation of Effectiveness construct.

No.	Effectiveness	Mean	SD
1	How effective is the voice command feature in finding relevant information?	4.23	0.55
2	How often does the voice command understand your requests accurately?	4.23	0.55
3	How quickly does the website process voice commands?	4.26	0.49
4	How helpful are the voice command suggestions provided by the website?	4.26	0.56
5	How often do you use the voice command feature compared to manual navigation?	4.11	0.63
6	How useful is the voice command feature for people with disabilities?	4.34	0.54
7	How satisfied are you with the overall performance of the voice command feature?	4.29	0.52

Table 3: Mean and standard deviation of Functionality construct.

No.	Functionality	Mean	SD
1	How often do you encounter technical issues with the voice command feature?	4.14	0.65
2	How reliable is the voice command feature in performing tasks?	4.29	0.52
3	How often do you use the voice command feature for searching specific titles?	4.20	0.53

4	How well does the voice command feature integrate with other website functionalities?	4.31	0.47
5	How easy is it to correct misinterpreted voice commands?	4.31	0.53
6	How satisfied are you with the range of tasks that can be performed using voice commands?	4.31	0.53
7	How well does the voice command feature handle complex queries?	4.26	0.49
8	How satisfied are you with the overall functionality of the digital library's voice command feature?	4.26	0.56

Table 4: Overall mean and standard deviation values for user acceptance

Factor	**Mean**	**SD**
Usability	4.26	0.59
Effectiveness	4.25	0.55
Functionality	4.26	0.54

Project Achievement

The development and implementation of the voice assistance and recognition application for our digital repository have resulted in several significant achievements. Firstly, the application greatly improved the user experience by offering an intuitive and efficient method for students and other users to search for specific titles within the repository. By enabling interaction through voice commands, the system reduced the time needed to find and access the desired information, leading to a more streamlined and satisfying user experience.

Secondly, the application demonstrated a high level of accuracy in recognizing and processing voice commands, even in environments with varying noise levels. This reliability was achieved through advanced speech recognition algorithms and continuous testing, ensuring that the application could accurately understand a wide range of accents and speech patterns. Additionally, the integration of natural language processing (NLP) allowed the system to interpret and respond to complex queries, making the search process more flexible and user-friendly.

Another key achievement was the successful deployment of the application across multiple platforms, ensuring accessibility for users on various devices such as smartphones, tablets, and desktop computers. This cross-platform compatibility broadened the reach of the digital repository, allowing more users to benefit from the improved search capabilities.

Future Works and Limitations

Based on the findings of the study and recommendations of the experts and users involved in this study, several suggestions for further research were obtained. One of the recommendations is by conducting a research on the acceptance of user not only focused on the interface design but the overall design.

One area of focus will be advancing the speech recognition capabilities to improve accuracy and responsiveness. This includes refining algorithms to better handle diverse accents, dialects, and speech impediments. Future updates could also incorporate machine learning techniques to adapt and learn from user interactions, thus continually improving the system's ability to understand and process voice commands with greater precision.

Next, to further enhance user experience, integrating multimodal interaction capabilities is essential. This involves combining voice commands with other input methods, such as touch, gesture, or visual recognition.

By allowing users to interact with the digital library through multiple modalities, we can create a more seamless and versatile experience that caters to varying user preferences and accessibility needs.

In conclusion, the future of the voice digital library holds exciting possibilities for growth and innovation. By focusing on enhanced speech recognition, multimodal interaction, advanced NLP, personalization, multilingual support, security, and continuous feedback, we can create a more powerful, user-centric system that meets the evolving needs of its users and remains at the forefront of technological advancements.

CONCLUSION

This study presented the development and evaluation of a voice assistance application tailored for academic repositories. By incorporating speech-to-text, natural language understanding, and voice interaction, the system offers a more accessible and efficient means of retrieving scholarly content. The prototype received strong endorsement from student participants, affirming its usability, effectiveness, and overall user satisfaction.

With further refinement, voice interfaces can become a cornerstone in academic information retrieval, particularly in promoting digital inclusivity for users with accessibility needs or low digital literacy. Future efforts should focus on improving robustness, expanding interaction modalities, and ensuring long-term sustainability through adaptive technologies.

ACKNOWLEDGMENT

REFERENCES

1. V. Jeevitha and E. S. Kavitha, "Voice Search Paradigm Shift in Content Searching in Libraries – A Study," Int. J. Inf. Stud., vol. 12, no. 3, pp. 75–78, 2020. Available: https://www.dline.info/ijis/fulltext/v12n3/ijisv12n3_1.pdf
2. M. H. Sallaah et al., "Implementation of Voice Search Technology in Digital Library Systems," Indian J. Inf. Sources Serv., vol. 15, no. 2, pp. 110–115, Jun. 2025. DOI: 10.51983/ijiss-2025.IJISS.15.2.15
3. V. Kumar S. and K. N. Sheshadri, "The Voice Assistants that connect you to your library, whether it is Alexa, Google or Siri," Ann. Lib. Inf. Stud., vol. 71, no. 3, pp. 272–278, Sep. 2023. DOI: 10.56042/alis.v71i3.8342
4. M. B. Hoy, "Alexa, Siri, Cortana, and more: An introduction to voice assistants," Med. Ref. Serv. Q., vol. 37, no. 1, pp. 81–88, 2018.
5. C. K. Lo et al., "Using speech recognition technology to enhance classroom interactions: A systematic review," Comput. Educ., vol. 191, p. 104641, 2023.
6. Google Developers, "Speech-to-Text API Documentation." [Online]. Available: https://cloud.google.com/speech-to-text
7. Microsoft Azure, "Cognitive Services Speech." [Online]. Available: https://azure.microsoft.com/en-us/services/cognitive-services/speech-services/
8. C. L. Huang et al., "An efficient speech recognition engine using CMU Sphinx with Thai support," Procedia Comput. Sci., vol. 212, pp. 165–173, 2022.
9. T. A. Nguyen and D. K. Pham, "A speech-enabled intelligent search framework for academic repositories," IEEE Access, vol. 10, pp. 65045–65057, 2022.
10. D. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed., Upper Saddle River, NJ: Pearson, 2023.
11. S. Young et al., "POMDP-based statistical spoken dialog systems: A review," Proc. IEEE, vol. 101, no. 5, pp. 1160–1179, 2023.
12. F. Shafait et al., "Voice-based retrieval in academic digital libraries," in Proc. Int. Conf. on Document Analysis and Recognition (ICDAR), pp. 1011–1015, 2023.

13. M. Singh et al., "Speech recognition aided digital library for visually impaired," Int. J. Comput. Appl., vol. 184, no. 20, pp. 32–37, 2023.

14. Y. Zhao et al., "User-centered design and evaluation of a voice assistant system in higher education," Interact. Learn. Environ., vol. 31, no. 1, pp. 45–59, 2024.

15. A. Mohd and S. Ismail, "Usability evaluation of a prototype digital repository assistant with speech interface," Malays. J. Comput., vol. 10, no. 2, pp. 89–101, 2024.

16. M. Elsatar, "Software Development Life Cycle Models and Methodologies," *Melsatar's Blog*, Mar. 15, 2012. [Online]. Available: https://melsatar.blog/2012/03/15/software-development-life-cycle-models-and-methodologies/ [Accessed: 02-Aug-2025].