

SUBTITLE GENERATOR AND TRANSLATOR USING AI

By-

OMPRAKASH 21BCE1950

ABSTRACT

1. Our research introduces an innovative AI solution tailored for digital content creators — a Subtitle Generator and Translator designed to address the growing need for accessible and multilingual content. Utilizing advanced Natural Language Processing (NLP) and machine translation techniques, including transformer architectures and recurrent neural networks, our system autonomously generates precise subtitles and seamlessly translates them into multiple languages.
2. Highlighted features include real-time processing, adaptability to diverse languages and accents, and an intuitive user interface for straightforward customization. The AI models effectively manage intricate audiovisual contexts, such as background noise and overlapping speech, thereby enhancing accuracy and linguistic fluency in comparison to conventional methods.
3. As the digital landscape continues to evolve, this AI-driven solution emerges as a transformative tool for content creators, ensuring efficient subtitling and widespread accessibility for diverse global audiences. This abstract underscores the pivotal role of AI in promoting inclusivity and advancing the accessibility of multimedia content in our ever-expanding digital era.

OBJECTIVE

Efficient Subtitling: Develop an AI-driven system that can rapidly and accurately generate subtitles in real-time, ensuring an efficient subtitling process for various audiovisual content types.

Multilingual Translation: Implement advanced machine translation techniques to enable the automatic translation of subtitles into multiple languages, facilitating broader global accessibility and audience reach.

Adaptability to Diverse Content: Create a versatile system capable of handling diverse audiovisual contexts, including varying speaking rates, background noise, and overlapping speech, ensuring high-quality subtitles in complex scenarios.

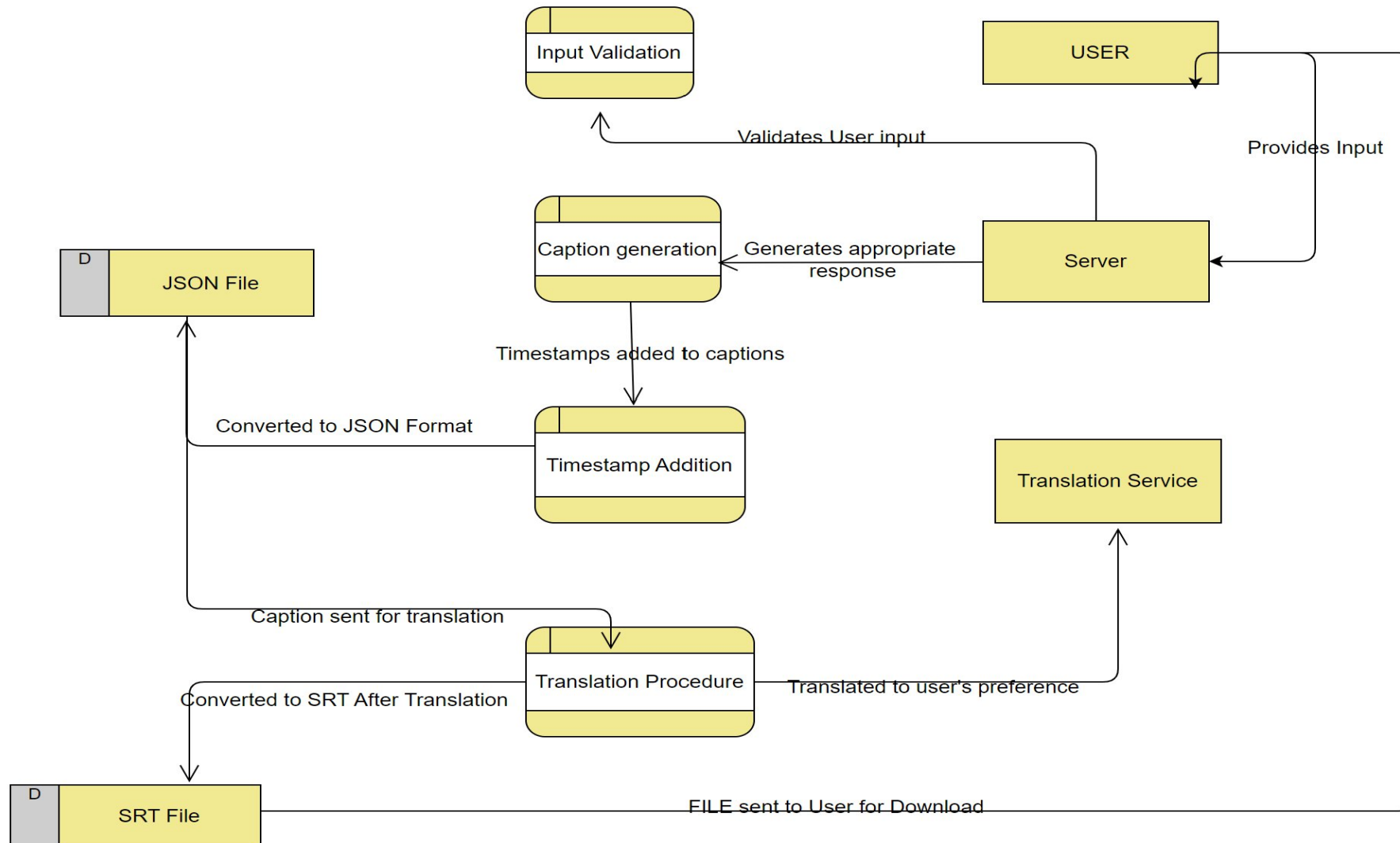
User-Friendly Interface: Design an intuitive user interface that allows content creators, filmmakers, and producers to easily customize and fine-tune subtitles according to their preferences, ensuring flexibility and creativity in the subtitling process.

Enhanced Accuracy: Utilize state-of-the-art NLP and machine translation models to achieve a higher level of accuracy in transcribing spoken words and translating them into different languages, surpassing the limitations of traditional subtitling methods.

Accessibility Improvement: Address language barriers and enhance the overall accessibility of digital content by providing automated, accurate, and contextually relevant subtitles, catering to a diverse global audience.

Technological Advancement: Stay at the forefront of technological advancements in AI, incorporating the latest models and techniques to continually improve the performance and capabilities of the Subtitle Generator and Translator.

WORKING PROCEDURE



MODEL DESCRIPTION

- **User Input:**
 - Users provide input, including Google Drive or YouTube links.
- **API Authorization and Server Processing:**
 - APIs validate the link for authorized access.
 - The server processes the validated link for content insights.
- **Response Generation and Timestamp Addition:**
 - The server generates a response based on the content.
 - Timestamps are added to enhance content accessibility.
- **JSON File Creation:**
 - The timestamped response is converted into a JSON file.
- **Language Translation and User Input:**
 - Users choose one of 20 supported languages for translation.

MODEL DESCRIPTION (CONTINUED)

- **SRT File Generation and Download Option:**
 - The model creates SubRip Subtitle (SRT) files.
 - Users decide whether to download or discard the generated SRT file.
- **User-Friendly Interface and Multilingual Support:**
 - The model maintains a user-friendly interface throughout.
 - It supports translation into 20 languages for a diverse user base.
- **Automation and Efficiency:**
 - Automation enhances efficiency in tasks like timestamp addition and translation.
- **Enhanced Accessibility and Dynamic Output:**
 - Timestamps and translations contribute to content accessibility.
 - The output adapts dynamically to user preferences.
- **Educational and Entertaining:**
 - The model's processing of content adds an educational and entertaining dimension.

S.No	Title of the Paper	Name of the Authors	YOP	Name of the journal	Methodology	Inference
I)	NLP-Driven Ensemble-Based Automatic Subtitle Generation and Semantic Video Summarization Technique	V. B. Aswin, Mohammed Javed, Parag Parihar, K. Aswanth, C. R. Druval, Anupam Dagar & C. V. Aravinda	2021	AIDE 2019: Advances in Artificial Intelligence and Data Engineering	Speech recognition can be used to generate subtitles for videos without them, and natural language processing (NLP) algorithms can be employed for summarization. The proposed model uses NLP-based text summarization algorithms and an Ensemble Technique to summarize videos using subtitles, ensuring a concise and accurate summary.	The use of NLP Algorithms proved to be a very efficient way to form abstracts of videos. The case of no subtitles was by using subtitle generation method to convert speech to text, which turned out to be of great use in normal day to day usage and were successful in attracting more users to the site

S.No	Title of the Paper	Name of the Authors	YO P	Name of the journal	Methodology	Inference
2)	Subtitle Generation and Video Scene Indexing using Recurrent Neural Networks	Sajjan Kiran; Umesh Patil; P Siddarth Shankar; Poonam Ghuli	2021	IEEE	Implementing sequence-to-sequence recurrent neural networks for audio data preprocessing, and network architectures. The use of the CTC (Connectionist Temporal Classification) algorithm for backpropagation of errors through time are explored.	sequence-to-sequence recurrent neural networks has the potential to improve accessibility, simplify subtitle creation, and enhancing the search and indexing capabilities of video files. This approach could positively impact the way users interact with video content and improve their overall experience.

S.No	Title of the Paper	Name of the Authors	YOP	Name of the journal	Methodology	Inference
3)	An Integrated Model for Text to Text, Image to Text and Audio to Text Linguistic Conversion using Machine Learning Approach	Aman Raj Singh; Diwakar Bhardwaj; Mridul Dixit; Lalit Kumar	2023	ISCON, IEEE	The model utilizes machine learning techniques for text-to-text, image-to-text, and audio-to-text conversions, with a focus on Indian languages. The model has been tested on large datasets of various Indian languages and uses machine learning, computer vision, and speech recognition.	the model is effective in accurately translating input data, with potential applications including language learning, accessibility, and cross-language communication. The proposed model aims to bridge communication across different language backgrounds.

S.No	Title of the Paper	Name of the Authors	YOP	Name of the journal	Methodology	Inference
4)	Artificial Intelligence Technology Supports the Following Research on the Generation of Subtitles for College English Teaching	Guangming Zou; Yingying Qiu	2022	ACAIT (IEEE)	The self-attention mechanism from artificial intelligence technology is then applied to construct the subtitle generation model. Following model training, its performance is evaluated based on accuracy, recall, and F1-Score	The application of the self-attention mechanism in constructing a subtitle generation model has shown good results in enhancing English teaching. The model demonstrates high accuracy, recall, and F1-Score, indicating its effectiveness in completing subtitle generation tasks.

S.No	Title of the Paper	Name of the Authors	YOP	Name of the journal	Methodology	Inference
5)	Deep-Sync: A novel deep learning-based tool for semantic-aware subtitling synchronization	Alejandro Martín, Israel González-Carrasco, Victor Rodriguez-Fernandez, Mónica Souto-Rico, David Camacho & Belén Ruiz-Mezcua	2021	S.I. : Data Fusion in the era of Data Science	The methodology involves integrating a deep language representation model and real-time voice recognition software to create Deep-Sync, an intelligent alignment tool for subtitles and audio-visual content.	Deep-Sync, offers a promising solution to the challenges faced in real-time subtitle generation. By integrating a deep language representation model and real-time voice recognition software, it successfully improves synchronization and accessibility of content

S.No	Title of the Paper	Name of the Authors	YOP	Name of the journal	Methodology	Inference
6)	Application of Translation Technology based on AI in Translation Teaching	Yu Yuxiu	2024	Systems and Soft Computing	<ol style="list-style-type: none">Utilized neural machine translation (NMT) and statistical machine translation (SMT) algorithms.Comparative analysis showed a 97% accuracy rate, surpassing traditional methods, improving student scores, and garnering teacher satisfaction.	<ol style="list-style-type: none">AI-based translation technology profoundly impacts language teaching, offering innovative strategies for addressing challenges and opportunities.Experimental tests in Chinese-English bilingual translation demonstrate the technology's efficacy, with future research aiming for broader applications.

S.No	Title of the Paper	Name of the Authors	YOP	Name of the journal	Methodology	Inference
7)	End-to-End Learning of Video-Based Text Generation From Multimodal Inputs	Xudong Lin, Gedas Bertasius, Jue Wang, Shih-Fu Chang, Devi Parikh, Lorenzo Torresani;	2021	Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 7005-7015	<ol style="list-style-type: none"> 1. Applied a differentiable tokenization mechanism for end-to-end training. 2. Integrated an autoregressive decoder for open-ended text generation, surpassing existing models. 	<ol style="list-style-type: none"> 1. Presents a unified framework for multimodal text generation using transformer networks, outperforming state-of-the-art models. 2. Demonstrates superior performance in video-based text-generation tasks like captioning, question answering, and audio-visual scene-aware dialog.

S.No	Title of the Paper	Name of the Authors	YOP	Name of the journal	Methodology	Inference
8)	A Speech-to-Subtitles corpus	Alina Karakanta, Matteo Negri, Marco Turchi	2020	Fondazione Bruno Kessler, Via Sommarive 18, Povo, Trento- Italy 2 University of Trento, Italy	<ol style="list-style-type: none"> 1. Leverages NMT for automated subtitling, emphasizing the need for high-quality, task-specific training data. 2. Introduces MuST-Cinema as a multilingual speech translation corpus for comprehensive NMT solutions. 	<ol style="list-style-type: none"> 1. Introduces MuST-Cinema to address gaps in subtitling corpora for Neural Machine Translation (NMT) solutions. 2. Proposes an automatic method for annotating other corpora, showcasing the ability to segment full sentences into subtitles.

S.No	Title of the Paper	Name of the Authors	YOP	Name of the journal	Methodology	Inference
9)	An attention enhanced sentence feature network for subtitle extraction and summarization	Chalothon Chootong, Timothy K. Shih, Ankhtuya Ochirbat, Worapot Sommool, Yung-Yu Zhuang	2021	Expert Systems with Applications	<ol style="list-style-type: none"> 1. Leverages both CNNs and Bi-LSTM Networks to capture critical information for sentence importance identification. 2. Salient sentence scores guide the summary generation process, enhancing the efficiency of video retrieval systems. 	<ol style="list-style-type: none"> 1. Introduces a novel multiple attention mechanism for video subtitle summarization using CNNs and Bi-LSTM Networks. 2. Experimental results showcase the model's superiority over baseline and state-of-the-art methods in terms of ROUGE-1, ROUGE-2, and ROUGE-L scores.

S.No	Title of the Paper	Name of the Authors	YOP	Name of the journal	Methodology	Inference
10)	Direct Speech Translation for Automatic Subtitling	Sara Papi, Marco Gaido, Alina Karakanta, Mauro Cettolo, Matteo Negri, Marco Turchi	2023	Transactions of the association for computational linguistics	<ol style="list-style-type: none">Utilizes GloVe for converting text data into sentence matrices and introduces a groundbreaking multiple attention mechanism.Experiments demonstrate superior performance over baseline and state-of-the-art models, affirming the effectiveness of the proposed approach.	<ol style="list-style-type: none">Introduces a multiple attention mechanism, combining Convolution Neural Networks (CNN) and Bidirectional Long Short-Term Memory (Bi-LSTM) for subtitle summarization in educational videos.The novel approach outperforms previous methods in educational video summarization, marking a significant advancement.

S.No	Title of the Paper	Name of the Authors	YOP	Name of the journal	Methodology	Inference
11)	Integrated Framework for Automatic Bilingual Subtitle Generation for MOOC Lecture Videos.	Xiaoyin Che; Sheng Luo; Haojin Yang; Christoph Meinel	2017	IEEE 17th International Conference on Advanced Learning Technologies (ICALT)	<ol style="list-style-type: none">Utilizes Automatic Speech Recognition (ASR), Sentence Boundary Detection (SBD), and Machine Translation (MT).Quantitatively evaluates auto-generated subtitles and manually produced subtitles for accuracy and time expenditure.	<ol style="list-style-type: none">Auto-generated subtitles in the original language save 54% of working time with a 54.3% error rate reduction.Proposed framework shortens bilingual subtitle preparation time by approximately 1/3 without quality decline.

S.No	Title of the Paper	Name of the Authors	YOP	Name of the journal	Methodology	Inference
12)	Subtitle Generation and Video Scene Indexing using Recurrent Neural Networks.	Sajjan Kiran, Umesh Patil, P Siddarth Shankar, Poonam Ghuli.	2021	Third International Conference on Inventive Research in Computing Applications (ICIRCA).	<ol style="list-style-type: none"> Utilizes sequence-to-sequence recurrent neural networks for video subtitle generation and scene indexing. Includes preprocessing audio data, extracting features, network architectures, CTC algorithm, and suitable evaluation metrics. 	<ol style="list-style-type: none"> Enhances video accessibility and indexing through sequence-to-sequence models. Addresses challenges in designing and training phase for improved video understanding.

S.No	Title of the Paper	Name of the Authors	YOP	Name of the journal	Methodology	Inference
13)	Automated Subtitle Generation Using Speech Recognition.	Prachi Sharma, Manasi Raj, Pooja Jangam, Sana Bhati, Prof. Neelam Phadnis	2019	SSRG International Journal of Computer Science and Engineering (SSRG - IJCSE) - Volume 6 Issue 4	<ol style="list-style-type: none"> 1. Follows three stages: audio extraction, speech recognition, and subtitle generation. 2. Focuses on automated subtitle generation using speech recognition to eliminate manual efforts. 	<ol style="list-style-type: none"> 1. Proposes an automated solution for video subtitles. 2. Emphasizes on the importance of metadata and automatic classification for video understanding.

S.No	Title of the Paper	Name of the Authors	YOP	Name of the journal	Methodology	Inference
14)	Artificial Intelligence Technology Supports the Following Research on the Generation of Subtitles for College English Teaching.	Guangming Zuo, Yingying Qiu.	2021	5th Asian Conference on Artificial Intelligence Technology (ACAIT).	<ol style="list-style-type: none"> Utilizes the self-attention mechanism for constructing an English teaching subtitle generation model. Applies the model to sequence and paragraph segmentation in a news dataset. 	<ol style="list-style-type: none"> Self-attention mechanism-based model effectively completes tasks with high accuracy, recall, and F1-Score. Provides strong support for enhancing English teaching through subtitle generation.

S.No	Title of the Paper	Name of the Authors	YOP	Name of the journal	Methodology	Inference
15)	A Novel Approach in the Automatic Generation of Regional Language Subtitles for Videos in English.	Bilva Teja R, Nischitha P Chinnari, Prerana Hadadi, Surabhi Shanbhogh, V R Badri Prasad.	2023	IEEE 8th International Conference for Convergence in Technology (I2CT).	<ol style="list-style-type: none"> 1. Addresses automated creation of Kannada subtitles for videos. 2. Highlights the use of Natural Language Processing (NLP) and Speech Recognition in computer science for subtitle generation. 	<ol style="list-style-type: none"> 1. Focuses on the significance of subtitles in regional languages. 1. Emphasizes on the automation of subtitle generation to save time, cost, and labor.

REFERENCES

- <https://ieeexplore.ieee.org/document/9574684> (An Approach for Audio/Text Summary Generation from Webinars/Online Meetings)
- https://link.springer.com/chapter/10.1007/978-981-15-3514-7_1 (NLP-Driven Ensemble-Based Automatic Subtitle Generation and Semantic Video Summarization Technique)
- <https://ieeexplore.ieee.org/document/9731269> (Artificial Intelligence Technology Supports the Following Research on the Generation of Subtitles for College English Teaching)
- <https://ieeexplore.ieee.org/document/9544837> (Subtitle Generation and Video Scene Indexing using Recurrent Neural Networks)
- <https://link.springer.com/article/10.1007/s00521-021-05751-y> (Deep-Sync: A novel deep learning-based tool for semantic-aware subtitling synchronisation)
- <https://ieeexplore.ieee.org/abstract/document/8001709> (Integrated Framework for Automatic Bilingual Subtitle Generation for MOOC Lecture Videos)
- <https://ieeexplore.ieee.org/abstract/document/9544837> (Subtitle Generation and Video Scene Indexing using Recurrent Neural Networks)
- <https://www.internationaljournalssrg.org/IJCSE/2019/Volume6-Issue4/IJCSE-V6I4P103.pdf> (Automated Subtitle Generation Using Speech Recognition)

REFERENCES

- <https://ieeexplore.ieee.org/document/9731269> (Artificial Intelligence Technology Supports the Following Research on the Generation of Subtitles for College English Teaching)
- <https://ieeexplore.ieee.org/document/10126282> (A Novel Approach in the Automatic Generation of Regional Language Subtitles for Videos in English)
- <https://www.sciencedirect.com/science/article/pii/S2772941924000012> (Application of Translation Technology based on AI in Translation Teaching)
- https://openaccess.thecvf.com/content/CVPR2021/html/Lin_Vx2Text_End-to-End_Learning_of_Video-Based_Text_Generation_From_Multimodal_Inputs_CVPR_2021_paper.html (End-to-End Learning of Video-Based Text Generation From Multimodal Inputs)
- <https://arxiv.org/abs/2002.10829> (MuST-Cinema: a Speech-to-Subtitles corpus)
- https://direct.mit.edu/tac1/article/doi/10.1162/tac1_a_00607/118115 (Direct Speech Translation for Automatic Subtitling)
- <https://www.sciencedirect.com/science/article/abs/pii/S0957417421003870> (An attention enhanced sentence feature network for subtitle extraction and summarization)