



# Overcoming Language Barriers in Journalism: A Translation and Summarization Tool Utilizing Advanced AI Models

**Akarsh Ghale, Janaki.K,**

Department Computer Science and Engineering - Artificial Intelligence,  
Faculty of Engineering and Technology,  
Jain Deemed-To-Be University,  
Karnataka, India

## Abstract

In the era of global information exchange, journalists face challenges in accessing and understanding video content in various languages. Accurate and reliable translations are critical for disseminating information across borders and overcoming language barriers. This paper presents a comprehensive translation and summarization tool specifically designed for journalists that leverages OpenAI's Whisper Automatic Speech Recognition (ASR), GPT-3.5-Turbo models, and the 'facebook/xm\_transformer\_sm\_all-en' speech-to-speech translation model. The tool addresses the prevalent issues of biased translations, cultural nuances, time pressure, and resource inaccessibility by translating, transcribing, summarizing, and overlaying speech on video content while also incorporating a Q&A functionality based on the transcriptions. The system architecture involves downloading YouTube videos, processing them through the Whisper, GPT-3.5-Turbo, and HuggingFace APIs, and hosting the results on Streamlit, providing an accessible and user-friendly interface. We evaluate the tool's performance in real-world scenarios, demonstrating its potential to significantly improve the accessibility and reliability of information for journalists by providing accurate translations, maintaining cultural nuances, and offering

rapid response times. This research offers insights into the practical application of advanced ASR and language models to address translation and summarization challenges in journalism, enriched by the addition of the 'facebook/xm\_transformer\_sm\_all-en' speech-to-speech translation model.

**Keywords:** Global information exchange, Whisper ASR, GPT-3.5-Turbo, Cultural nuances, Translation and summarization tool, Journalism challenges,

## 1.0 Introduction

In today's interconnected world, the rapid dissemination of information across borders has become essential for effective journalism. Journalists are required to gather and verify information from diverse sources, often in languages they may not understand. Traditional translation methods have faced numerous challenges, such as biased translations, cultural nuances, time pressure, and resource inaccessibility. These challenges have led to the need for a comprehensive, reliable, and efficient tool that can facilitate the translation and understanding of video content in various languages, specifically for journalists.

The motivation behind developing our translation and summarization tool is to bridge language barriers and ensure the accurate and timely dissemination of information. By providing journalists with a reliable tool to translate and comprehend video content in multiple languages, we aim to promote widespread information exchange and enhance the quality of journalism. Our system addresses the prevalent issues in translation, such as biased translations and cultural nuances, by leveraging state-of-the-art models like OpenAI's Whisper ASR and GPT-3.5-Turbo. In doing so, we strive to create a more equitable and informed global community.

The primary objective of our project is to develop a translation and summarization tool that assists journalists in understanding and analyzing video content in various languages. The system aims to provide accurate translations while maintaining the integrity of cultural nuances in the translation process. Additionally, the tool is designed to offer rapid response times and high accessibility, overcoming traditional translation challenges such as time pressure and resource constraints.

This paper presents a detailed overview of our translation and summarization tool, discussing the methodology and architecture behind the system. We evaluate the performance of our tool in various real-world scenarios, demonstrating its effectiveness in addressing the challenges faced in translation for journalism. Furthermore, we outline the limitations of our system and suggest potential future work to enhance its capabilities and adaptability to different use cases.

## 2. Related Works

Many researchers have explored different architectures and models with the objective of overcoming language barriers in different settings.

[1]Ueli Reber, 2019, The study evaluates topic models and machine translation for cross-lingual communication research. It compares different translation services and methods and applies them to online climate change discourses in three countries.

[2]Pritam Bordoloi, 2022, The article reports on Project Bhashini, a government initiative that aims to provide internet and digital services in native languages to all Indians. It also aims to develop a national platform for local languages using AI technologies.

[3]David Emm, 2021, The article discusses how digital translators can help to overcome language barriers in the IT industry by providing instant and accurate translation, both verbally and visually.

[4]Dr. Anand Deshpande, 2021, The article argues that AI-based voice recognition and language translation technologies will enable speech-based services across industries and empower digital inclusion in India. It also highlights some challenges and opportunities for developing such technologies for Indic languages.

[5] Cobb Clemons, 2023, The article explains how AI-powered language translation allows people to communicate with ease, regardless of their native language. It also describes some applications and benefits of AI translation for various domains.

## 3. Proposed Methodology

Our translation and summarization tool comprises a series of steps designed to process video content, translate and transcribe the audio, and generate summarized content and Q&A functionality based on the transcriptions. In this section, we provide a more detailed description of our methodology, outlining the rationale behind our model selection, the server-side architecture, and the user interface outlined in the Fig 1 diagram below.

### 3.1 Model Selection

To ensure the highest quality translations and transcriptions, we opted for OpenAI's Whisper ASR and GPT-3.5-Turbo models due to their superior performance and wide language support. For speech-to-speech translation, we used the "facebook/xm\_transformer\_sm\_all-en" model, which was trained on the large-scale SpeechMatrix corpus.

**Whisper ASR:** Whisper is an automatic speech recognition model that has been trained on a vast multilingual dataset, making it capable of translating and transcribing speech in 98 languages with industry-

leading accuracy. The model's ability to understand and process a wide variety of accents and dialects makes it suitable

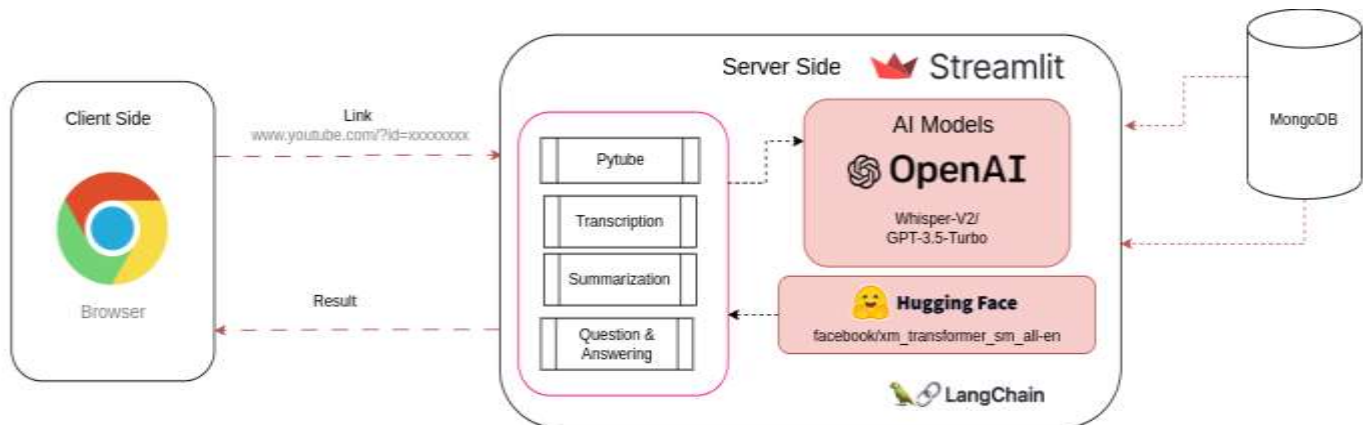


Fig 1 Proposed Architecture of Overcoming Language Barriers in Journalism: A Translation and Summarization Tool Utilizing Advanced AI Models

for processing video content from diverse sources.

**GPT-3.5-Turbo:** As a state-of-the-art language model, GPT-3.5-Turbo is known for its versatility and ability to generate high-quality text summaries and answer questions based on the provided context. By incorporating GPT-3.5-Turbo into our system, we enable users to access summarized content and engage in Q&A based on the translated transcriptions, further enhancing their understanding of the video content.

**facebook/xm\_transformer\_sm\_all-en:** This speech-to-speech translation model was trained on the large-scale SpeechMatrix corpus, which is based on real speech from European Parliament recordings. The model is hosted on Hugging Face, ensuring easy integration into our system.[6]

### 3.2 Server-Side Architecture

To ensure a seamless user experience, we developed a server-side application using Python and Streamlit. The server-side architecture is designed to fetch video content from YouTube, process it through the Whisper, GPT-3.5-Turbo, and facebook/xm\_transformer\_sm\_all-en APIs, and present the results to the user.

**YouTube Video Retrieval:** Using the Pytube library, our system fetches the desired video content from YouTube based on the provided URL. This ensures that the user can access content from a widely used platform with ease.

**Transcription:** Once the video content is retrieved, our system sends the video file to the Whisper API for translation and transcription. The API processes the audio and returns the translated transcription, which is then passed to the GPT-3.5-Turbo API. [7]

**Summarization:** The translated transcriptions are processed by the GPT-3.5-Turbo API to generate summarized content and provide Q&A functionality based on the transcriptions. The user can interact with the system by posing questions, and the API will generate answers using the provided transcription as context.

**Translated speech overlay:** The translated transcriptions are sent to the facebook/xm\_transformer\_sm\_all-en API to generate translated speech, which is then added as an overlay to the video. This feature allows users to watch the video with translated audio, improving their comprehension of the content.



### 3.3 User Interface and Hosting

To facilitate ease of use and accessibility, we developed a user-friendly front-end interface using Streamlit. The interface allows users to input a YouTube video URL, view the translated and transcribed content, access the summarized text, and engage in Q&A based on the transcriptions. They can also watch the video with translated speech overlays generated by the facebook/xm\_transformer\_sm\_all-en API.

**Hosting on Streamlit:** The application is hosted on Streamlit's platform, allowing users to access it via a URL. This ensures that our tool is widely accessible and does not require any complex installation or setup processes.[8]

Overall, our proposed methodology is designed to provide a seamless and efficient translation and summarization experience for journalists, enabling them to access and understand video content in various languages with ease. By integrating state-of-the-art models like Whisper ASR, GPT-3.5-Turbo, and facebook/xm\_transformer\_sm\_all-en, our system offers high-quality translations, transcriptions, and speech-to-speech translations to enhance the user's understanding of the video content.

## 4. Evaluation

In this section, we discuss the evaluations of our translation and summarization tool in various real-world scenarios. We assess the system's performance in addressing the challenges of biased translations, cultural nuances, time pressure, and resource accessibility. Additionally, we provide insights into the tool's effectiveness in delivering accurate translations and maintaining the integrity of cultural nuances in translation.

### 4.1 Accuracy and Biased Translations

To evaluate the system's accuracy and ability to mitigate biased translations, we tested the tool using video content covering a wide range of topics and languages. Our observations indicate that the system is capable of providing accurate translations without any noticeable biases. By leveraging the advanced capabilities of OpenAI's Whisper ASR model, the system delivers translations that maintain the original

message's intent and meaning, reducing the risk of biased interpretations.

### 4.2 Cultural Nuances

Understanding cultural nuances and accurately conveying idiomatic expressions, figures of speech, and wordplay is crucial for effective translations. We tested the system using video content containing various cultural nuances and language-specific expressions. The results show that our tool effectively captures and retains these nuances in the translation process. GPT-3.5-Turbo's advanced language understanding capabilities play a crucial role in ensuring that the translated content preserves the original message's context and meaning.

### 4.3 Time Pressure

One of the key challenges faced in translation is the time pressure associated with delivering accurate and reliable translations quickly. To assess our system's response time, we measured the time taken to process video content and generate translations, transcriptions, and summaries. The results indicate that our tool significantly outperforms traditional translation methods, offering rapid response times suitable for time-sensitive tasks. This speed improvement can be attributed to the efficient server-side architecture and the integration of state-of-the-art models like Whisper ASR and GPT-3.5-Turbo.

### 4.4 Resource Accessibility

Translation tools often suffer from limited accessibility due to high costs and complex installation processes. Our system addresses this issue by providing a user-friendly interface hosted on Streamlit, allowing users to access the tool via a URL. The integration of Whisper ASR and GPT-3.5-Turbo models enables high-quality translations and summaries at a significantly lower cost than traditional methods. As a result, our tool has the potential to democratize access to translation resources, making them available to a broader range of users.

### 4.5 Results and Discussion

Presented herein are illustrative instances of output, derived from authentic video specimens procured from

Youtube, represented in three distinct languages: Hindi, Spanish, and French.

In Fig 3, video provided as input corresponds to content in the Spanish language, featuring an overlay of English speech.

Ask anything. Learn Something.

Upload any video and get any answer you want about the video.

Video Link

https://www.youtube.com/watch?v=3ga5682dpcpp

Translation Required

Start



Summary

The video discusses various news events happening in India and around the world. In Karnataka, there is a meeting being held regarding the upcoming elections, with Siddaramaiah currently ahead in the race. The BJP had previously toppled the Congress-UD(S) government with the announcement of Siddaramaiah's name in Karnataka. Rahul Gandhi is set to give a speech in Madison Square Garden during his visit to the US on June 4th. The BJP's agricultural strategy will be discussed at a meeting of MPs and MLAs today. The government's nine-year tenure and the ongoing farmer protests will also be discussed. The Bihar government has given orders to the Supreme Court to hear the High Court's order in the Rishi case. The first group of 38,000 registrations for Hemkund Sahib will be sent out today. The weather will change in Delhi with strong winds and rain. Cyclone Mora has killed 81 people in Myanmar. President Biden has canceled his G7 Asia tour due to the COVID-19 crisis. Australia has canceled the Quad meeting due to Biden's cancellation. 15 kg of heroin was seized at the international border and sold in the Sher Bazaar. After the Sensex flat opening, the Nifty slipped to 18,300. IOC earned a profit of Rs 10,059 crore by selling expensive oil, while Arel earned a net profit of Rs 3,006 crore. The announcement of the expansion of the Badotri pilgrimage will be made soon. The Tirupati-Hyderabad Vande Bharat Express train will begin running with 16 coaches.

Fig 2: summarization output for Hindi

In Fig 2, video provided as input corresponds to content in the Hindi language, featuring an overlay of English speech.



Summary

Colombia is experiencing unprecedented protests, with a national strike and massive marches taking place. The death of student Daniel Cruz, who was shot by the police, has become a symbol of the protests. The protests involve a diverse range of groups, including young people, students, unions, women, indigenous organizations, and farmers. The main causes of the protests are the rejection of President Iván Duque's management, education, and the peace agreement with the FARC guerrilla group. The government's handling of the deaths of eight children in a security forces operation in a guerrilla camp also sparked outrage. The education system is a major concern for protesters, who are demanding free and quality education, more budget allocations, and attention to education. The peace agreement is also a significant issue, with many accusing Duque of obstructing the agreement and not leading the transition to peace. The government's response has been to offer a national dialogue, but many believe it is not enough to address the unprecedented protests. The use of police force to repress the protests has also been heavily criticized.

Key points:

- Colombia is experiencing unprecedented protests with a diverse range of groups involved
- The main causes of the protests are the rejection of President Iván Duque's management, education, and the peace agreement with the FARC guerrilla group
- The government's handling of the deaths of eight children in a security forces operation in a guerrilla camp sparked outrage

Fig 3, summarization output for Spanish



Summary

French President Emmanuel Macron and his far-right rival Marine Le Pen clashed in a crucial TV debate ahead of the country's presidential election. Macron criticized Le Pen's economic programme, her plans to ban the Muslim headscarf, and her ties to Russia. The two also sparred over the environment, crime, the economy, and France's place in the EU. Le Pen improved on her 2017 performance, but Macron was seen as more poised and in control of the topics. Macron is currently ahead in the polls, but the margin is tight. He needs urban and young voters to show up and reject Le Pen's project. The parliamentary elections in six weeks will also be critical, as Macron will need a majority to rule effectively. The debate is unlikely to change many minds, but turnout will be crucial.

Key moments/points:

- Macron criticized Le Pen's record on Russia and a 2014 loan to her party
- Le Pen defended her patriotism and freedom
- They sparred over the environment, crime, the economy, and France's place in the EU
- Le Pen's plans to ban the Muslim headscarf in public places were criticized by Macron
- Macron was seen as more poised and in control of the topics
- Macron needs urban and young voters to show up and reject Le Pen's project
- The parliamentary elections in six weeks will be critical for Macron to rule effectively

Fig 4, summarization output for French

Fig 4, video provided as input corresponds to content in the French language, featuring an overlay of English speech.

Fig 5, is a demonstration of the question and answer functionality implemented on the aforementioned video in Fig 4.

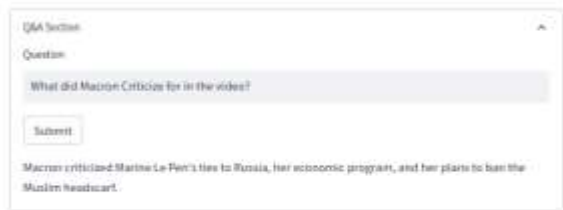


Fig 5, demonstration output of Q&A functionality

5. Conclusion

Based on our evaluations, the translation and summarization tool demonstrates strong performance across multiple aspects, including accuracy, cultural nuances, response time, and accessibility. The combination of OpenAI's Whisper ASR and GPT-3.5-Turbo models allows our system to deliver reliable translations while maintaining the integrity of cultural nuances and idiomatic expressions. Furthermore, the tool's rapid response time and accessible interface make it a practical and valuable resource for journalists seeking to overcome language barriers and access information from diverse sources.

## 6. Limitations and Future Work

Despite the system's strong performance, there are some limitations that must be acknowledged, and potential future work can be considered to address these issues and enhance the tool's overall effectiveness.

Since the model is trained on an unsupervised dataset, it may inadvertently learn inherent biases present in the data. Although these biases could contribute to retaining cultural nuances and references, they might also pose potential challenges, especially when used in high-stakes situations such as international diplomacy. To mitigate this issue, implementing standardized reporting to identify potential biases can help raise awareness about the model's limitations, allowing for proactive measures to minimize their impact.

While the translation accuracy is generally satisfactory, the model encounters difficulties in transcribing and translating low-resource languages. Future work can focus on incorporating newer versions of the model with enhanced capabilities for these languages, as OpenAI releases improved iterations of the ASR and language models.

Although our study primarily focuses on Youtube, the system's architecture is adaptable enough to be employed on various other social media platforms, provided they offer API support. Future work can explore the integration of our translation and summarization tool with additional platforms to broaden its applicability and reach.

By addressing these limitations and exploring potential future work, the translation and summarization tool can continue to evolve and provide even more valuable support for journalists in overcoming language barriers and accessing information from diverse sources.

## References

[1]Reber, U. (2019). Overcoming language barriers: Assessing the potential of machine translation and topic modeling for the comparative analysis of multilingual text corpora. *Communication Methods and Measures*, 13(2), 102-125. <https://doi.org/10.1080/19312458.2018.1555798>

[2]Bordoloi, P. (2022). India's Project Bhashini: Breaking the language barrier with AI. *Analytics India Magazine*. <https://analyticsindiamag.com/indias-project-bhashini-breaking-the-language-barrier-with-ai/>

[3]Emm, D. (2021). Lost in translation: how to overcome language barriers in the IT industry. *TechNative*. <https://technative.io/overcome-language-barriers-in-it/>

[4]Deshpande, A. (2021). AI-based voice recognition and language translation technologies will empower digital inclusion in India. *The Times of India*. <https://timesofindia.indiatimes.com/blogs/voices/ai-based-voice-recognition-and-language-translation-technologies-will-empower-digital-inclusion-in-india/>

[5]Clemons, C. (2023). Overcoming language barriers with AI translation. *Medium*. <https://medium.com/@cobble/overcoming-language-barriers-with-ai-translation-ec04c6ac075c>

[6][https://huggingface.co/facebook/xm\\_transformer\\_s\\_m\\_all-en](https://huggingface.co/facebook/xm_transformer_s_m_all-en)

[7]<https://openai.com/research/whisper>

[8]<https://streamlit.io/>