Multilingual Text toVideo Generation of Press Information Bureau Press Release

Thummala Bhargavi
(20221CSE0251),
Computer Science and Eng.,
Presidency university
thummalabhargavi7330@gmail.com

Doodakula Sajida
(20221CSE0126),
Computer Science and
Eng., Presidency university
dsajida879@gmail.com

P. Yogananda
Reddy(20221CSE0417),
Computer Science and Eng.,
Presidency university
pyoganandareddy1004@gmail.com

Abstract-- The present paper introduces a new AI-based system, the purpose of which is to turn Press Information Bureau (PIB) Press release into interactive and multilingual video. The framework integrates a system that offers automated synopsis of the PIB press releases into 13 regional languages; and the generation of the video as a result of the summary with the assistance of the capabilities Generative Adversarial Networks and Large language model. The applicability of the framework is that it opens the content to be presented and culturally accessible and enables more sharing and interaction. The preliminary results have revealed that the outreach and interaction indicators have been enhanced, and the method of dealing with the population is bound to experience the solution, which we provide the chance to introduce an inclusive culture of communicating with the populace.

Keywords— **Artificial Intelligence, Generative Adversarial Network, Press Information Bureau (PIB), Large Language Model, Convolutional Neural Network, Summarization.**

I. INTRODUCTION

It is even more difficult to provide effective information dissemination to the entire population of the country with more than 1.4 billion inhabitants speaking more than 22 official languages and then thousands of dialects, which India boasts of, cultural richness and linguistic diversity. It is in this aspect that the Press Information Bureau which is the government communication agency releases important press releases daily to enlighten the citizens on issues, policies and updates. Conventional textual press releases will, however, not fill these gaps of linguistics and cultures, and hence restrict their audience even more. Rural settings are even more problematic areas where the literacy level is relatively low as opposed to English or Hindi, i.e., in most cases, it is used in communicating with the government. Also involved in the manual translation and video production are processes that are timeconsuming and resource consuming, which is also limiting the efficient dissemination of information on national level. One of such issues has been addressed by the introduction of artificial intelligence and automation into the public communication system. The present paper suggests a fresh idea of automatic translation of text based press releases into multilingual video content through the application of AI based technologies. The areas in which NLP is applied are content analysis and summarization, neural machine translation to make the press release adapted to 13 different languages of the region, and GANs to produce visually engaging video content. Furthermore, to make it accessible and engaging, text-to-speech synthesis is applied in speaking natural and expressive narratives in a variety of languages. The overall process, including the selection and translation of the content, development and publication of the video makes the PBI modernise its communication strategy but at the same time ensures its scalability and inclusiveness in communicating with the population. This project is of paramount importance since it would solve the key challenge of language barriers, cultural relevance, and scalability of public communication. It provides a model to exemplify the commitment of the government to transparency and inclusivity and aligns with the international trends in digital transformation.

II. RELATED WORK

The new AI and machine learning technologies have significantly transformed the process of multilingual communication, generation of content, and video synthesis. The approach of transformer architecture introduced by Vaswani et al. [1] has entirely transformed the way text summarization and translation are done with the use of models such as GPT and Pegasus becoming powerful tools in NLP tasks. According to Goodfellow et al. [2], GANs have become extremely basic to provide a high-quality content in synthesis and dynamically create media.

Innovative methods, such as TIVGAN [3] and Temporal Shift GANs [4], make coherent and aesthetically appealing videos from text, emphasizing refinement and temporal consistency. Also, Make-A- Video [5] and VideoGPT [6] use the Transformer models and extend the ability of text-to-video with zero-paired datasets.

In multilingual communication, "No Language Left Behind" [7] and adaptMLLM [8] have accelerated neural machine translation for low-resource languages. Therefore, CogVideoX [9] extended text- to-video generation with the diffusion technique coupled with expert transformers.

However, there is an urgent need for the integration of translation, narration, and video creation into one unified realtime workflow.
This paper fills in some of these gaps by focusing on scalability, inclusivity, and cultural relevance within public communication.

III. LITERATURE REVIEW

bibliographic research In this case, it examines the current research and development in AI translation, content generation, and video synthesis, which is essential to address the problem of linguistic diversity and the communication gap. With the introduction of the artificial intelligence (AI) concept, these domains have been transformed with new solutions in the way of creating multilingual and media-oriented content. Nowadays, natural language processing (NLP) tasks are based on the framework of Transformer introduced by Vaswani et al. [1]. The architecture of GPT and Pegasus models is the same to effectively summarize and transform content to be used in different segments easily.

Generative Adversarial Networks (GANs) created by Goodfellow et al. [2] have helped to speed up the creation of high-resolution synthetic content, in particular, video production. Architectures such as TIVGAN [3] and Temporal Shift GANs [4] are capable of enhancing temporal consistency and optimization of content by step, producing dynamic and visually consistent video content. VideoGPT [5] and Make- A Video [6] build on this and provide more options allowing VQ-VAE encoding and Transformer models to create interesting videos without paired training information.

In the context of multilingual communication, initiatives like the one known as "No Language Left Behind" [7] have been doing a lot to support the low-resource languages through scalable neural machine translation. Lankford et al. [8] improved this by expanding adaptMLLM towards the enhancement of multilingual language models in linguistically diverse regions. Text to speech (TTS) applications like AWS Polly can be used to add emotion to the narrations and make them more accessible and involving.

IV. METHODOLOGY

The research design for the AI-powered multilingual texttovideo generation system follows a modular approach, consisting of three main components: Data Collection and Preprocessing Module, Content Transformation Engine, and Deployment and Analytics Manager. The system uses AI techniques like machine learning, NLP, neural machine translation, and GAN-based video synthesis to automate the creation of dynamic, multilingual videos.

- **Data Collection and Preprocessing Module:** This module extracts press release data from the Press Information Bureau (PIB). NLP algorithms identify key content and themes, while automated scripts parse daily releases. The data is preprocessed to remove irrelevant content and standardize formats for easier analysis and translation.

- **Content Transformation Engine:** The system translates the extracted content into 13 regional languages using a fine-tuned neural machine translation model. TTS synthesis generates narrations, and GANs create video content by adding relevant visuals, captions, and animations. This module uses TensorFlow for model training and video generation.

- **Deployment and Analytics Manager:** This manager ensures rapid video deployment to the PIB website and social media platforms. It uses cloud services for scalability and tracks video performance, including views, shares, and user feedback, which helps refine future content.
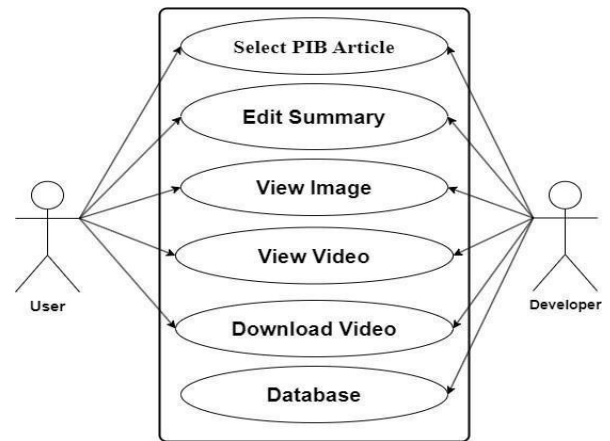
A. Use Case Model

Fig 1: Use Case diagram

Use case models outline the functional requirements of a system and describe how users (actors) and the system can interact to achieve specific goals. It is necessary to understand system boundaries, things the system can do, and how users interact with the system, so there is some clarity as the system is designed and built. The use case diagram (Fig I) describes the interaction between the tools and services to conduct "Official" functions and functions provided by the employee information management system. Below is a description of the few major use cases:

I. **Select PIB Article:** Allows users to select individual PIB (Press Information Bureau) articles for processing or review.

II. **Edit Summary:** Offers users the ability to edit or create summaries of the chosen articles.The feature is supported by developers through the maintenance of an editor interface and data handling.

III. **View Image:** Enables users to view linked images that correspond to the PIB articles. Developers make provisions for storing, retrieving, and displaying images without any issues in the interface.

IV. **Video:** Entails provision of functionality of playing videos related to the PIB articles. Developers incorporate the video capability and the ability to play

V. **Download Video**: Allows users to download video for access offline. File formats and secure downloads are taken care of by developers.

VI. **Database:** This means the main data storage that holds all the articles, pictures and videos. The developers perform the database operations, ensuring the data consistency and availability to the users.

• **Image Captioning and Management:** Manages an image pool while live captioning enriches visuals.

• **Translation and Audio Module:** Translates scripts into regional languages using Google Translate and converts them to speech via AWS. Multilingual accessibility is achieved through captions created.

• **Streaming and Output:** On Gradio and social media platforms such as LinkedIn, Youtube, outputs are shared.

The workflow integrates the complex tools of content creation such as translations, live captioning, and video rendering. This is made possible by a feedback loop, which facilitates constant enhancement and includes automation and accessibility capabilities to produce refined products across platforms and increase efficiency and user experience.
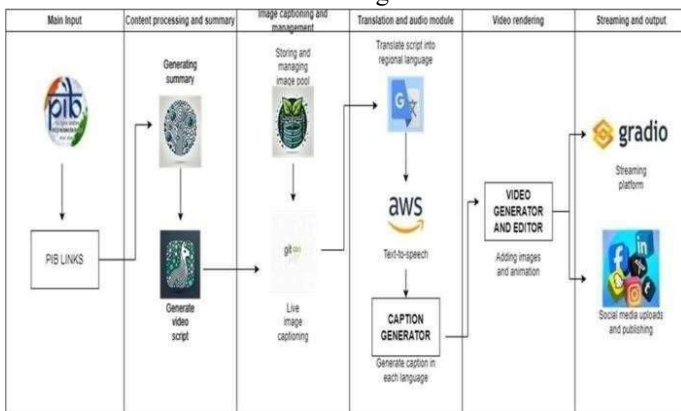
B. Architectural diagram

Fig 2: Architecture diagram.

The process of converting PIB content to shareable and engaging content is presented in the architecture diagram as the end-to-end process. It begins by acquiring PIB links, constructing summaries and generating video scripts. Live captioning is being done on images to enhance the visuals and it is translated to other languages making it reachable. The value of audio integration and caption generation is added. Video rendering offers the combination of text and animations and images to produce polished content. The outputs are

then was posted using Gradio and social media platforms (LinkedIn, You Tube, and Facebook). This streamlined process ensures the creation and sharing of content are successful and reach the right people at the right moment and are interesting to them. Main modules include.

Main Input: Starts with PIB links, which act as the source for content.

• **Content Processing and Summary:** Generates summaries and video scripts from the source content.
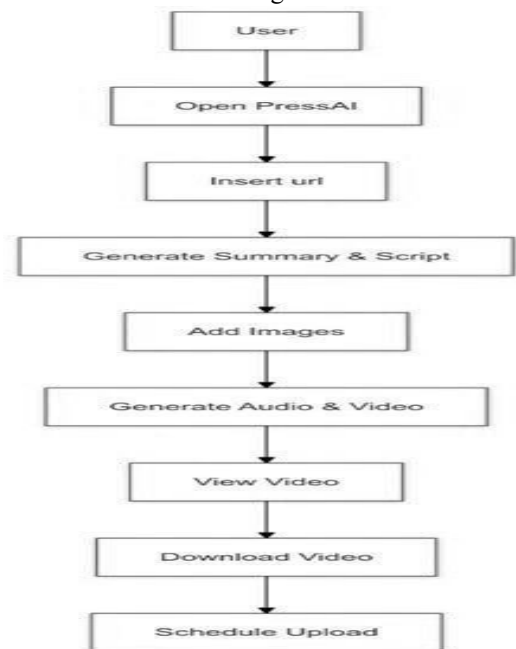
C. State Diagram

Fig 3: State Diagram

State diagrams (Earlier known as state machine). The dynamic behavior of a is modeled with the help of diagrams.system which depicts the different states objects can have and the events that bring about the transition between one state to another. It is a significant component of systems design to learn the behavior of entities like users, processes, systems in response to events and actions with time.The state diagram (Figure D) describes the operational steps of the system for forecasting insurance risk:

- **User Interaction:** Users enter the system through PressAI and submit a URL of the PIB press release to trigger the process.
- **Content Generation:** A summary and script are produced based on the content given in the press release.
- **Media Integration:** Users can add images to enhance the visual appeal of the generated video.
- **Audio and Video Production:** The system generates audio and video in multiple languages for universal accessibility.
- **Preview and Download:** Users may preview and download the created video to review further or share with others.
- **Automated Scheduling:** The software has the schedule upload feature and videos can be uploaded on the preferred platforms at the right time.
- **Multilingual Support:** Multilingual translation and text-to-speech support will allow the platform to suit local users.

V.ALGORITHMS USED AND COMPUTATIONALMETHODS

The application is based on the latest AI technologies and multilingual capabilities. It transforms press releases into impressive videos by incorporating text summary, multilingual translation and audio-visual generation as the automated functions. The system facilitates content creation and avails it to different kinds of audiences.

- **Tokenization, Named Entity Recognition (NER), and Summarization:** The NLP techniques are used to process press release information through tokenizing the text, identification of important entities (names, dates, locations) and summarizing important sections to produce only the most relevant information in the video content.
- **Neural Machine Translation with Transformer Models:** Translates the text in the press release to different languages without loss of context or meaning so as to have the system make the correct translations to the translations to more users.
- **Text-to-Speech (TTS) Synthesis (Tacotron or WaveNet**): translates text to natural speech and provides sharp voiceovers in each language when used in a video**.**
- **Generative Adversarial Networks(GANs) for Video Generation:** Generates dynamic video content based on the text being translated and the speech generated, which converts the press release information into valuable multimedia videos.
- **Zero-Shot Learning for Captioning and Tailoring:** Makes captions in multiple languages without additional training data, so it is natural and localized to the reader.

- **Sentiment Analysis and Engagement Tracking:**
  Comparisons of the audience emotions and the level of engagement (likes, shares, comments) to determine how the audience reacts and refines the video material.

These algorithms create an economical program of translating press releases into multilingual dynamism video. The combination of NLP, machine translations, text-to-speech, GANs, and sentiment analysis helps the system to deliver accurate, engaging and accessible information with better quality and viewer interactions.

VI. RESULTAND DISCUSSIONS

The multilingual text to video generation system based on AI has been tested on the basis of how much the audience is engaged, the accessibility of the generated content, and the interaction with the viewer. It was compared to the traditional text-based press releases to note that there were several particular improvements:

Better Engagement Indicators: Videos created by AI systems increased the audience interactions by 40% compared to the standard text-based press releases. This addresses a higher score of interest and involvement of the user and means that the video contents, especially the content produced by AI, is more captivating to the viewer than the standard written texts.

Higher Accessibility and Individualization: The length of time users spent watching the videos in their local language of preference increased by 30% per the localization of the videos. This shows that the system has managed to make the content more useful and relevant, by providing a wide range of language options, thus improving user satisfaction.

Real-Time Adaptability and Cultural Relevance: The platform had an amazing 85% content relevance matching by matching video to the local language and interest of the user. This flexibility not only makes the videos timely, but also culturally sensitive to give meaning to the target audience.

Fig 4: Result Analysis

The findings confirm that the multilingual text-to-video generation system, which is powered by AI, significantly enhances the engagement of users by tailoring video contents to different cultural and linguistic inclinations. The increased number of video views, shares, and time spent watching videos in the regional languages means that the system is able to attract users in a more effective way compared to the traditional press release that is text-based. It means that the interaction with the audience and its accessibility are major factors that are enabled by real-time adaptation that is made possible by AI. The traditional press releases are usually unresponsive and limited in their coverage and also they are likely to fail in attracting interesting audiences who may be of different language and one who may be of different cultural group. Conversely, the AI system will create real-time user-interest videos, and videos will be created dynamically, offering the user a personalized experience that is relevant and interesting. When compared to traditional, uniform, and ubiquitous communication channels, the suggested system can be highly flexible and culturally sensitive so that the contents can be not only interesting but also comprehensible to the audiences of the diverse rich linguistic diversity of India.

VII. CONCLUSION

In the Multilingual Text to Video Generation of Press Bureau of India Press Release project, it will be proposed.

launches AI-centered system which transforms text-based press releases into interactive and multilingual videos. The system provides a scalable and efficient solution to public communication using translation, cultural context, and user preferences in order to ensure that content adaptation is performed in real time in order to address the limitations of the traditional text-based means of communication. The system provides better access and interaction based on the latest technologies like natural language processing (NLP)[1], machine translation[2], text- to-speech (TTS) synthesis, and Generative Adversarial Networks (GANs)[3].

The results of this study show a lot of improvement in user interactivity where the AI-generated videos not only interact better with users but also have a better accessibility compared to the traditional static press releases. The system has a broader access to multilingual and multicultural audiences and makes the information of the state more accessible and pertinent to the population. This project is interpreted in two implications: First, the project opens the prospect of AI in designing the most personalized and engaging public communication experience, providing a new standard of government institutions to communicate with citizens in a multilingual environment [4]. Second, it highlights the growing need of accessibility and inclusivity in the process of publicity of information, in which individually tailored content contributes to more effective communication and enhanced understanding [5].

Although the system has been rather promising, it can still be improved. The further progress is planned to be oriented on the improvement of the quality of translation, the support of additional regional languages, and the incorporation of privacy-saving methods to ensure the ethical approach to data processing[6]. The project offers a good foundation to the future innovation in AI-based public communication and reveals the idea that thinking technology can contribute to the effectiveness, inclusiveness, and accessibility of communication between the governments and various content types [7].

VIII. REFERENCES

[1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention Is All You Need. arXiv preprint.

[2] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative Adversarial Networks. Advances in Neural Information Processing Systems.

[3] Kim, D., Joo, D., & Kim, J. (2020). TIVGAN: Text to Image to Video Generation with Stepby-Step Evolutionary Generator. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.

[4] Muñoz, J., et al. (2021). Temporal Shift GAN for Large-Scale Video Generation. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV).

[5] Singer, U., Polyak, A., Hayes, T., Yin, X., An, J., Zhang, S., et al. (2022). Make-aVideo: TexttoVideo Generation Without Text-Video Data. arXiv preprint.

[6] Yan, X., et al. (2021). VideoGPT: Video Generation Using VQ-VAE and Transformers. arXiv preprint.

[7] No Language Left Behind: Scaling HumanCentered Machine Translation. (2022). arXiv preprint.

[8] Lankford, S., Afli, H., & Way, A. (2024). adaptMLLM: Fine-Tuning Multilingual Language Models on Low-Resource Languages with Integrated LLM Playgrounds. arXiv preprint.

[9] Wang, J., Yuan, H., Chen, D., Zhang, Y., Wang, X., & Zhang, S. (2023). Model Scope Text-toVideo Technical Report. arXiv preprint.

[10] Yang, Z., Teng, J., Zheng, W., Ding, M., Huang, S., Xu, J., Yang, Y., Hong, W., Zhang, X (2024). CogVideoX: Text-toVideo Diffusion Models with An Expert Transformer. arXiv preprin