JELS Journal of Education and Learning Sciences
E-ISSN : 2808-3695

Vol 06 No 02 2026
http://jurnal.gerakanedukasi.com/index.php/gerasi
https://doi.org/10.56404/jels.v6i2.524

Design And Development Of A Web-Based AI Captioning System For Online Video Advertisements: A Functional Evaluation

[1] Putu Eka Suryadana, [2] Gede Pradiva Adiningrat, [3] Nyoman Indah Kusuma Dewi & [4] Muhammad Panji Mahardika

[1234]Jurusan Administrasi Bisnis, Politeknik Negeri Bali, Indonesia.

[1]esuryadana@pnb.ac.id, [2]gedepradiva@pnb.ac.id. [3]ikdewi@pnb.ac.id, [4]panjimahardika@pnb.ac.id

ABSTRACT: The lack of subtitles in online video advertisements limits access to audiovisual information for deaf and hard-of-hearing users. Manual subtitle creation is often time-consuming, costly, and requires specialized software, making it impractical for many content creators. This study develops a web-based automatic captioning system by integrating the OpenAI Whisper Automatic Speech Recognition (ASR) model to automatically generate time-stamped subtitles from online video advertisements. The proposed system enables users to generate subtitles by simply providing a video URL and selecting the transcription language through a browser-based interface. The backend automatically retrieves the video, extracts its audio, transcribes speech using the Whisper API, and converts the transcription into a SubRip Subtitle (SRT) file with timestamps. The developed application was evaluated using functional black-box testing to verify its core features, including video URL input, language selection, subtitle generation, subtitle display, and subtitle download. All five tested functions passed the predefined functional test scenarios on 12 Indonesian-language advertisement videos. The system is designed to support digital accessibility for deaf and hard-of-hearing users; however, the current evaluation does not establish transcription accuracy or the effectiveness of the system for these users. These aspects require further evaluation using human-verified reference transcripts and direct user assessment.

Key words: automatic captioning, OpenAI Whisper, automatic speech recognition, web application, digital accessibility, subtitle generation, deaf users

INTRODUCTION

Individuals with hearing impairments continue to encounter considerable difficulties in accessing audiovisual content disseminated through digital platforms. While hearing audiences can interpret spoken dialogue, narration, and audio cues simultaneously, deaf and hard-of-hearing users depend largely on visual representations to comprehend video-based information (Hatami & Chegini, 2024). In the context of online advertising, verbal communication frequently conveys important details regarding product characteristics, promotional campaigns, and brand identity, and captions of adequate quality are therefore essential for deaf and hard-of-hearing viewers to access this information (Fathallah et al., 2024). Nevertheless, captioning practices for online videos remain uneven, and many videos lack captions that convey

their full audio content (May et al., 2024), which can prevent deaf audiences from fully understanding the conveyed information and restrict their equal participation in digital commerce. Although subtitles are widely recognized as an essential accessibility feature, their implementation remains inconsistent because they are commonly produced through manual transcription and synchronization, a process that demands considerable time, financial resources, technical expertise, and specialized software, making it impractical for many small businesses and independent content creators (Mocanu & Tapu, 2021).

The rapid development of Artificial Intelligence (AI), particularly in the field of Automatic Speech Recognition (ASR), has opened new possibilities for addressing these accessibility issues. Contemporary ASR models based on deep learning techniques are capable of converting spoken language into text across different languages, accents, and recording environments, thereby reducing both the effort and cost required for subtitle production (Hasini et al., 2025), although recognition error rates have been reported to vary considerably across recordings (Rai et al., 2023). Among the available ASR technologies, OpenAI Whisper has become a widely used option because of its multilingual transcription capability and automatic language identification, and it has been reported to be robust to accents and background noise (Saraf et al., 2025). Whisper produces segment-level output that has been used for speaker-labelled transcription and for subtitle synchronization (Lyu et al., 2024; Thara et al., 2024), which makes it a candidate for video captioning applications.

Despite these advancements, practical challenges remain in applying automatic subtitle generation to online advertisements. Related studies have addressed individual components of the problem. Thara et al. (2024) used Whisper to synchronize subtitles with audio; Fathallah et al. (2024) used large language models to correct ASR-generated captions for deaf and hard-of-hearing viewers and evaluated them using Word Error Rate; and Striuk and Hordiienko (2025) described an AI-based subtitle management system in which users upload video files and obtain subtitles in SRT or WebVTT format. Other studies have examined subtitle generation for video creators (Karad et al., 2024) and subtitle synchronization and positioning for deaf and hearing-impaired viewers (Mocanu & Tapu, 2021). Table 1 compares the studies whose content was examined in detail with the system developed in this study. Within these sources, none describes a workflow that retrieves an online advertisement video directly from its URL, transcribes it with Whisper, and returns a downloadable SRT file through a browser interface, and the evaluations reported differ widely, from word error rates to no experimental evaluation. This observation is limited to the sources examined and does not exclude the existence of other tools.

Table 1. Comparison of Related Studies and the Proposed System

Aspect	Thara et al. (2024)	Fathallah et al. (2024)	Striuk & Hordiienko (2025)	This study
Main focus	Subtitle synchronization using Whisper	LLM-based correction of ASR-generated captions for deaf and hard-of-hearing viewers	AI-based subtitle management system	Subtitle generation for online video advertisements
Whisper used	Yes	Not stated	Not stated (ASR model not named)	Yes
Video input via URL	Not stated	Not stated	No (uploaded video files)	Yes
Browser-based	Not stated	Not stated	Not stated	Yes
SRT output	SRT files used for input and synchronization	Not stated	Yes (SRT or WebVTT)	Yes
Evaluation reported	Not described in the source consulted	WER on a new dataset; no user study mentioned	No experimental evaluation reported	Functional black-box testing; WER/CER and user evaluation not yet performed

Note: "Not stated" means that the information was not stated in the source consulted. The comparison covers only the studies whose content could be examined and does not exclude other existing tools.

To address this gap, this research proposes the design and implementation of a web-based automatic captioning system that incorporates the OpenAI Whisper ASR model to generate subtitles for online video advertisements. AI-based processing of advertisement videos has been explored in other contexts, such as the summarization of televised campaign advertisements (Breuer et al., 2025). The proposed application allows users to create subtitles directly from a video URL using a browser-based interface, reducing the need for software installation and simplifying the subtitle creation process. Automatic captioning of images for web accessibility has been studied separately (Leotta et al., 2023); it addresses a different accessibility domain from the speech-based subtitles considered here. By integrating AI-powered speech recognition with a web interface, the system is designed to support the accessibility of online video content.

The contribution of this study lies in the implementation of a browser-based workflow that integrates video retrieval from a URL, audio extraction, Whisper-based speech recognition, and SRT generation in a single web-based process, and in the evaluation of its main functions. Rather than proposing a new speech recognition algorithm, the study focuses on the functional implementation of the workflow and its evaluation. The distinction between functional feasibility, transcription quality, and

accessibility outcomes is important because successful system execution alone does not demonstrate accurate subtitles or improved user comprehension.

This study aims to design and implement a web-based automatic captioning application for online video advertisements by integrating the OpenAI Whisper Automatic Speech Recognition (ASR) model to automatically generate time-stamped subtitles through AI-based speech-to-text technology. Furthermore, the study evaluates the functionality of the developed application through functional black-box testing to verify that its core features operate according to the predefined specifications. Through the development of this web-based captioning workflow and its functional evaluation, this research aims to provide a foundation for supporting more inclusive online advertising for deaf and hard-of-hearing communities.

Accordingly, the scope of the present study is limited to functional performance. Functional testing is used to verify whether the application operates according to its technical requirements, whereas transcription accuracy measured against human-verified reference transcripts (for example, using Word Error Rate) and evaluation with deaf and hard-of-hearing users are outside the current scope and represent important directions for further validation.

METHODS

2.1 System Design

The proposed system is implemented as a web-based automatic captioning application that enables users to generate subtitles for online video advertisements through a browser-based interface. To automate the speech-to-text conversion process, the application integrates the OpenAI Whisper Automatic Speech Recognition (ASR) model, a multilingual speech model (Saraf et al., 2025), with a web-based platform. The system architecture is represented through three key design models, including the use case diagram, activity diagram, and sequence diagram, which describe the interactions among the user, web application, backend server, and Whisper API. Instead of manually uploading video files or creating subtitles, users simply provide the URL of an online advertisement video to initiate the subtitle generation process. The application automatically retrieves the video, extracts the audio, processes the speech using the Whisper API, and produces subtitle files that can be viewed or downloaded directly from the web interface. By automating the entire workflow, the proposed system is intended to reduce user effort and to support the accessibility of online video advertisements for deaf and hard-of-hearing audiences.

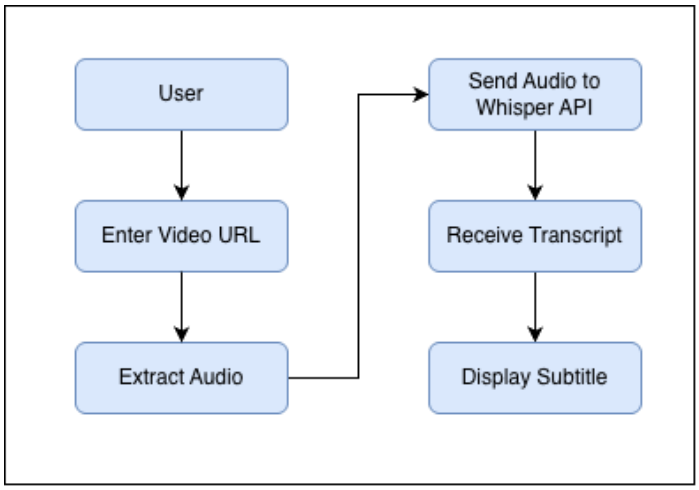

Figure 1. Automatic Subtitle Generation Workflow

Figure 1 illustrates the system design of the proposed AI-based automatic captioning application for online video advertisements. The process begins when the user provides the URL of an online video through the web interface. After receiving the input, the system automatically retrieves the video and extracts its audio stream, which serves as the input for the speech recognition process. Instead of requiring users to manually upload media files or perform transcription, the proposed design automates the entire preprocessing stage, thereby simplifying the subtitle generation workflow.

2.2 Automatic Captioning Process

The proposed automatic captioning workflow comprises a series of processing stages that convert an online advertisement video into time-stamped subtitle text by utilizing the OpenAI Whisper Automatic Speech Recognition (ASR) model. The process is initiated when a user enters the URL of an online video advertisement through the web-based interface. The backend application first verifies the validity of the submitted URL before retrieving the corresponding video from its online source. Once the video has been successfully obtained, the system extracts its audio stream to produce a speech-only audio file, which serves as the input for the subsequent speech recognition process. In this article, transcription refers to the conversion of speech into text, subtitle generation additionally includes arranging the text with timing information for display or export as an SRT file, and captioning is used as the general term for the overall function of the system.

The extracted audio is then forwarded to the OpenAI Whisper API, where the spoken content is automatically transcribed into text (Lyu et al., 2024). Along with the transcription, the API generates timestamp information for each recognized speech segment, which is used to align the subtitle text with the corresponding segments of the video. The backend system subsequently converts the transcription into a SubRip

Subtitle (SRT) file, which is returned to the web application for display. Users can preview the generated subtitles directly through the interface and download the resulting subtitle file for future use. By automating every stage of the subtitle generation pipeline, the proposed system is intended to reduce the need for manual transcription and to streamline subtitle production. The overall workflow of the proposed automatic captioning system, including URL validation, audio extraction, transcription, subtitle generation, and file download, is illustrated in the flowchart shown in Figure 2.

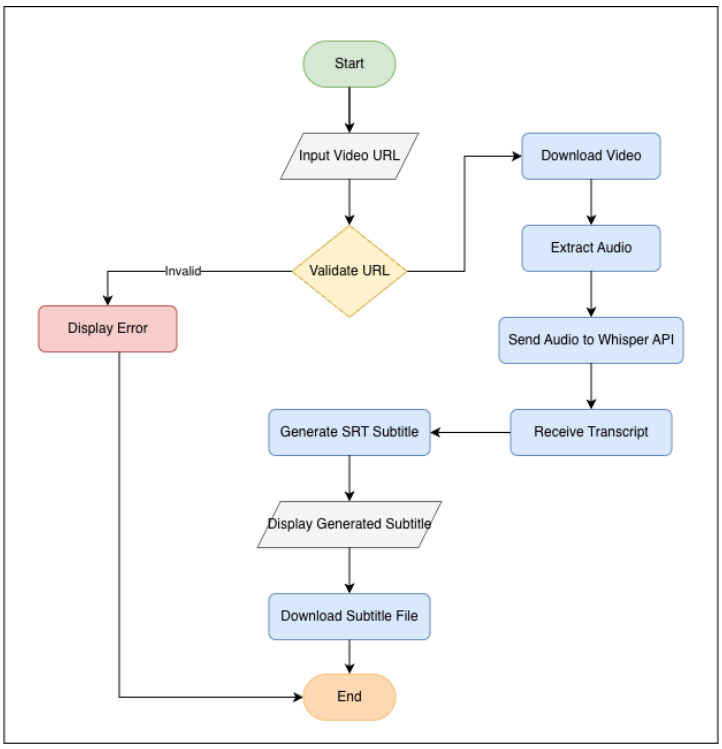

Figure 2. Flowchart of the Automatic Subtitle Generation Process

The proposed automatic captioning algorithm operates through a structured sequence of seven processing stages to generate subtitles from online advertisement videos. The workflow is initiated when the user provides the URL of a target video via the web-based interface. After validating the submitted URL, the backend application retrieves the corresponding video and downloads it for further processing. The system then extracts the audio track from the video, producing an audio file that serves as the input for the speech recognition module. This extracted audio is subsequently sent to the OpenAI Whisper API, where the spoken content is automatically converted into text along with timestamp information for each recognized speech segment (Thara et al., 2024).

Upon receiving the transcription output, the backend processes the generated text and transforms it into a SubRip Subtitle (SRT) file while maintaining the timestamp information to support synchronization with the original video. The completed subtitle file is then returned to the web application, where it is presented to the user for preview and can be downloaded for future use or integrated into the source video. By automating each stage of the subtitle generation process, the proposed algorithm streamlines the overall workflow, reduces manual effort, and provides a practical workflow for producing time-stamped subtitles from online video advertisements.

2.3 Functional Testing Procedure

Functional testing was conducted by the researchers using the black-box testing approach, in which the system is evaluated from the user's perspective without examining its internal implementation or source code. Five test scenarios were defined to cover the core functions of the subtitle generation workflow: video URL input, language selection, subtitle generation, subtitle display, and subtitle download. For each scenario, a valid input was provided through the web interface, the observed system response was recorded as the actual result, and it was compared with the expected result defined beforehand. A scenario was assigned a Pass status when the actual result matched the expected result and a Fail status otherwise. A Pass status therefore indicates only that a function ran and produced the expected output; it does not verify that the generated text or timestamps are correct. Because the tests were performed by the researchers rather than by end users, the testing verifies functional behavior only and does not constitute User Acceptance Testing (UAT). Transcription accuracy against reference transcripts, the quality of subtitle timestamps, and system performance under load were not evaluated in this study. The technology, test data, and test conditions are described in Section 2.4.

2.4 Test Environment and Test Data

The web-based AI captioning system was developed using Laravel/PHP and integrated the Whisper API for automatic speech recognition and subtitle generation from online video advertisements. Google Chrome was used to access and test the web application.

The system was functionally tested using 12 Indonesian-language video advertisements, each with a duration ranging from 30 to 60 seconds. The videos were obtained from YouTube and Facebook/Instagram. Following the black-box procedure described in Section 2.3, each of the five functional test scenarios was executed once on each of the 12 videos, giving 60 scenario executions in total, and the results were documented using screenshots of the testing process, which served as supporting evidence for the recorded test outcomes.

Some details were not recorded during development and testing and are therefore not reported rather than inferred: the allocation of frontend and backend responsibilities and any additional backend technology, the Whisper model or API version, the criteria used to select the videos, the audio characteristics of the videos (such as background noise, music, speaker accents, and speech rate), and the browser version, operating system, device specifications, and network conditions. This limits the reproducibility of the testing environment, and the audio characteristics cannot be used as a basis for comparative analysis. Because each scenario was executed only once on each video, the evaluation also does not establish the consistency of the system's behavior across repeated trials on the same video. In addition, this functional evaluation did not measure transcription accuracy, subtitle timing quality, or usability through direct testing with deaf or hard-of-hearing participants.

RESULTS AND DISCUSSION

3.1 Result

This section presents the implementation of the proposed AI-based automatic captioning system as a web application. The system was developed to provide a simple browser-based interface that enables users to generate subtitles automatically from online video advertisements. The implementation integrates a web-based interface with the OpenAI Whisper Automatic Speech Recognition (ASR) model through server-side processing, allowing subtitle generation to be performed with minimal user interaction.

Figure 3 illustrates the homepage of the developed application. The homepage serves as the primary interface where users interact with the automatic captioning system. The interface was intentionally designed with a clean and straightforward layout to simplify the subtitle generation process for users with different levels of technical expertise. The application requires only three main interactions before subtitle generation begins: entering the video URL, selecting the transcription language, and initiating the subtitle generation process.

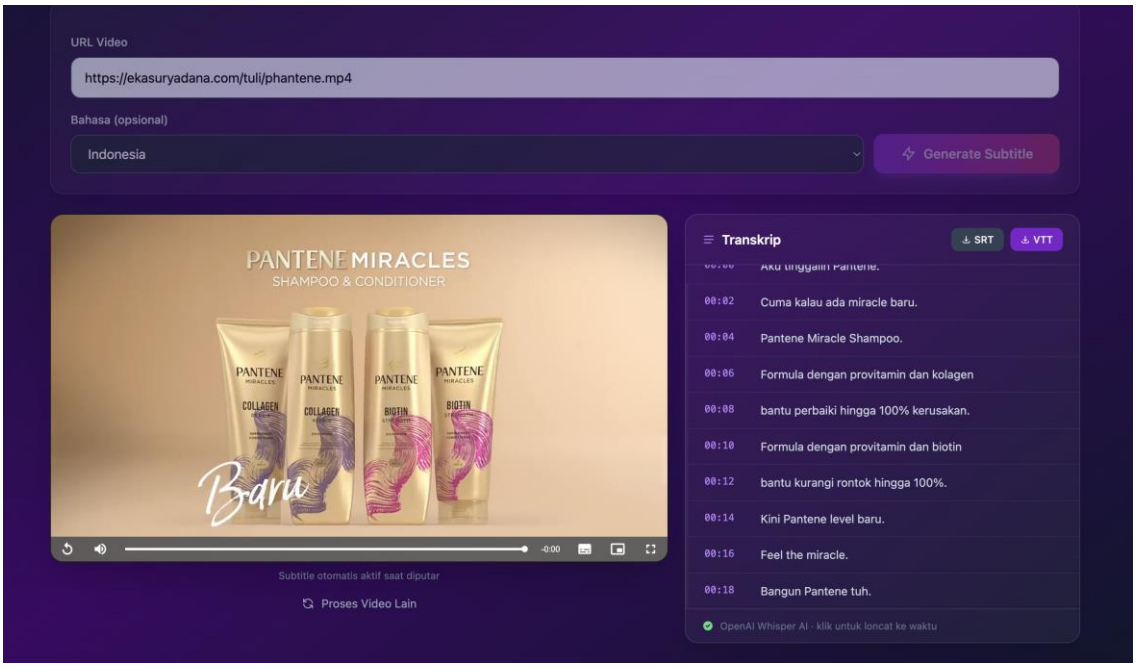

Figure 3. Homepage Interface of the Proposed AI-Based Automatic Captioning System

The implementation demonstrates that the proposed web-based application integrates the major components required for automatic subtitle generation. The interface minimizes user interaction by providing a simple workflow consisting of video URL input, language selection, subtitle generation, and result presentation. These features are supported by backend processes that include video retrieval, audio extraction, speech transcription, subtitle generation, and subtitle display.

The developed application was evaluated using functional black-box testing, conducted by the researchers according to the procedure and test conditions described in Sections 2.3 and 2.4. As presented in Table 2, which summarizes the outcomes of the 60 scenario executions (five scenarios on each of the 12 test videos), all five tested functions, including video URL input, language selection, subtitle generation, subtitle display, and subtitle download, operated as expected and received a Pass status, with all 60 scenario executions (12 of 12 for each scenario) passing and none failing. These findings indicate that the tested functions operated according to the predefined test scenarios. The results do not constitute User Acceptance Testing and do not, by themselves, demonstrate transcription accuracy or improved accessibility for deaf and hard-of-hearing users.

Table 2. Summary of Functional Black-box Testing Results (Five Scenarios on 12 Videos) of the Proposed Automatic Captioning System

Feature	Test Scenario	Expected Result	Actual Result	Status (Pass/Total)
Video URL Input	Enter a valid advertisement video URL	The system accepts the URL and proceeds to processing	URL successfully accepted and processed	Pass (12/12)
Language Selection	Select the transcription language	Selected language is applied during transcription	Language successfully selected	Pass (12/12)
Generate Subtitle	Click the Generate Subtitle button	The system automatically generates subtitles	Subtitle successfully generated using OpenAI Whisper	Pass (12/12)
Subtitle Display	Display transcription results	Generated subtitles are displayed on the web interface	Subtitle displayed correctly	Pass (12/12)
Download Subtitle	Download the generated subtitle	Subtitle file is downloaded in SRT format	Subtitle successfully downloaded	Pass (12/12)
Total	All five scenarios on 12 videos			Pass (60/60)

In each scenario, the actual result matched the expected result recorded in Table 2: the entered URL was accepted and processed, the selected language was applied, subtitles were generated using OpenAI Whisper, the generated subtitles were displayed on the web interface, and the subtitle file was downloaded. Accordingly, the results support the functional feasibility of the implemented workflow within the scope of the tests conducted. A Pass status for all five scenarios means that these scenarios met the predefined criteria; it does not mean that the system is free of errors. Because the tests were limited to these five scenarios, they do not provide evidence of transcription accuracy, subtitle timing quality, or the effect of the subtitles on deaf and hard-of-hearing users.

3.2 Discussion

The results demonstrate the functional feasibility of the proposed web-based captioning workflow: within the five scenarios tested, the integrated components, including the OpenAI Whisper ASR service (Lyu et al., 2024), operated together as

designed. They do not show that the generated transcriptions are accurate or that the timestamps are well aligned with the speech, because transcription accuracy (for example, Word Error Rate) and timing quality were not measured. A Pass status therefore indicates functional behavior only and does not imply that the system is free of errors.

As summarized in Table 1, the closest studies examined also use speech recognition for subtitle-related tasks: Thara et al. (2024) apply Whisper to subtitle synchronization, and Striuk and Hordiienko (2025) describe a subtitle management system that produces SRT or WebVTT files from uploaded videos. The present study differs in its focus on online video advertisements and in retrieving the video from a URL within a browser-based Whisper workflow that returns a displayable and downloadable SRT file. With regard to evaluation, Fathallah et al. (2024) reported Word Error Rate for LLM-corrected captions, whereas the present study reports functional testing only; its results therefore cannot yet be compared quantitatively with studies that report recognition error rates (Rai et al., 2023; Fathallah et al., 2024) or subtitle synchronization quality (Mocanu & Tapu, 2021; Thara et al., 2024). The contribution of this study is thus one of implementation and functional evaluation rather than of improved recognition performance.

The system is designed to support deaf and hard-of-hearing users by reducing the effort needed to produce subtitles for online video advertisements, but its accessibility value remains a potential rather than a demonstrated outcome. Evaluating subtitle quality requires human-verified reference transcripts, from which Word Error Rate and, where appropriate, Character Error Rate can be computed, together with an assessment of subtitle timing. Evaluating accessibility requires direct assessment with deaf and hard-of-hearing users, for example of readability, comprehension of the advertised information, and usability. Factors such as brand names, background music, accents, and rapid speech could potentially affect speech recognition in advertisement videos; this study did not test their influence, and they are proposed as variables to be examined in future evaluations.

From a practical perspective, the browser-based implementation allows users to generate subtitles without installing additional software. The design models describe the system as interactions among the user, web application, backend server, and Whisper API (Section 2.1), but the allocation of frontend and backend responsibilities in the implementation was not documented (Section 2.4). Claims regarding modularity, maintenance effort, scalability, processing performance, or ease of integrating other speech recognition technologies therefore cannot be made from the present results and would require dedicated documentation and testing.

This study has several limitations. First, testing was conducted by the researchers, and no evaluation with deaf and hard-of-hearing users was carried out. Second,

transcription accuracy was not measured against reference transcripts, and the study does not report a systematic analysis of how language or audio conditions affect the results; the multilingual capability attributed to Whisper (Saraf et al., 2025) was likewise not assessed across languages. Third, the quality of subtitle timestamps and the performance of the system were not evaluated, and each scenario was executed only once on each of 12 Indonesian-language advertisements of 30 to 60 seconds, so consistency across repeated trials, videos of other durations, and other languages was not established. Fourth, the application depends on an external API whose availability and behavior are outside the control of the developers, and potential problems in accessing some online video sources were not systematically examined. Future work should compute Word Error Rate and Character Error Rate against human-verified reference transcripts for a documented set of advertisement videos, evaluate timestamp accuracy using a predefined tolerance, and conduct a user evaluation with deaf and hard-of-hearing participants, with appropriate ethical approval and informed consent, to assess usability, readability, and comprehension.

CONCLUSION

This study developed a web-based automatic captioning application for online video advertisements by integrating the OpenAI Whisper ASR service with a browser-based workflow that retrieves a video from its URL, extracts its audio, transcribes the speech, and produces a downloadable SRT file. Functional black-box testing showed that all five tested core functions operated according to their predefined test scenarios, which demonstrates the functional feasibility of the implemented application within the scope of the tests conducted. The study does not yet establish transcription accuracy, the quality of subtitle timestamps, or the effectiveness of the generated subtitles for deaf and hard-of-hearing users; the system is designed to support accessibility, but this remains a potential rather than a demonstrated outcome. Future work should evaluate subtitle quality against human-verified reference transcripts, assess timestamp accuracy and system performance, and examine usability and accessibility through direct evaluation with deaf and hard-of-hearing users.

Funding Statement

This research was funded by the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia through the BIMA 2026 research funding program under the Regular Fundamental Research (Penelitian Fundamental Reguler) scheme. The funding supported the implementation of the research and the preparation of this scientific publication. The authors gratefully acknowledge the financial support provided through the BIMA 2026 program.

Declaration of AI Usage

The authors used artificial intelligence (AI) tools, including ChatGPT, to assist with paraphrasing sentences and improving the clarity and readability of the manuscript. AI tools were also used to assist in identifying and refining an appropriate and relevant title for the study. In addition, AI tools were used to review grammar, spelling, academic writing style, and manuscript formatting. AI tools were also used to assist in developing and reviewing program code used in the research; however, all AI-assisted outputs were critically reviewed, modified where necessary, and verified by the authors, who retain full responsibility for the accuracy, originality, validity, and integrity of the manuscript and its results.

Conflict of Interest

The authors declare that there are no financial, institutional, professional, or personal conflicts of interest that could have influenced the conduct of this research. The authors further declare that there is no conflict of interest related to the preparation and publication of this article.

REFERENCES

Breuer, A., Dietrich, B. J., Crespin, M. H., Butler, M., Pryse, J., & Imai, K. (2025). Using AI to summarize US presidential campaign TV advertisement videos, 1952–2012. Scientific Data, 12(1), 1552. https://doi.org/10.1038/s41597-025-05558-9

Fathallah, N., Bhole, M., & Staab, S. (2024). Empowering the Deaf and Hard of Hearing community: Improving video captions with large language models. Proceedings of the 11th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion, 345–353. https://doi.org/10.1145/3696593.3696617

Hasini, Lokesh, K., P, A., Venkataramana, A., Kumar, M. M., & Kumar, D. P. (2025). Speech recognition using deep learning techniques: A comparative study. 2025 Innovations in Power and Advanced Computing Technologies (i-PACT), 1–8. https://doi.org/10.1109/i-PACT65952.2025.11307948

Hatami, M., & Chegini, M. (2024). Enhancing digital content accessibility for the hearing impaired through AI-driven visual representations. 2024 10th International Conference on Artificial Intelligence and Robotics (QICAR), 322–328. https://doi.org/10.1109/QICAR61538.2024.10496621

Karad, S., Lahane, A., Dhamdhere, P., Thosar, D., Mane, S., & Kshirsagar, S. (2024). Generation of video subtitles for all video creators. 2024 2nd DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAIEI), 1–5. https://doi.org/10.1109/IDICAIEI61867.2024.10842907

Leotta, M., Mori, F., & Ribaudo, M. (2023). Evaluating the effectiveness of automatic image captioning for web accessibility. Universal Access in the Information Society, 22(4), 1293–1313. https://doi.org/10.1007/s10209-022-00906-7

Lyu, K.-M., Lyu, R., & Chang, H.-T. (2024). Real-time multilingual speech recognition and speaker diarization system based on Whisper segmentation. PeerJ Computer Science, 10, e1973. https://doi.org/10.7717/peerj-cs.1973

May, L., Ohshiro, K., Dang, K., Sridhar, S., Pai, J., Fuentes, M., Lee, S., & Cartwright, M. (2024). Unspoken sound: Identifying trends in non-speech audio captioning on YouTube. Proceedings of the CHI Conference on Human Factors in Computing Systems, 1–19. https://doi.org/10.1145/3613904.3642162

Mocanu, B., & Tapu, R. (2021). Automatic subtitle synchronization and positioning system dedicated to deaf and hearing impaired people. IEEE Access, 9, 139544–139555. https://doi.org/10.1109/ACCESS.2021.3119201

Rai, A. K., Jaiswal, S. D., & Mukherjee, A. (2023). A deep dive into the disparity of word error rates across thousands of NPTEL MOOC videos. arXiv. https://arxiv.org/abs/2307.10587

Saraf, A., Bhardwaj, R., Bhumika, & Thakur, R. (2025). Multilingual translation for speech and text using Whisper AI: A deep learning approach. International Journal for Research in Applied Science and Engineering Technology, 13(7), 1895–1901. https://doi.org/10.22214/ijraset.2025.73288

Striuk, A. M., & Hordiienko, V. V. (2025). Research and development of a subtitle management system using artificial intelligence. CEUR Workshop Proceedings, 3917, 415–427. https://ceur-ws.org/Vol-3917/paper61.pdf

Thara, P., Azneed, M., Sanas, M., Jithu Prakash, P. M., Naik, H. P., & Divya, B. (2024). Subtitle synchronization using Whisper ASR model. 2024 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), 1–6. https://doi.org/10.1109/ICPECTS62210.2024.10780268