

# Research and development of a subtitle management system using artificial intelligence

Andrii M. Striuk<sup>1,2,3</sup>, Vladyslav V. Hordiienko<sup>1</sup>

<sup>1</sup>Kryvyi Rih National University, 11 Vitalii Matusevych Str., Kryvyi Rih, 50027, Ukraine

<sup>2</sup>Kryvyi Rih State Pedagogical University, 54 Universytetskyi Ave., Kryvyi Rih, 50086, Ukraine

<sup>3</sup>Academy of Cognitive and Natural Sciences, 54 Universytetskyi Ave., Kryvyi Rih, 50086, Ukraine

## Abstract

Subtitles play a vital role in making video content accessible to a wider audience, including individuals with hearing impairments and those who do not understand the spoken language. However, the manual creation of subtitles is a time-consuming and labor-intensive process. This paper proposes an AI-powered subtitle management system that automates the generation and management of subtitles for video content. The system leverages state-of-the-art automatic speech recognition (ASR) and machine translation (MT) technologies to generate accurate and synchronized subtitles in multiple languages. The proposed system architecture consists of a speech recognition module, a machine translation module, a subtitle segmentation and formatting module, and a user-friendly interface. The paper provides a comprehensive literature review of the related work in the field of AI-based subtitle generation, covering key aspects such as speech recognition techniques, machine translation approaches, multimodal methods, and evaluation methodologies. The implications of the proposed system for subtitle generation pipelines are discussed, highlighting its potential to enhance efficiency, scalability, and accessibility. The limitations of the current system and directions for future research are also outlined. This research contributes to the advancement of AI-powered subtitle generation and aims to make video content more inclusive and accessible to a global audience.

## Keywords

subtitles, artificial intelligence, speech recognition, machine translation, video accessibility

## 1. Introduction

### 1.1. Background and motivation

In today's digital age, video content has become an integral part of communication, education, and entertainment. However, the accessibility of video content remains a challenge for individuals with hearing impairments or those who do not understand the spoken language. Subtitles play a crucial role in making video content more inclusive and accessible to a wider audience [1].

Despite the importance of subtitles, the process of manually creating them is time-consuming, labor-intensive, and prone to errors [2]. It requires skilled human translators to listen to the audio, transcribe the dialogue, and synchronize the subtitles with the video timestamps. This manual process often results in delays in the availability of subtitles and limits the scalability of subtitle generation for large volumes of video content.

Advancements in artificial intelligence (AI) technologies, particularly in the fields of automatic speech recognition (ASR) and machine translation (MT), have opened up new possibilities for automating the subtitle generation process [3, 4]. AI-powered systems can significantly reduce the time and effort required for subtitle creation while maintaining high levels of accuracy and quality.

*CS&SE@SW 2024: 7th Workshop for Young Scientists in Computer Science & Software Engineering, December 27, 2024, Kryvyi Rih, Ukraine*

✉ andrey.n.stryuk@gmail.com (A. M. Striuk)

🌐 <http://mpz.knu.edu.ua/andrij-stryuk/> (A. M. Striuk)

🆔 0000-0001-9240-1976 (A. M. Striuk)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

## 1.2. Research objectives

The primary objective of this research is to develop an AI-powered subtitle management system that automates the generation and management of subtitles for video content. The proposed system aims to leverage state-of-the-art ASR and MT technologies to generate accurate and synchronized subtitles in multiple languages.

Furthermore, this research aims to provide a comprehensive literature review of the existing techniques, applications, and evaluation methodologies in the field of AI-based subtitle generation. By synthesizing the current state of knowledge, we aim to identify the challenges, opportunities, and future research directions in this domain.

## 1.3. Paper contributions and organization

The main contributions of this paper are as follows:

- We propose an AI-powered subtitle management system that automates the generation and management of subtitles for video content, leveraging state-of-the-art ASR and MT technologies.
- We provide an extensive literature review of the related work in the field of AI-based subtitle generation, covering key aspects such as speech recognition, machine translation, multimodal approaches, and evaluation methodologies.
- We present the architecture and key components of the proposed subtitle management system, including the speech recognition module, machine translation module, subtitle segmentation and formatting module, and user interface.

The remainder of this paper is organized as follows: section 2 provides a comprehensive overview of the related work in the field of AI-based subtitle generation. Section 3 describes the proposed AI-powered subtitle management system, including its architecture, components, and functionalities. Finally, section 4 concludes the paper and summarizes the key findings and contributions.

## 2. Related work

Extensive research has been conducted in the field of AI-based subtitle generation, spanning various techniques, applications, and evaluation methodologies. This section provides a comprehensive overview of the related work, focusing on key aspects such as speech recognition, machine translation, multimodal approaches, and subtitle evaluation metrics.

### 2.1. Speech recognition for subtitle generation

Automatic speech recognition (ASR) plays a crucial role in the subtitle generation pipeline by converting spoken audio into textual transcripts. Researchers have explored various ASR techniques to improve the accuracy and efficiency of subtitle generation. Radha and Pradeep [5] proposed an automated subtitle generation system using hidden Markov models (HMMs) for speech recognition. They demonstrated the effectiveness of their approach on English-language videos and highlighted the importance of accurate speech recognition for subtitle quality.

Convolutional neural networks (CNNs) [6] have also been employed for ASR in subtitle generation tasks. Ramani et al. [7] developed an automatic subtitle generation system using CNNs for speech recognition, achieving promising results on real-time video subtitling. They emphasized the significance of audio preprocessing techniques and the choice of media player for seamless subtitle integration.

Recurrent neural networks (RNNs), particularly long short-term memory (LSTM) networks, have gained popularity in ASR for subtitle generation. Kiran et al. [8] proposed a subtitle generation system using sequence-to-sequence RNNs for speech recognition and video scene indexing. Their approach demonstrated improved accuracy and the ability to handle longer video sequences compared to traditional methods.

The application of ASR techniques to specific domains, such as lecture videos, has also been explored. Che et al. [9] developed an automatic lecture subtitle generation system using ASR and evaluated its performance against manual subtitling. They found that the ASR-generated subtitles significantly reduced the time and effort required for subtitle creation while maintaining comparable quality. Similarly, Sridhar et al. [10] proposed a hybrid approach combining acoustic and linguistic features for subtitle generation in computer science lecture videos, achieving improved accuracy in detecting discourse boundaries.

## 2.2. Machine translation for multilingual subtitles

Machine translation (MT) is essential for generating subtitles in multiple languages, enabling video content to reach a wider global audience. Researchers have investigated various MT approaches, including statistical and neural models, to improve the quality and efficiency of multilingual subtitle generation.

Karakanta et al. [11] conducted a comparative study of different MT approaches for subtitle generation, including phrase-based statistical MT and neural MT. They evaluated the performance of these approaches on a multilingual subtitle corpus and highlighted the challenges in preserving linguistic and cultural nuances in translated subtitles.

Neural MT architectures, such as sequence-to-sequence models with attention mechanisms, have shown promising results in subtitle translation tasks. Du and Lu [4] proposed a neural MT system specifically designed for subtitle translation, incorporating features such as character-level encoding and domain adaptation. Their system achieved significant improvements in translation quality compared to traditional MT approaches.

The quality of AI-generated subtitles compared to human translations has also been a focus of research. Calvo-Ferrer [12] conducted a study comparing the quality of subtitles generated by machine translation systems with those created by human translators. They found that while MT systems have made significant progress, human translators still outperform them in terms of accuracy and contextual understanding.

## 2.3. Multimodal and end-to-end subtitle generation

Multimodal approaches that leverage both visual and linguistic information have emerged as promising directions for subtitle generation. These approaches aim to capture the contextual and visual cues present in the video to enhance the accuracy and coherence of the generated subtitles.

Shanmugam et al. [13] proposed a multimodal subtitle generation system that combines visual features extracted from the video frames with linguistic information from the audio transcripts. Their approach demonstrated improved synchronization and contextual relevance of the generated subtitles compared to unimodal methods.

Martín et al. [14] developed a multimodal subtitle generation framework that incorporates visual, acoustic, and linguistic features using deep learning techniques. They evaluated their system on a dataset of educational videos and showed significant improvements in subtitle quality and alignment.

End-to-end subtitle generation, where the entire process from speech recognition to subtitle generation is performed by a single model, has also gained attention. Valor Miró et al. [15] proposed an end-to-end subtitle generation system that directly translates speech into subtitles in multiple languages. Their approach achieved comparable performance to pipeline-based methods while reducing the complexity and error propagation.

Hotta et al. [16] developed an end-to-end speech-to-text translation system specifically designed for subtitle generation. Their system incorporated techniques such as attention mechanisms and beam search to improve the quality and fluency of the generated subtitles.

## 2.4. Subtitle evaluation metrics and methodologies

Evaluating the quality and effectiveness of AI-generated subtitles is crucial for assessing their usability and acceptability. Researchers have proposed various evaluation metrics and methodologies to measure the performance of subtitle generation systems.

Automatic evaluation metrics, such as word error rate (WER) and bilingual evaluation understudy (BLEU), have been widely used to assess the accuracy and fluency of generated subtitles. Ramani et al. [7] employed WER as a metric to evaluate the performance of their CNN-based subtitle generation system, demonstrating its effectiveness in measuring transcript accuracy.

Kaulage et al. [17] utilized the BLEU score to evaluate the quality of machine-translated subtitles in their multilingual subtitle generation system. They highlighted the importance of considering both the accuracy and fluency of the translations when assessing subtitle quality.

Human evaluation methodologies have also been employed to assess the subjective quality and user experience of AI-generated subtitles. Al Sawi and Allam [18] conducted a comparative analysis of human-generated and AI-generated Arabic subtitles, using qualitative and quantitative approaches to evaluate the subtitle quality and viewer comprehension.

Kuroiwa et al. [19] proposed a human-in-the-loop approach for subtitle generation, combining AI techniques with human intervention to improve the accuracy and cultural appropriateness of the generated subtitles. They emphasized the importance of human expertise in overcoming the limitations of AI systems in understanding cultural nuances.

## 2.5. Applications in education and entertainment

AI-based subtitle generation has found significant applications in various domains, particularly in education and entertainment. Researchers have explored the benefits and challenges of deploying AI-powered subtitle systems in these contexts.

In the educational domain, Qiu [20] proposed an automatic subtitle generation system for teaching videos using cloud computing techniques. They demonstrated the effectiveness of their approach in reducing the time and effort required for subtitle creation, thereby enhancing the accessibility of educational content.

Martín et al. [14] developed an automatic subtitle generation system specifically tailored for educational videos produced by the Government of La Rioja, Spain. Their system aimed to improve the accessibility of important educational content for individuals with hearing impairments.

In the entertainment industry, Malakul and Park [1] investigated the effects of using an auto-subtitle system in educational videos on learning comprehension, cognitive load, and satisfaction. They found that AI-generated subtitles significantly improved the learning experience for non-native speakers and individuals with hearing impairments.

Kuroiwa et al. [19] explored the challenges and opportunities of AI-based subtitle generation for anime content. They proposed a hybrid approach combining AI techniques with human intervention to improve the accuracy and cultural appropriateness of the generated subtitles, highlighting the importance of human expertise in overcoming linguistic and cultural barriers.

The related work discussed in this section highlights the diverse techniques, applications, and evaluation methodologies in the field of AI-based subtitle generation. Researchers have made significant strides in developing effective speech recognition, machine translation, and multimodal approaches for subtitle generation. However, challenges remain in terms of improving the accuracy, fluency, and contextual understanding of AI-generated subtitles, particularly in handling linguistic and cultural nuances. The evaluation of subtitle quality using both automatic metrics and human assessment is crucial for ensuring the usability and acceptability of AI-generated subtitles in real-world applications.

As the demand for accessible and multilingual video content continues to grow, AI-based subtitle generation systems are expected to play an increasingly important role in facilitating the creation and dissemination of subtitles. Future research directions include the development of more advanced and integrated AI techniques, the incorporation of domain-specific knowledge, and the exploration of

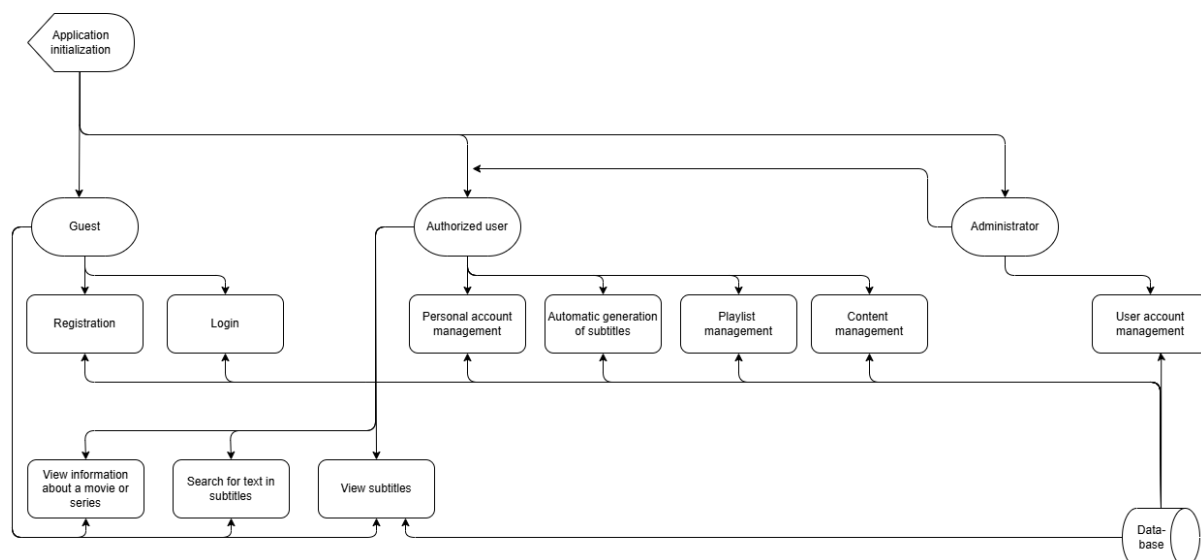
user-centric evaluation methodologies to ensure the effectiveness and user satisfaction of AI-generated subtitles.

### 3. Proposed subtitle management system

This section presents the proposed AI-powered subtitle management system, which aims to automate the generation and management of subtitles for video content. The system leverages state-of-the-art speech recognition and machine translation technologies to generate accurate and synchronized subtitles in multiple languages. The proposed system architecture, key components, and functionalities are described in detail.

#### 3.1. System architecture overview

The proposed subtitle management system follows a modular architecture, consisting of several interconnected components that work together to achieve automated subtitle generation and management. Figure 1 provides an overview of the system architecture, highlighting the main modules and their interactions.



**Figure 1:** Proposed subtitle management system architecture.

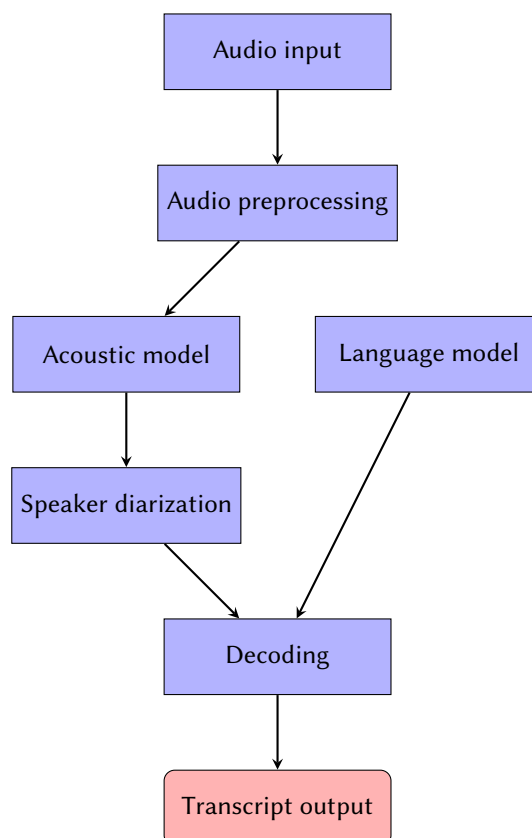
The system architecture consists of the following main components:

- Speech recognition module is responsible for converting the audio content of the video into textual transcripts. It employs advanced acoustic and language models to accurately recognize speech and generate time-aligned transcripts.
- Machine translation module takes the transcripts generated by the speech recognition module and translates them into the desired target languages. It utilizes state-of-the-art neural machine translation techniques to produce high-quality translations while preserving the context and meaning of the original content.
- Subtitle segmentation and formatting module handles the segmentation of the translated text into appropriate subtitle blocks and applies proper formatting and styling to ensure readability and compliance with subtitle standards.
- The system includes a user-friendly interface that allows users to upload videos, select target languages, and manage generated subtitles.
- The database stores video metadata, transcripts, translations, and subtitle files for efficient retrieval and management.

The modular architecture of the proposed system enables flexibility, scalability, and ease of maintenance. Each component can be independently developed, tested, and updated, allowing for continuous improvement and adaptation to advancements in speech recognition and machine translation technologies.

### 3.2. Speech recognition module

The speech recognition module plays a crucial role in the subtitle generation pipeline by accurately converting the spoken audio content into textual transcripts. Figure 2 illustrates the workflow of the speech recognition module.



**Figure 2:** Speech recognition module workflow.

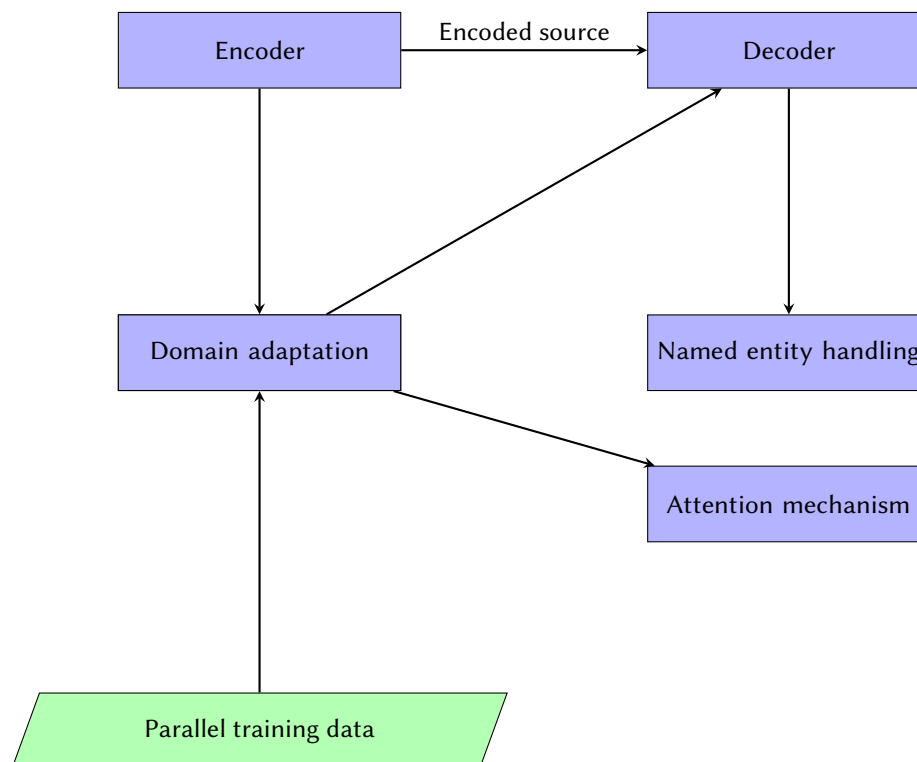
The speech recognition module incorporates the following key components and techniques:

- The acoustic model is trained on a large dataset of speech samples and their corresponding transcriptions. It learns the relationship between audio features and phonemes, enabling it to recognize speech patterns and map them to textual representations.
- The language model captures the statistical properties of the target language, including word sequences and grammar. It helps in improving the accuracy of speech recognition by providing contextual information and constraining the search space of possible transcriptions.
- Speaker diarization is the process of segmenting the audio stream into speaker-specific segments. It allows the system to identify and differentiate between multiple speakers in the video, enabling accurate attribution of subtitles to the corresponding speakers.
- Before feeding the audio content into the speech recognition module, various preprocessing techniques are applied to enhance the quality and remove noise. These techniques include audio normalization, noise reduction, and speaker adaptation.

The speech recognition module employs state-of-the-art deep learning architectures, such as CNN and RNN, to achieve high accuracy in transcribing speech. The module is trained on diverse speech datasets, including various accents, dialects, and languages, to ensure robustness and generalization capabilities.

### 3.3. Machine translation module

The machine translation module is responsible for translating the transcripts generated by the speech recognition module into the desired target languages. Figure 3 depicts the architecture of the machine translation module.



**Figure 3:** Machine translation module architecture.

The machine translation module utilizes an encoder-decoder architecture, which has become the de facto standard in neural machine translation. The key components of the machine translation module are as follows:

- The encoder takes the source language transcript as input and converts it into a fixed-length vector representation. It employs techniques such as word embeddings and recurrent neural networks to capture the semantic and syntactic information of the input sequence.
- The decoder takes the encoded representation produced by the encoder and generates the target language translation. It uses attention mechanisms to selectively focus on relevant parts of the input sequence during the decoding process, enabling the generation of accurate and fluent translations.
- The machine translation module incorporates techniques to handle out-of-vocabulary words and named entities. This includes subword tokenization, which breaks down rare words into smaller units, and named entity recognition, which identifies and preserves named entities during the translation process.
- To improve translation quality for specific domains, such as educational or entertainment content, the machine translation module can be fine-tuned on domain-specific parallel corpora. This

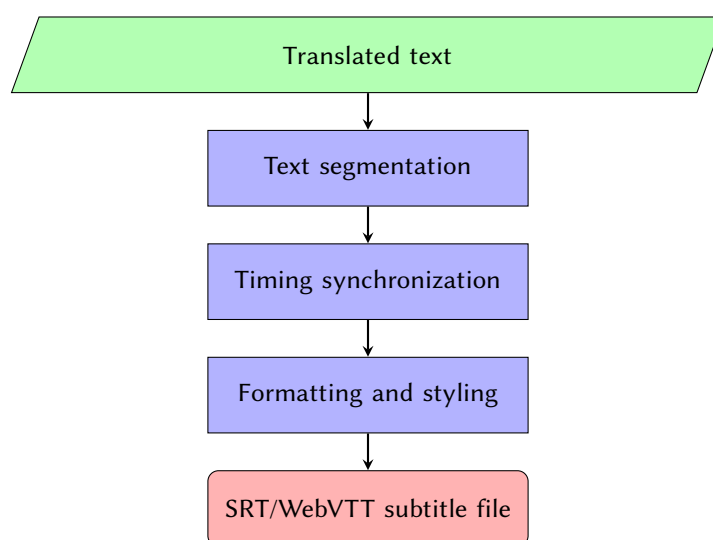


allows the module to learn domain-specific terminology and style, resulting in more accurate and contextually relevant translations.

The machine translation module is trained on large-scale parallel corpora, consisting of sentence pairs in the source and target languages. Advanced training techniques, such as teacher forcing and back-translation, are employed to improve the quality and fluency of the generated translations.

### 3.4. Subtitle segmentation and formatting

The subtitle segmentation and formatting module takes the translated text and performs necessary segmentation and formatting to generate properly structured subtitle files. Figure 4 illustrates the process of subtitle segmentation and formatting.



**Figure 4:** Subtitle segmentation and formatting process.

The subtitle segmentation and formatting module incorporates the following key steps:

1. The translated text is segmented into appropriate subtitle blocks based on factors such as sentence boundaries, dialogue turns, and reading speed. The segmentation ensures that each subtitle block is concise, readable, and synchronized with the audio.
2. Timing synchronization module aligns the segmented subtitle blocks with the corresponding timestamps in the video. It takes into account the start and end times of each subtitle block, ensuring that the subtitles appear at the appropriate moments and remain synchronized with the audio.
3. Formatting and styling module applies proper formatting and styling to the subtitle text, following established subtitle standards and guidelines. This includes setting font properties, such as size and color, and applying text formatting, such as italics or bold, to emphasize specific words or phrases.
4. The segmented and formatted subtitle blocks are combined to generate standard subtitle file formats, such as SubRip Text (SRT) or Web Video Text Tracks (WebVTT). These subtitle files can be easily integrated with video players and streaming platforms.

The subtitle segmentation and formatting module ensures that the generated subtitles adhere to industry standards and best practices, enhancing the readability and usability of the subtitles for viewers.



### 3.5. User interface and interaction design

The proposed subtitle management system includes a user-friendly interface that allows users to seamlessly interact with the system and manage the subtitle generation process. Figure 5 presents a high-level overview of the user interface and interaction design.

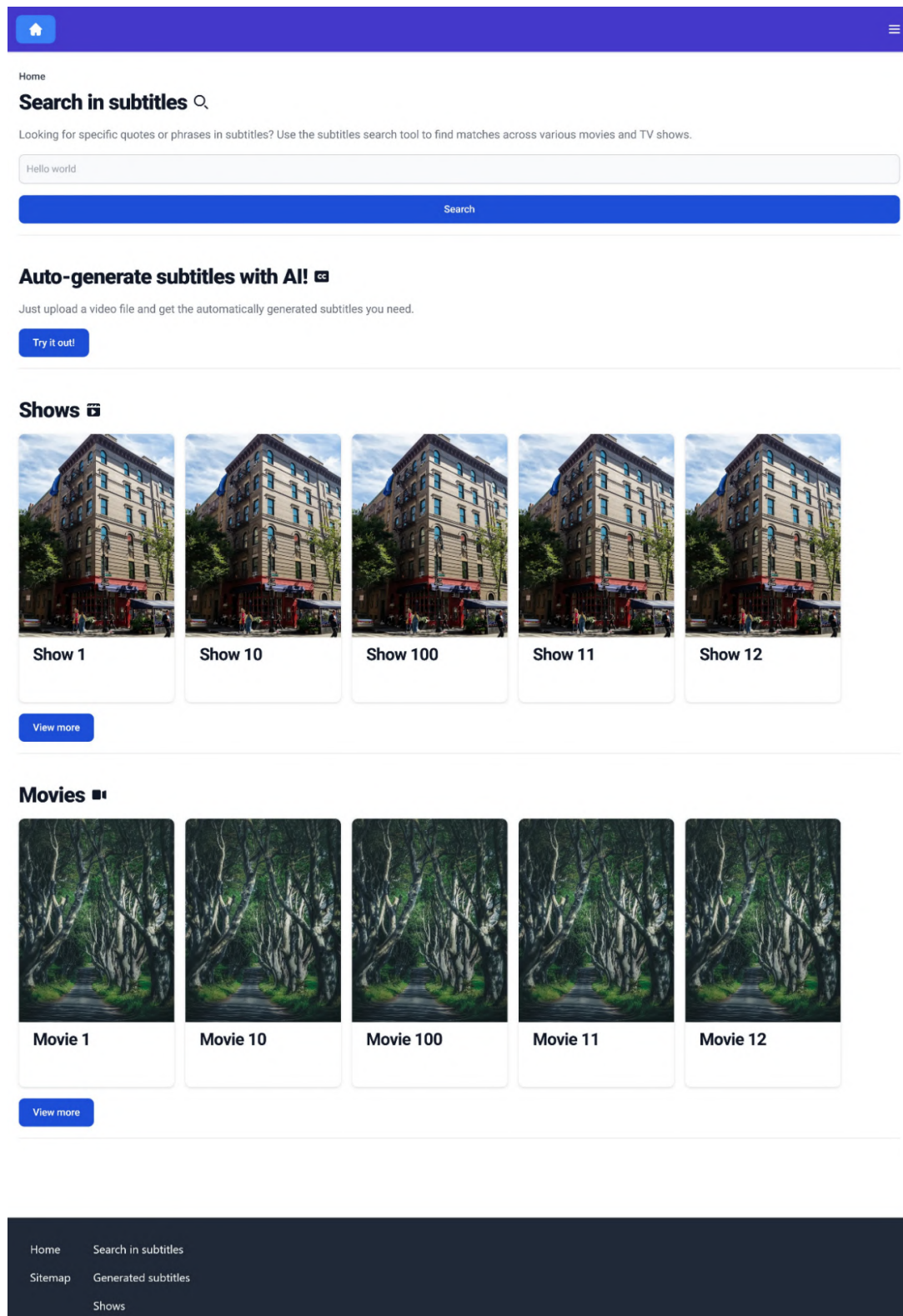


Figure 5: User interface and interaction design.

The user interface incorporates the following key features and functionalities:

- *Video upload*: users can easily upload their video files to the system through a simple and intuitive interface. The system supports various video formats and provides options for selecting the desired target languages for subtitle generation.
- *Language selection*: the interface allows users to choose the target languages for subtitle generation. Users can select multiple languages simultaneously, enabling the creation of multilingual subtitles for their videos.
- *Subtitle preview and editing*: the system provides a subtitle preview feature that allows users to review the generated subtitles alongside the video. Users can make necessary edits and adjustments to the subtitles, ensuring their accuracy and synchronization with the video content.
- *Subtitle download and integration*: once the subtitles are generated and reviewed, users can easily download the subtitle files in standard formats. The interface provides instructions and guides on how to integrate the subtitle files with popular video players and platforms.
- *Subtitle management*: the system offers a centralized subtitle management feature, allowing users to organize, search, and manage their generated subtitles. Users can view their subtitle history, update existing subtitles, and delete unwanted subtitle files.

The user interface is designed with usability and accessibility in mind, ensuring that users with varying technical backgrounds can easily navigate and utilize the subtitle management system. The interface incorporates responsive design principles, enabling access from different devices, including desktops, laptops, tablets, and smartphones.

## 4. Conclusion and future work

### 4.1. Summary of findings

This research presented an AI-powered subtitle management system that automates the generation and management of subtitles for video content. The proposed system leveraged state-of-the-art speech recognition and machine translation technologies to generate accurate and synchronized subtitles in multiple languages.

The system architecture was designed to be modular, scalable, and adaptable to advancements in AI technologies. It consisted of key components such as the speech recognition module, machine translation module, subtitle segmentation and formatting module, and user interface.

The speech recognition module utilized advanced acoustic and language models, along with techniques like speaker diarization and audio preprocessing, to accurately convert spoken audio into textual transcripts. The machine translation module employed an encoder-decoder architecture with attention mechanisms to translate the transcripts into desired target languages while preserving context and meaning.

The subtitle segmentation and formatting module ensured that the translated text was properly segmented, synchronized, and formatted according to subtitle standards and guidelines. The user interface provided a user-friendly and intuitive platform for users to upload videos, select target languages, preview and edit subtitles, and manage their subtitle files.

### 4.2. Implications for subtitle generation pipelines

The proposed AI-powered subtitle management system has significant implications for the efficiency and scalability of subtitle generation pipelines. By automating the process of speech recognition, translation, and subtitle formatting, the system can greatly reduce the time and effort required for manual subtitle creation.

The modular architecture of the system allows for easy integration with existing video platforms and workflows. It enables content creators, educational institutions, and entertainment providers to generate high-quality subtitles for their video content quickly and cost-effectively.

The system's ability to generate subtitles in multiple languages opens up new opportunities for content localization and global accessibility. It facilitates the dissemination of educational and entertainment content to a wider audience, breaking down language barriers and promoting inclusivity.

### 4.3. Limitations and directions for further research

While the proposed subtitle management system demonstrates promising results, there are certain limitations and areas for further research:

- *Language coverage*: the current system focuses on a limited set of languages for subtitle generation. Expanding the language coverage to include more diverse and low-resource languages would enhance the system's applicability and reach.
- *Domain adaptation*: the performance of the speech recognition and machine translation modules can be further improved by fine-tuning them on domain-specific datasets. Investigating techniques for domain adaptation, such as transfer learning and unsupervised adaptation, would enhance the system's effectiveness in various domains like education, entertainment, and specialized fields.
- *Contextual understanding*: although the system incorporates techniques to handle named entities and preserve context during translation, there is room for improvement in capturing and conveying subtle nuances, idiomatic expressions, and cultural references. Exploring advanced natural language processing techniques, such as contextual embeddings and knowledge graphs, could enhance the system's ability to generate more contextually accurate and culturally appropriate subtitles.
- Incorporating *user feedback and interaction* mechanisms into the system could greatly improve its usability and adaptability. Allowing users to provide feedback on generated subtitles, suggest corrections, and contribute to the system's learning process would lead to continuous improvement in subtitle quality and user satisfaction.
- Exploring the *integration of visual and acoustic cues from the video content*, such as scene changes, speaker identification, and emotion recognition, could further enhance the accuracy and synchronization of the generated subtitles.

Future research directions could focus on addressing these limitations and expanding the capabilities of the AI-powered subtitle management system. Collaborations between researchers, language experts, and industry stakeholders would be crucial in driving innovation and advancing the state of the art in automated subtitle generation.

**Declaration on Generative AI:** During the preparation of this work, the authors used Claude 3 Opus in order to: Drafting content, Generate literature review. After using this service, the authors reviewed and edited the content as needed and takes full responsibility for the publication's content.

## References

- [1] S. Malakul, I. Park, The effects of using an auto-subtitle system in educational videos to facilitate learning for secondary school students: learning comprehension, cognitive load, and satisfaction, *Smart Learning Environments* 10 (2023) 4. doi:10.1186/s40561-023-00224-2.
- [2] A. Mathur, T. Saxena, R. Krishnamurthi, Generating subtitles automatically using audio extraction and speech recognition, in: *Proceedings - 2015 IEEE International Conference on Computational Intelligence and Communication Technology, CICT 2015*, Institute of Electrical and Electronics Engineers Inc., 2015, pp. 621–626. doi:10.1109/CICT.2015.46.
- [3] V. B. Aswin, M. Javed, P. Parihar, K. Aswanth, C. R. Druval, A. Dagar, C. V. Aravinda, NLP-Driven Ensemble-Based Automatic Subtitle Generation and Semantic Video Summarization Technique, in: N. N. Chiplunkar, T. Fukao (Eds.), *Advances in Artificial Intelligence and Data Engineering*, volume 1133 of *Advances in Intelligent Systems and Computing*, Springer Nature Singapore, Singapore, 2021, pp. 3–13. doi:10.1007/978-981-15-3514-7\_1.

- [4] J. Du, J. Lu, A Comparative Study on the Translation Quality between Human and Machine-Generated Subtitles, in: 2024 6th International Conference on Natural Language Processing, ICNLP 2024, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 62–66. doi:10.1109/ICNLP60986.2024.10692675.
- [5] N. Radha, R. Pradeep, Automated subtitle generation, *International Journal of Applied Engineering Research* 10 (2015) 24741–24746.
- [6] V. Mukovoz, T. Vakaliuk, S. Semerikov, Road Sign Recognition Using Convolutional Neural Networks, in: E. Faure, Y. Tryus, T. Vartiainen, O. Danchenko, M. Bondarenko, C. Bazilo, G. Zaspá (Eds.), *Information Technology for Education, Science, and Technics*, volume 222 of *Lecture Notes on Data Engineering and Communications Technologies*, Springer Nature Switzerland, Cham, 2024, pp. 172–188. doi:10.1007/978-3-031-71804-5\_12.
- [7] A. Ramani, A. Rao, V. Vidya, V. R. B. Prasad, Automatic Subtitle Generation for Videos, in: 2020 6th International Conference on Advanced Computing and Communication Systems, ICACCS 2020, Institute of Electrical and Electronics Engineers Inc., 2020, pp. 132–135. doi:10.1109/ICACCS48705.2020.9074180.
- [8] S. Kiran, U. Patil, P. S. Shankar, P. Ghuli, Subtitle Generation and Video Scene Indexing using Recurrent Neural Networks, in: *Proceedings of the 3rd International Conference on Inventive Research in Computing Applications, ICIRCA 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 847–854. doi:10.1109/ICIRCA51532.2021.9544837.
- [9] X. Che, S. Luo, H. Yang, C. Meinel, Automatic Lecture Subtitle Generation and How It Helps, in: R. Huang, R. Vasiu, Kinshuk, D. G. Sampson, N.-S. Chen, M. Chang (Eds.), *Proceedings - IEEE 17th International Conference on Advanced Learning Technologies, ICALT 2017*, Institute of Electrical and Electronics Engineers Inc., 2017, pp. 34–38. doi:10.1109/ICALT.2017.11.
- [10] R. Sridhar, S. Aravind, H. Muneerulhudaakalvathi, M. Sibi Senthur, A hybrid approach for Discourse Segment Detection in the automatic subtitle generation of computer science lecture videos, in: D. E. Comer, P. Mueller, B. Mallick, S. Mukherjea, S. M. Thampi, D. Krishnaswamy, A. Sikora (Eds.), *Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2014*, Institute of Electrical and Electronics Engineers Inc., 2014, pp. 284–287. doi:10.1109/ICACCI.2014.6968422.
- [11] A. Karakanta, F. Buet, M. Cettolo, F. Yvon, Evaluating Subtitle Segmentation for End-to-end Generation Systems, in: N. Calzolari, F. Bechet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odiijk, S. Piperidis (Eds.), 2022 Language Resources and Evaluation Conference, LREC 2022, European Language Resources Association (ELRA), 2022, pp. 3069–3078. URL: <https://aclanthology.org/2022.lrec-1.328/>.
- [12] J. R. Calvo-Ferrer, Can you tell the difference? A study of human vs machine-translated subtitles, *Perspectives: Studies in Translation Theory and Practice* 32 (2024) 1115–1132. doi:10.1080/0907676X.2023.2268149.
- [13] D. D. Shanmugam, S. F. Syed, S. Dinesh, S. Chitrakala, VAR: An Efficient Silent Video to Speech System with Subtitle Generation using Visual Audio Recall, in: *Proceedings of the 5th International Conference on Inventive Research in Computing Applications, ICIRCA 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 814–821. doi:10.1109/ICIRCA57980.2023.10220944.
- [14] M. S. Martín, J. Heras, G. Mata, Automatic Generation of Subtitles for Videos of the Government of La Rioja, in: B. Dorronsoro, F. Chicano, G. Danoy, E.-G. Talbi (Eds.), *Optimization and Learning*, volume 1824 of *Communications in Computer and Information Science*, Springer Nature Switzerland, Cham, 2023, pp. 393–402. doi:10.1007/978-3-031-34020-8\_30.
- [15] J. D. Valor Miró, J. A. Silvestre-Cerdà, J. Civera, C. Turró, A. Juan, Efficient Generation of High-Quality Multilingual Subtitles for Video Lecture Repositories, in: G. Conole, T. Klobučar, C. Rensing, J. Konert, E. Lavoué (Eds.), *Design for Teaching and Learning in a Networked World*, volume 9307, Springer International Publishing, Cham, 2015, pp. 485–490. doi:10.1007/978-3-319-24258-3\_44.
- [16] M. Hotta, C. S. Leow, N. Kitaoka, H. Nishizaki, Evaluation of Speech Translation Subtitles Generated by ASR with Unnecessary Word Detection, in: *GCCE 2024 - 2024 IEEE 13th Global Conference on*

- Consumer Electronics, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 815–819. doi:10.1109/GCCE62371.2024.10760522.
- [17] A. Kaulage, A. Walunj, A. Bhandari, A. Dighe, A. Sagri, Edu-lingo: A Unified NLP Video System with Comprehensive Multilingual Subtitles, in: 2nd IEEE International Conference on Data Science and Information System, ICDSIS 2024, Institute of Electrical and Electronics Engineers Inc., 2024. doi:10.1109/ICDSIS61070.2024.10594128.
- [18] I. Al Sawi, R. Allam, Exploring challenges in audiovisual translation: A comparative analysis of human- and AI-generated Arabic subtitles in *Birdman*, PLoS ONE 19 (2024). doi:10.1371/journal.pone.0311020.
- [19] S. Kuroiwa, C. Oshima, T. Koita, Exploring a Hybrid System Combining AI and Human Intervention for Subtitle Creation in Entertainment Content, in: N. C. Callaos, E. Gaile-Sarkane, N. Lace, B. Sanchez, M. Savoie (Eds.), Proceedings of World Multi-Conference on Systemics, Cybernetics and Informatics, WMSCI, volume 2024-September, International Institute of Informatics and Cybernetics, 2024, pp. 72–73. doi:10.54808/WMSCI2024.01.72.
- [20] X. Qiu, Study on Automatic Generation of Teaching Video Subtitles Based on Cloud Computing, *Smart Innovation, Systems and Technologies* 156 (2020) 309–314. doi:10.1007/978-981-13-9714-1\_34.