AI Weaving Words into Vision: Multilingual Videos for a New India

Dr. Namrata Tapaswi
Acropolis Institute of Technology &n Research Indore, India

ABSTRACT

The diversified linguistic landscape of India is responsible for breaking down boundaries across languages and exhibiting the country's rich past. This project aims to provide information in many Indian languages and convert written text into engaging films. According to the research, press releases will be transformed into interactive videos in 18 indigenous languages of India in addition to English. Employing state-of-the-art technology aims to meet the evolving preferences of a demographic that is more drawn to video content, to reimagine how information is shared. The method is automated and utilizes artificial intelligence to transform text into visually captivating films with audio and seamless transitions. These videos are designed to fulfill the demands of the modern digital age. Their seamless dissemination on social media platforms enables effective engagement with a diverse audience. The innovative endeavor aims to transform information transmission by using cutting-edge technology to match audience demands for engaging video content. The project recognizes and celebrates the great variety of languages, ensuring that information is accessible to a broader audience. The project exclusively utilizes photographs and video clips that are devoid of copyright restrictions to ensure genuineness and adherence to legal requirements.

Keywords: text image, video image, India’s regional languages, Text-to-Video conversion.

1 INTRODUCTION

It is essential to modify communication tactics to effectively include individuals in the ever-changing media environment. As people's attention spans continue to decrease, it is necessary to convert press releases, currently written in text style, into captivating films [1]. The proposed program aims to generate videos automatically from press releases written in English and 18 regional languages, such as Bengali, Assamese, Hindi, Punjabi, Tamil, Sindhi, Santali, Gujarati, Odia, Konkani, Marathi, Kashmiri, Telugu, Kannada, Bodo, Dogri, Malayalam, Maithili, Assamese, and Manipuri [19]. An essential component to incorporate is the functionality for keeping an assortment of photographs and video clips that are necessary to produce the film. Before being published, the film should undergo a thorough evaluation by the appropriate public relations representative. The technology will streamline the management of a wide-ranging collection of visuals, by press communication guidelines. Before being published, every video produced will undergo a thorough assessment by the designated public relations representative to ensure precision and adherence to communication guidelines.

The program is equipped with a notification mechanism that informs the relevant public relations representative about the requirement for final approval. Upon authorization, the system will instinctively publish the video to the relevant social media platforms, ensuring prompt and extensive dissemination[4]. By utilizing this program, Press will have the capability to transform press releases into visually captivating and comprehensible video content, thereby bridging the disparity between conventional text-based communication and modern audience inclinations[6]. The update preserves the integrity and reliability of media communication standards while simultaneously addressing the evolving requirements of the audience.

2. LITERATURE REVIEW

Recent publications have commonly utilized methodologies such as cut-based segmentation, color-based indexing, k-means-based dimensionality reduction, and data clustering. The entirety of the data obtained from these documents is accessible in a spreadsheet that is publicly accessible [1]. This article aims to include both geographical and temporal information from a video into embeddings, and subsequently compare videos based on these embeddings. We propose an architectural framework that enables the exploration of various embeddings of a video to get other movies related to the same occurrence or event [2]. The objective of this study is to address the issue of retrieving videos on a big scale using a query image. To begin with, we establish the problem of querying the top-k images in a video. Next, we integrate the advantages of convolutional neural networks and the Bag of Visual Word module to create a model for extracting and representing information from video frames. To fulfill the demands of large-scale video retrieval, we suggest using a visually weighted inverted index and its corresponding algorithm to enhance the efficiency and precision of the retrieval process [3]. A method is introduced whereby a video sample is matched with a text sentence from a corpus, and vice versa. Typically, video and text matching involve creating a common embedding space, and the encoding of one modality is not influenced by the other. In this study, we encode the dataset data by considering the important information of the query. The efficacy of the strategy is shown to stem from aggregating relationship data between words and frames [6]. We showcase the wide applicability and effectiveness of our noise estimation technique by achieving comparable outcomes to the most advanced methods on five distinct benchmark datasets for two demanding multimodal tasks: Video Question Answering and Text-To-Video Retrieval. In addition, we offer a theoretical probabilistic error bound that supports our empirical findings and examines instances of failure [8]. This study examines the video-to-text challenge, which aims to establish a connection between an input video and its corresponding textual description. The association can be primarily established by extracting the most pertinent descriptors from a corpus or developing a new one based on a contextual video. These two methods are fundamental jobs for the Computer Vision and Natural Language Processing fields. They are known as the text retrieval from video tasks and the video captioning/description tasks. These two challenges are significantly more intricate than the act of guessing or recovering a solitary sentence from an image [9]. Suggest integrating generative processes into the cross-modal feature embedding to get both global abstract features and local grounded features [12]. Suggest an innovative hybrid deep learning architecture that demonstrates excellent efficiency in sentiment analysis for languages with limited resources. Assess our suggested method for analyzing sentiment at both a broad and detailed level on four Hindi datasets that encompass different subject areas [16]. The objective of this study is to enhance previous surveys by creating and implementing a systematic review of Content-Based Visual Information Retrieval (CBVIR). The SR approach specifically enables one to achieve a consistent and comprehensive examination of the pertinent literature while minimizing subjective bias [20].

The automation of this nature ensures both the prompt transmission of information and the adherence to legal and credibility standards, which is a crucial aspect in the field of press releases. The literature review acts as a basis for innovative software solutions by combining multiple research threads. The initiative seeks to bridge the divide between traditional text-based news releases and the evolving preferences of modern customers.

3 . ASPECTS OF METHODOLOGY

In a society characterized by the unrestricted flow of technology and information, there is a fascinating endeavor underway to transform verbose press releases into entertaining videos that can be understood in multiple languages[19]. The methodology outlines our utilization of machine learning and artificial intelligence to achieve this. By employing several techniques, we condense lengthy press releases into concise and informative video content [2,3]. We utilize video production techniques to provide visually engaging content, using relevant graphics and employing several languages for narration. Additionally, can condense the material into a concise summary. Subsequently, these movies are actively distributed and crafted to captivate people. Continual input facilitates the process of improvement. Talking head technology is utilized to increase the engagement of the information [4]. The comprehensive procedure elucidates the transformation of intricate textual content from English and 18 regional languages into visually captivating films. Additionally, it ensures that individuals with diverse linguistic backgrounds in India may readily comprehend the material. This approach is depicted in Figure 1, demonstrating the sequential progression from the written content to the ultimate visual output.

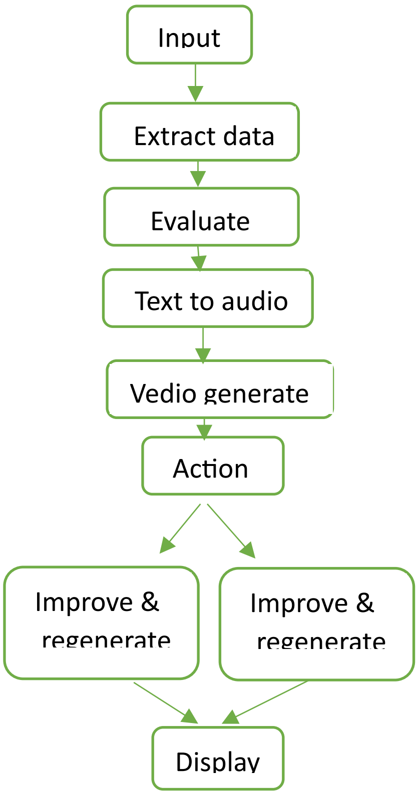


Figure 1: Video creation process map

3.1 Information Gathering and Preparation

The first tasks involving data collecting and preparation are essential for the complex process of converting news items into multilingual video material[7], as elucidated in this scientific publication. To begin, it is necessary to acquire press releases from authorized sources and confirm their validity, adherence to legal requirements, and genuineness. These factors are crucial for subsequent handling and examination. The subsequent Preprocessing phase is crucial for organizing and improving the textual data. Often, the gathered textual material comes in various formats, with inconsistencies and unnecessary features that impede further analysis and subsequent processing. The preprocessing stage encompasses essential procedures focused on improving and structuring the collected content. The tasks involve removing superfluous noise, standardizing text forms, and resolving anomalies in language [19]. This essential method ensures that the data is standardized, purified, and improved, establishing the basis for efficient processing in the next stages. Noise reduction involves the elimination of unwanted letters, punctuation, and unusual formats. Standardization duties focus on text encoding and addressing language-specific peculiarities encountered in press releases [5]. Ensuring linguistic coherence throughout the dataset is crucial for establishing a uniform and systematic approach, which will streamline the later stages of natural language processing and translation. The meticulous preparation establishes the foundation for subsequent stages in the conversion process from text to video, facilitating the creation of accurate and valuable multilingual video material derived from press releases[11], which remains the primary focus of this scholarly investigation.

3.2 Voice-over in multiple languages

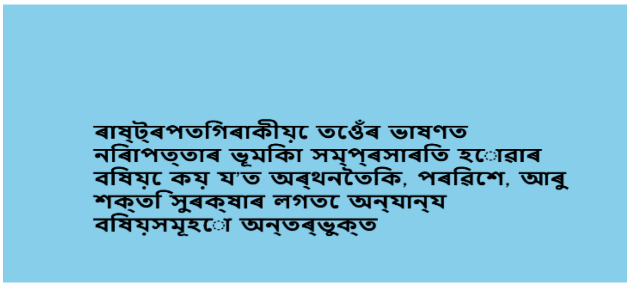
The Voice-over in multiple languages process is a crucial element of the comprehensive methodology developed for the transformation of written content from Press releases into video content in many languages, which is the main emphasis of this study. This phase involves the use of advanced Artificial Intelligence techniques [15], specifically speech synthesis, to translate and vocalize condensed text into multiple languages. These languages include English as well as 18 regional languages such as Bengali, Assamese, Hindi, Punjabi, Tamil, Sindhi, Santali, Gujarati, Odia, Konkani, Marathi, Kashmiri, Telugu, Kannada, Bodo, Dogri, Malayalam, Maithili, Assamese, and Manipuri, covering a wide range of linguistic diversity. AI-powered speech synthesis enables the conversion of condensed Press release text into realistic speech, preserving language-specific intricacies, accents, intonations, and pronunciation, guaranteeing an authentic and understandable watching experience. The creation of accurate voice-overs in regional languages relies on comprehensive datasets, which are enhanced by advanced AI algorithms. These algorithms can capture the cultural nuances and grammatical intricacies that are specific to each language. Iterative improvement cycles are used to continuously enhance speech synthesis models [16]. User feedback methods are integrated to update these models, assuring accurate and realistic voice-overs. This results in a more immersive and natural audience experience.

The procedure of Voice-over in multiple languages greatly improves the inclusiveness of video information obtained from Press releases[12] . It promotes wider access and understanding among diverse audiences, surpassing language obstacles, and fostering greater involvement among viewers with different linguistic origins. AI-powered speech synthesis enhances linguistic accuracy and increases the accessibility and authenticity of content, resulting in a more engaging and comprehensible viewing experience for diverse audiences with different preferences for language and socioeconomic backgrounds.

3.3 Images for the Background

The incorporation of Images for the Background is a crucial component in the comprehensive framework designed to transform textual data from news articles into multilingual video material[13,14], which is the main objective of this study. During this step, the textual information in movies is enhanced by including visual elements acquired through web scraping. Background photos are crucial for strengthening textual tales since they enable the smooth integration of scraped images with the text. The process commences by utilizing specialized libraries and tools for web scraping, ensuring a meticulous arrangement of images based on the primary concepts and themes conveyed in news articles. The purpose of these meticulously selected photographs is to visually depict the concepts[13], locations, or occurrences that are being addressed. This results in a coherent and captivating visual narrative.

Integrating Images of the Background obtained through web scraping greatly improves the effectiveness of content delivery, creating greater intimacy among text and a variety of audiences. The combination of these pictures and textual content synergistically enhances the video story, resulting in heightened viewer comprehension and engagement[15]. These visuals serve two purposes: they act as visual backgrounds and enhance the text, enabling a more comprehensive presentation of complex ideas and increasing audience involvement Figure 2. Furthermore, the intentional integration of text obtained by site scraping with graphics ensures a comprehensive and engaging press release delivery in many geographical areas, providing a wide range of languages. Viewers derive advantages from a more captivating and instructive viewing experience due to the implementation of this inclusive method. By integrating certain frameworks and resources for online scraping and text-image production, the visuals presented are improved in terms of relevancy and accuracy. This promotes a greater level of comprehension and connection among different language communities.



ৰাষ্ট্ৰপতগিৰাকীয়ে তওঁৰ ভাষণত
নৰিাপত্তাৰ ভূমকিা সম্প্ৰসাৰতি হ◌োৱাৰ
বষিয়ে কয় য'ত অৰ্থনৈতকি, পৰৱিশে, আৰু
শক্তি সুৰক্ষাৰ লগতে অন্যান্য
বষিয়সমূহ◌ো অন্তৰ্ভুক্ত

Figure. 2: Generated text image for Marathi.

3.4 Video creation

The Video Creation process, which is a crucial and captivating feature of this complete program, involves transforming Press Releases into films as part of a larger transformation effort. This step skilfully combines multimedia elements derived from the modified textual material, creating customized and captivating visual stories[1,6]. This stage utilizes advanced artificial intelligence, machine learning algorithms, generative adversarial networks (GANs), and advanced video processing procedures to meticulously compile video footage that captures the core message of the news releases while incorporating different linguistic contexts. The method begins by arranging pertinent textual information, alongside photos and graphic elements obtained through web scraping, within a platform for generating videos.

The Video Development procedure utilizes state-of-the-art techniques and structures to handle multilingual content. This ensures the production of culturally appropriate videos for each language. Moreover, the integration of regional accents in the speech synthesis process leads to the generation of authentic-sounding voice-overs in all 18 languages. Furthermore, the movies are enriched by using background images obtained through web scraping, resulting in a visual setting that complements and amplifies the written content. The algorithms arrange and time the textual and visual elements to create visually captivating videos that effectively capture the essence of the press releases. To enhance engagement, the movie incorporates a voice actor (Fig. 3), which serves as a representation of a human-like entity[8,11]. The movies are enhanced by incorporating iterative refinement methods, that include consumer surveys and optimization processes, to ensure correctness, cultural relevance, and accessibility across different linguistic contexts. The creation of videos serves as a bridge between written knowledge and captivating visual narratives, offering viewers a comprehensive and comprehensive observing encounter in English as well as many regional languages.

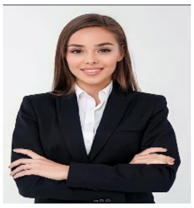


Figure 3: Voice Actor

3.5 Publication and availability

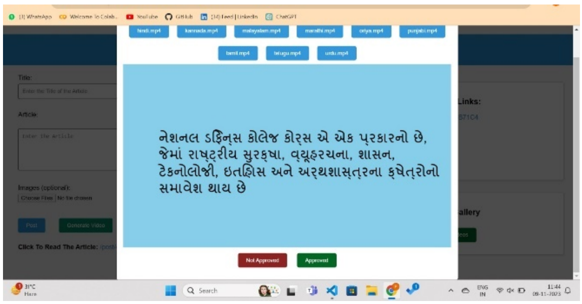
The phase plays a crucial role in addressing evolving user preferences. To effectively engage users with decreasing attention spans, it is crucial to provide news releases in video format. The program design intends to autonomously produce videos using public Press Releases, while strictly sticking to the utilization of copyright-free images and clips from credible sources[13]. The system incorporates a library that enables the storage of an array of photos and clips, hence facilitating the process of video generation.

Before publishing, the video footage produced is subjected to a rigorous evaluation process by the relevant public relations representative to verify its authenticity and ensure its high quality[15]. The program incorporates a notification mechanism that promptly informs the officer to evaluate and authorize, hence optimizing the approval process. Once accepted, the program features an automated upload option that disseminates it to relevant social media platforms. The architecture of the software is specifically engineered to automate the process of video production, ensuring the maintenance of content integrity and strict adherence to copyright regulations. The necessity of official endorsement and screening ensures the pertinence and accuracy of the information. The seamless notifications and automatic uploading capabilities ensure efficient dissemination, ensuring that Press video material is widely accessible on social media platforms. The system effectively delivers information promptly and engagingly, while also adjusting to the evolving user preferences for ingesting video-based material [18].

3.6 User-Driven Enhancement

An essential element in transforming press releases into films involves the incorporation of user feedback and iterative enhancements. In today's world, wherein the attention spans of individuals are diminishing, it has become crucial to convey information through visually captivating video formats[7,9]. The application, devised to generate videos autonomously from these publications, utilizes copyright-free photographs and videos sourced from an extensive library to streamline the content creation process. Each film (Fig. 4) is meticulously reviewed by an appointed public relations representative to guarantee precision and relevance. The software expedites this procedure by promptly notifying the officer for swift review and approval. Once the content has been accepted, the program instantly publishes it on the relevant social media platforms.

Importantly, the program employs a strategy of continually soliciting user feedback. This iterative method aims to enhance the video content by incorporating user recommendations and preferences[14]. The program guarantees the videos remain relevant and impactful by incorporating user feedback and adapting to the changing requirements and preferences of the audience. The flexible method of incorporating user feedback results in ongoing improvements in the video material, ensuring that it closely matches the ever-changing expectations of the audience. Strategy plays a crucial role in enhancing the audience's viewing experience by making it more captivating and pertinent.



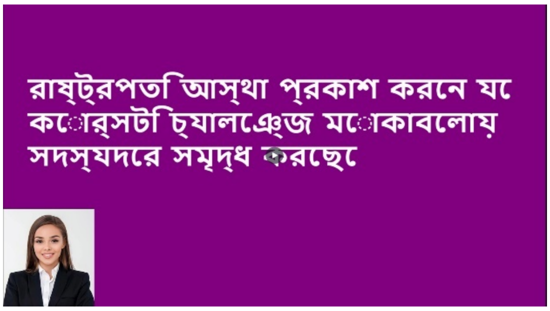
નેશનલ ડિફેન્સ કોલેજ કોર્સ એ એક પ્રકારનો છે, જેમાં રાષ્ટ્રીય સુરક્ષા, વ્યૂહરચના, શાસન, ટેકનોલોજી, ઇતિહાસ અને અર્થશાસ્ત્રના ક્ષેત્રોનો સમાવેશ થાય છે

Figure 4: Video critique

4 EMPIRICAL FINDINGS

This research presents the experimental findings of the Text-to-Video conversion, highlighting the culmination of a resilient and groundbreaking procedure. This section presents the results obtained from the complex combination to convert news written content into dynamic visual stories in English and 18 regional languages. This paper examines the effectiveness and viability of the proposed technology, demonstrating both the successful aspects and the obstacles faced in transforming text into captivating video content. The results presented here provide useful insights into the efficacy, difficulties, and flexibility of the Text-to-Video translation system in meeting the varied linguistic landscape and ever-changing audience preferences.

The movie was generated by utilizing an English literature input taken as an experiment input. The article titled "63rd National Defence College Course visited the President of India, Smt. Draupadi Murmu at Rashtra Pati Bhavan" had a word count of 1202. The text was condensed into a concise synopsis consisting of 300 words. Subsequently, the summary is translated into many languages, both in written and auditory formats, with a high degree of realism[19]. The combination of visuals, audio, and a person speaking is transformed into a video. The photos obtained by web scraping were highly pertinent and precise. The film produced exhibited a captivating visual depiction accompanied by an authentic voiceover. The film in Figure 5 features a talking head that accurately synchronizes its lip movements with the article being presented.



রাষ্ট্রপতি আস্থা প্রকাশ করেন যে কোর্সটি চ্যালেঞ্জ মোকাবেলায় সদস্যদের সমৃদ্ধ করেছে

Figure 5: Video Output

5. USE CASES

It serves as a portal to the adaptability of this state-of-the-art technology, offering a viewpoint that comprehends its practicality and versatility across various scenarios. This section presents a study that aims to provide insight into the various uses of Text-to-Video technology[2,13]. It emphasizes how this technology has the potential to revolutionize audience engagement, information sharing, and communication in the current dynamic and multicultural setting.

5.1 Collaborations with Regional News Agencies

Incorporating footage from local news channels into videomaking is a strategic method that improves the richness and legitimacy of the narrative. The movie achieves a sense of vitality and relevancy by smoothly integrating news updates, articles, and regionally specific information obtained from local or regional news outlets[5,18]. The inclusion of up-to-the-minute, location-specific news not only informs the audience about ongoing occurrences but also makes a direct link between the material and the viewer's environment. The film stimulates attention and discourse among viewers by addressing themes that directly influence the community, promoting a sense of regional belonging and involvement. In addition, making use of local news updates offers significant context, guaranteeing that the content stays current and relevant.

In the realm of storytelling, the incorporation of trustworthy news sources enriches the narrative by introducing a diverse range of viewpoints and expert analysis [17]. By incorporating additional levels of depth and authenticity, it enhances the overall quality of the text. The video serves as a reliable source of information that caters to the passions as well as issues of the audience by including local news components. This integration serves to educate the audience [13]. The effortless incorporation of these elements into videos enables content creators to craft a remarkable viewer experience and cultivate a robust connection between their audience and the material.

5.2 Audiovisual

Incorporating multimedia components is essential for crafting a visually captivating and immersive narrative. The process involves seamlessly integrating diverse visual and auditory components, such as images, audio files, and video clips, into the video content [8,9,11]. The film surpasses the constraints of textual content by adeptly integrating different multimedia resources to generate a dynamic and visually compelling experience. Images and videos offer a visual framework, simplifying intricate ideas into easily comprehensible images, and augmenting the audience's comprehension. Concurrently, the inclusion of audio components enhances the information by providing a sense of depth and imbuing it with emotional and resonant qualities. The seamless integration of audio and visual elements not only enhances the storytelling but also enhances audience involvement, making the material more easily understandable and memorable.

The incorporation of multimedia integration allows for a wider audience to connect with the information on a personal level, as it caters to diverse learning styles. Interactive learners can engage in audiovisual exercises and tests, listeners may communicate with spoken tales, and visual learners benefit from visuals and animations[12]. Due to the versatility of multimedia elements, creators can employ diverse storytelling strategies and innovative forms of artistic expression, resulting in a captivating and unique watching experience. Ultimately, a skilfully implemented multimedia strategy enhances the overall influence of the film by guiding viewers through a compelling, instructive, and visually engaging experience that creates a lasting impact and fosters a deeper understanding of the subject matter.

5.3 Educational Material Videos

Educational videos are created by transforming conventional text-based instructional content into engaging and interactive video formats. The recognition that aural and visual inputs often enhance the process of learning acts as the primary catalyst for this phenomenon. Educational films encompass a variety of instructional methods, including animated clarifications, hands-on demonstrations, and lectures supplemented with supporting graphics. These movies accommodate a diverse range of learners, such as those who learn best through visual or auditory means, by presenting educational content in a more accessible and appealing manner. By employing pictures and illustrations, they can enhance comprehension of abstract ideas and streamline complex concepts. Furthermore, educational videos can foster active participation by incorporating quizzes, interactive components, and the option to pause and review material, thus enabling students to acquire knowledge at their preferred speed. The process of transforming educational material into films is in line with contemporary intellectual trends, wherein multimedia content plays a crucial role in enhancing the experience of learning and facilitating knowledge retention. The remarkable flexibility of videos for learning is worth mentioning. They can be readily disseminated via several online platforms, reaching a worldwide audience. Consequently, they possess significant value as instruments for courses offered online, distance learning, and mixed learning environments. The ability to amend and edit videos enables the refinement of content, ensuring that the material remains up-to-date and pertinent. Creating films from instructional content enables educators to produce captivating and efficient learning materials that cater to various learning preferences, foster profound understanding, and support the wider dissemination of knowledge in a constantly evolving educational environment.

5.4 Product Advertisement Conversions

Converting written specifications or commercials into engaging video advertising is an innovative approach in the field of text-to-video projects. This innovative approach employs multimedia elements to animate otherwise static text, resulting in visually captivating narratives that captivate readers. The initial stage of the process involves the meticulous selection of key characteristics, advantages, and distinctive selling factors, as well as the thorough curation of product-related language. The video advertisement's hierarchical foundation is based on this text [15]. The product description incorporates vibrant visuals, attention-grabbing text overlays, and enticing music to create an engaging and immersive experience. Within the realm of text-to-video endeavors, the transformation of textual descriptions of products or marketing information into vibrant video commercials is a revolutionary strategy. This innovative technique utilizes multimedia elements to animate text that would normally remain static, resulting in visually captivating narratives that captivate readers. The process commences with the conversion process and is significantly influenced by visual components [17]. The product is displayed in high-resolution photos that emphasize its processes and features from different perspectives. Strategically positioned text overlays serve to highlight important information, promotions, or time-limited offers, thus improving the viewer's comprehension and persuading them to act. Strategically incorporating audio elements such as music, voice-overs, or sound effects in advertisements amplifies their emotional impact by evoking emotions or thoughts that strongly resonate with the intended audience.

The resultant videos are crafted to endorse a diverse range of products and services, encompassing trends, electronic goods, and even experiential offerings such as vacation packages. These movies not only display the product's visual attractiveness and practicality but also evoke intense desires and a sense of urgency in potential customers by developing an emotional connection [10]. Video advertising significantly enhances conversion rates by transforming written material into a visually captivating and emotionally persuasive medium. Potential clients perceive the product as more valuable when it incorporates compelling pictures, concise writing, and impactful music.

5.5 Instructional Video for a Product

The use of text-to-video technology to create explanatory or mentoring videos for products represents notable progress in improving user experience and understanding of the product. This novel method entails the effortless transformation of written guidelines or instructions into a dynamic video format, providing step-by-step visual demonstrations on how to use the product, assemble it, or utilize its features[3,7,17]. Through the utilization of multimedia components, which include animations, graphics, and captivating narration, these movies convert mundane written directions into interactive and captivating visual tutorials.

By providing meticulous visual cues and clear demonstrations, customers acquire a thorough comprehension of the product's functionality, hence eradicating any confusion or potential mistakes. These movies function as both accessible learning resources and cater to varied learning preferences, particularly benefiting visual learners who comprehend things more effectively through examples. Tutorial videos enhance the onboarding process for customers, guaranteeing that they feel self-assured and powerful when utilizing the product [14]. The interactive graphics, combined with succinct and comprehensible explanations, boost users' assurance, resulting in a more favorable engagement with the product. These movies are especially advantageous for items that have intricate assembly procedures, detailed technical specifications, or distinctive characteristics, as they offer consumers vital knowledge and guidance for problem-solving.

Through the process of converting instructions in writing into visually integrated guidance, these movies optimize customer happiness and reduce the need for support requests. They function as indispensable assets, assisting users in resolving problems autonomously and enhancing their product experience. Moreover, instructional videos play a substantial role in fostering brand loyalty by positioning the brand as customer-centric and easy to use.

5.6 Tourism Videos

Text-to-video technology is a potent tool for converting written information about regional areas or tourist sites into fascinating and interactive video experiences in the context of tourism. The technique involves transforming descriptive text, sometimes including details about popular tourist destinations, landmarks, cultural sites, and geographical highlights, into visually captivating video footage [1,6]. This technology converts textual material into a dynamic video format, creating virtual tours or advertising videos that display the beauty, distinctiveness, and main attractions of a particular region or site. These videos offer prospective passengers a vivid glimpse into the attractions and amenities of the area, making them a compelling tool for promoting tourism. By employing captivating imagery, compelling storytelling, and melodic accompaniment, they may craft a multimodal encounter that allures spectators to explore the area. These videos, encompassing expansive aerial perspectives of natural landscapes and detailed shots of cultural festivities, possess the capacity to accentuate the multifaceted and lively characteristics of a place, rendering it an enticing tourist destination. By integrating interactive features, such as hyperlinks or integrated maps, users can access further information, organize their journeys, and engage deeply with the local culture, thereby enriching the overall visitor experience[20]. The utilization of text-to-video technology additionally enhances the accessibility and appeal of tourism content, but also contributes to the wider marketing of local destinations, hence facilitating a rise in tourism and economic development in these regions.

6. CONCLUSION

The establishment and implementation of the framework described in this study signify significant progress in the enhancement of a better-informed, inclusive, and actively involved Indian society. This creative initiative recognizes and considers the diverse language perspectives inside the country, aiming to not only translate the news but also adapt it, ensuring that it is easily understandable for a broad audience. The capacity of this kind of technology to convert printed news into captivating, visually striking videos is a clear indication of its transformational potential. These videos offer a captivating and informative experience that effectively engages the audience [15]. The mode of news dissemination is undergoing transformation, which has an opportunity to significantly enhance the total influence of information diffusion.

The consequences of this endeavour are significant, going beyond just connecting different languages to promote equal access to information. This initiative facilitates seamless comprehension and interaction with news material by strengthening speakers of regional languages. In addition, the Press's progressive stance in embracing innovation establishes an admirable benchmark for efficient communication tactics, especially in governmental settings. The advancement of Text-to-Speech technology overcomes linguistic obstacles in the distribution of information, enhancing communication and fostering social cohesion[18]. With the increasing prevalence of this state-of-the-art technology, it offers a captivating and immersive experience while also facilitating more accessible communication. This sets the stage for a future where information effortlessly crosses language barriers, fostering greater interconnectedness and unity among all members of society. Furthermore, it signifies a significant advancement towards the achievement of a society where comprehension and relationships surpass linguistic obstacles.

7. FUTURE DEVELOPMENT

The future potential of Text-to-Video technology is highly promising for progress and creativity. To enhance the capabilities and broaden the scope of this innovation, future research could focus on many critical areas:

7.1 Improved Multilingual Capabilities

Extending the range of Text-to-Video technology to include additional regional languages is a vital domain for future investigation. The expansion seeks to close linguistic barriers and appeal to a more extensive clientele. To expand the technical reach and application in India, it is necessary to incorporate more regional languages to cater to the heterogeneous linguistic terrain. The progress in speech recognition and natural language processing is going to have a substantial impact on faithfully reproducing various linguistic subtleties, accents of the region, and speech patterns. This advancement will greatly aid in the democratization of information access and ensure that folks of all languages may effortlessly interact with content.

7.2 Templates and Styles that can be customized according to individual preferences.

Expanding the text-to-video technology to include a wide range of customizable themes and styles will greatly enhance the user experience. Users can select the most appropriate theme or structure for their individual content or brand identification from a variety of prepared templates and visual styles. This feature enables modification without requiring advanced technological expertise. This feature will not only increase the flexibility of the video content development process but also improve brand coherence across different media platforms.

7.3 Feedback from many languages

The inclusion of a feature that allows users to provide feedback in any regional language of India is a substantial improvement. Enabling the submission of feedback in multiple regional languages caters to a wider range of users, fostering inclusivity and facilitating conversation. Enabling users to offer reviews in the language of their choice promotes increased engagement and guarantees that a wide range of viewpoints contribute to the enhancement of the platform. This strategy not only improves user involvement but also guarantees a more thorough and inclusive process for collecting input.

Data Availability Statement

Data may be available based on request.

Declarations

Conflict of Interests/Competing Interests

The authors declare that they have no conflict of interest.

Not received any funding for the proposed work.

Authors' contributions

NT took care of the review of literature and methodology, formal analysis, data collection and investigation, initial drafting, and statistical analysis. Also has supervised the overall project.

Acknowledgments

The authors gratefully acknowledge the Management of Acropolis Institute of Technology & Research, Indore, M.P., India for providing the facilities to carry out the research work.

8. REFERENCE

1. Ting et. al. A multi-embedding neural model for incident video retrieval, Pattern Recognition, 2022, 130.
2. Chengyuan Zhanget. Al. CNN-VWII: An efficient approach for large-scale video retrieval by image queries, Pattern Recognition Letters, 2019,123, 82-88.

3. Yi-Hua Tina Wang, Wan-Hsuan Huang Phase II monitoring and diagnosis of autocorrelated simple linear profiles, Computers & Industrial Engineering, 2017,112, 57-70.
4. Y. Gu et al. Supervised recurrent hashing for large-scale video retrieval, In Proceedings of the 2016 ACM Conference on Multimedia Conference, MM 2016, Amsterdam, The Netherlands, 2016,15–19.
5. Ali A, Schwartz I, Hazan T, Wolf L Video and text matching with conditioned embeddings. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision.2022, 1565–1574.
6. Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In EMNLP, 2014.
7. Amrani E, Ben-Ari R, Rotman D, Bronstein A Noise estimation using density estimation for self-supervised multimodal learning. In Proceedings AAAI Conference Artificial Intelligence, 2021,35,6644–6652.
8. Jesus Martin et. al., A comprehensive review of the video-to-text problem, Published: Artificial Intelligence Review,2022,55,4165–4239.
9. Anderson P, Fernando B, Johnson M, Gould S SPICE: semantic propositional image caption evaluation. ECCV. Springer Nature, 2016,382–398.
10. Dong J, Li X, Snoek CGM Predicting visual features from text for image and video caption retrieval. IEEE Trans Multimedia,2018, 20(12),3377–3388.
11. A. Eisenschtat and L. Wolf Linking image and text with 2-way nets, In Proceedings IEEE Conference Computer Vison Pattern Recognition, 2017, 4601-4611.
12. J. Gu, J. Cai, S. Joty, L. Niu and G. Wang Look imagine and match: Improving textual-visual cross-modal retrieval with generative models, IEEE Conference Computer Vision Pattern Recognition,2018,7181-7189.
13. Ruoyu Liu et. al. Modality-Invariant Image-Text Embedding for Image-Sentence Matching, ACM Transactions on Multimedia Computing, Communications, and Applications,2018,15 (1), 27, 1–19.
14. Andrej Karpathy, Armand Joulin, and Li F. Fei-Fei Deep fragment embeddings for bidirectional image sentence mapping, In NIPS. MIT Press, 2014,1889--1897.
15. Jurafsky, D., and Martin, J. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2008.
16. Akhtar, MS., Kumar, A., Ekbal, A., Bhattacharyya P. A hybrid deep learning architecture for sentiment analysis. In Proceedings of COLING 2016, the 26th international conference on computational linguistics: technical papers, 2016, 482–493.
17. Junyan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017,2242–2251.
18. Tanya Marwah, Gaurav Mittal, and Vineeth N Balasubramanian Attentive Semantic Video Generation Using Captions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017,1435–144.
19. Namrata Tapaswi Spellchecker for Sanskrit Sentences based on Morphological Analysis , Kiranavali , 2023, XIV (III-IV),ISSN 0975-4067.
20. Newton et. al. A systematic review on content-based video retrieval, Engineering Applications of Artificial Intelligence, on Computer Vision and Pattern Recognition 2020,90,1435–1443.