



Text to Video of Various PIB Press Releases Using Artificial Intelligence in English and Regional Languages

¹Mr. N. Lakshmi Narayana, ²G. Sai Sri Bhavya, ³Harshini Hechina, ⁴Immadisetty Yugandhar Prasad

¹Assistant Professor, Dept of Computer Science and Engineering, St. Ann's College of Engineering and Technology, Chirala-523187, India.

^{2,3,4}U. G Student, Dept of Computer Science and Engineering, St. Ann's College of Engineering and Technology, Chirala-523187, India.

ABSTRACT- *The project "Text to Video of Various PIB Press Releases Using Artificial Intelligence in English and Regional Languages" is an AI-based system that automatically converts Press Information Bureau (PIB) press releases into informative videos in both English and regional languages by integrating text, visuals, voice narration, and subtitles to make government information more engaging and accessible. Existing PIB press releases are primarily text-based, making them difficult for many users to read and understand, while manual video creation is time-consuming, costly, and requires technical expertise, resulting in delays in information dissemination. To overcome these challenges, the proposed system automates the text-to-video generation process using Artificial Intelligence by summarizing press releases, translating the content into regional languages, generating relevant visuals, producing natural voice narration, and creating synchronized*

videos with subtitles. The project is developed using Python and employs Natural Language Processing, Deep Learning, Text-to-Speech, Neural Machine Translation, and Generative Artificial Intelligence. It utilizes BART for text summarization, MarianMT for multilingual translation, CLIP and BLIP for image-text understanding, and Coqui TTS for speech synthesis. The proposed system enables faster dissemination of multilingual government information, improves accessibility, reduces manual effort, and enhances public awareness through accurate and engaging AI-generated videos.

KEYWORDS- *Natural Language Processing, Artificial Intelligence, Deep Learning, Python, Neural Machine Translation, Text-To-Speech*

INTRODUCTION

Artificial Intelligence has significantly improved the way information is created, processed, and delivered. Government



organizations regularly publish press releases to communicate important policies, schemes, and announcements to the public, but these releases are usually available only in text format, making them less engaging and difficult for many users to understand. This project converts Press Information Bureau press releases into informative videos in English and regional languages by combining relevant visuals, natural voice narration, and subtitles. The system uses Natural Language Processing for text processing, Deep Learning for intelligent content generation, Neural Machine Translation for multilingual translation, Text-to-Speech technology for voice narration, and Generative Artificial Intelligence for creating visual content

LITERATURE SURVEY

Artificial Intelligence has gained significant attention in recent years due to its ability to automate complex tasks and generate intelligent multimedia content. Text-to-video generation has emerged as an important research area that combines Natural Language Processing, Deep Learning, Computer Vision, and speech synthesis to transform textual information into engaging videos. Government organizations, including the Press Information Bureau, regularly publish press releases to communicate important announcements; however, these releases

are predominantly text-based and may not effectively reach audiences with different language preferences or limited reading abilities. Recent advancements in multilingual translation, text summarization, image-text understanding, and Text-to-Speech technologies have enabled the automatic generation of videos from textual content. Existing research demonstrates that integrating these Artificial Intelligence techniques can significantly improve the accessibility, efficiency, and dissemination of information. Therefore, the proposed system leverages these technologies to generate multilingual videos from Press Information Bureau press releases, making official government information more understandable, engaging, and accessible to a wider audience.

RELATED WORK

Several researchers have developed Artificial Intelligence-based text-to-video generation systems to improve information accessibility and automate multimedia content creation. Earlier approaches mainly converted text into simple videos using predefined templates and rule-based methods, offering limited automation and multilingual support. Recent advancements in Natural Language Processing, Deep Learning, Neural Machine Translation, Text-to-Speech, and Computer Vision have



enabled automatic text summarization, translation, visual generation, and voice narration. However, most existing systems focus on general content and do not specifically generate multilingual videos from Press Information Bureau press releases. The proposed system addresses this limitation by automatically creating informative videos in English and regional languages from official government press releases.

EXISTING SYSTEM

The existing system mainly provides Press Information Bureau press releases in text format through websites and digital documents. Users must read the complete content to understand government announcements, which can be time-consuming and less engaging. Manual conversion of press releases into videos requires significant human effort, editing skills, and additional time. Most existing text-to-video systems support only general content and offer limited multilingual capabilities, making them less suitable for the rapid and accessible dissemination of official government information.

PROPOSED SYSTEM

The proposed system automatically converts Press Information Bureau press releases into informative videos in both English and regional languages using

Artificial Intelligence. It summarizes the press release, translates the content into the selected language, generates relevant visuals, creates natural voice narration, and synchronizes subtitles to produce a complete video. By automating the entire text-to-video generation process, the system reduces manual effort, improves accessibility, enables faster dissemination of government information, and provides an engaging way for users to understand official announcements.

SYSTEM ARCHITECTURE

The architecture converts Press Information Bureau (PIB) press releases into informative videos in English and regional languages using Artificial Intelligence. Initially, the user uploads a PIB press release through the application. The uploaded text is processed using Artificial Intelligence techniques, where BART performs text summarization and MarianMT translates the summarized content into the selected regional language. The translated text is then converted into natural voice narration using Coqui TTS. Simultaneously, relevant images are generated or selected using CLIP and BLIP to match the content. All processed data, including user details, text, audio files, and generated videos, is stored in the database for efficient management. Finally, the system combines images, voice narration,



subtitles, and background elements to generate a complete multilingual video in MP4 format, which can be viewed, downloaded, and shared by the user.

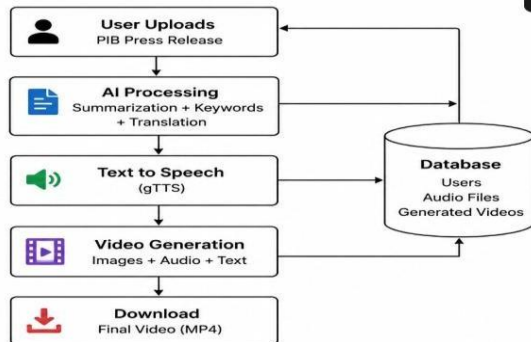


Fig 1.1: System Architecture

METHODOLOGY DESCRIPTION

The methodology with uploading a Press Information Bureau press release into the system. The input text is preprocessed and summarized using the BART model to extract key information. The summarized content is translated into the selected regional language using the MarianMT model. Relevant visual content is generated using CLIP and BLIP, while Coqui Text-to-Speech converts the translated text into natural voice narration.

The system synchronizes images, audio, and subtitles to generate a complete multilingual video. Finally, the generated video is stored in the database and made available for viewing, downloading, and sharing, ensuring faster and more accessible dissemination of government information.

RESULTS AND DISCUSSIONS

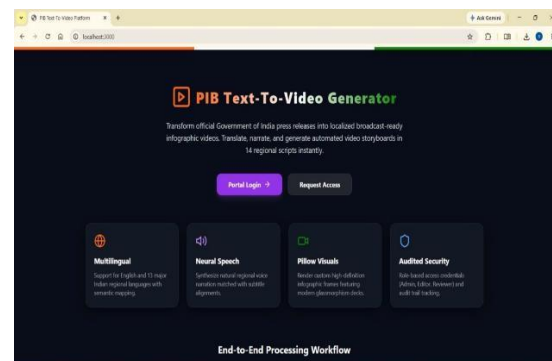


Fig 1.2: Home page

The Home Page serves as the main entry point of the PIB Text-to-Video Generator application. It introduces the purpose of the system, which is to transform official Government of India PIB press releases into AI-generated videos. The page provides options for Portal Login and Request Access, allowing users to securely access the platform.

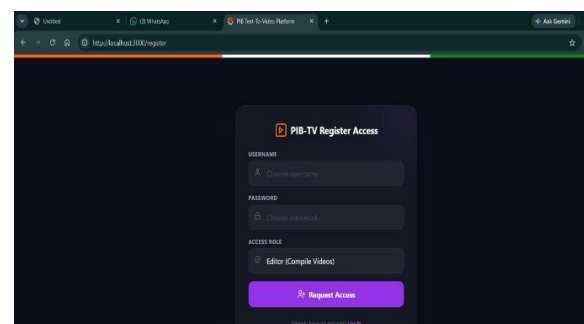


Fig 1.3: Sign Up page

The Sign Up page is the entry point for new users to access the PIB Text-to-Video platform. It allows users to create an account by entering a unique username, password, and selecting an appropriate access role.

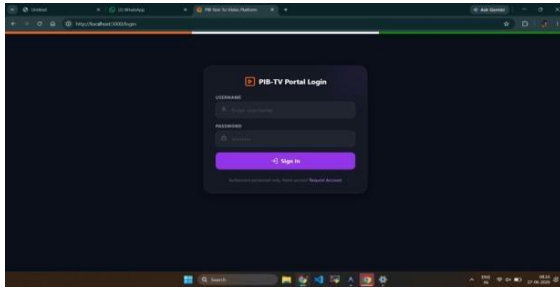


Fig 1.4: Login Page

The Login Page provides secure access to the PIB Text-to-Video Generator application. Authorized users can log in by entering their registered username and password.

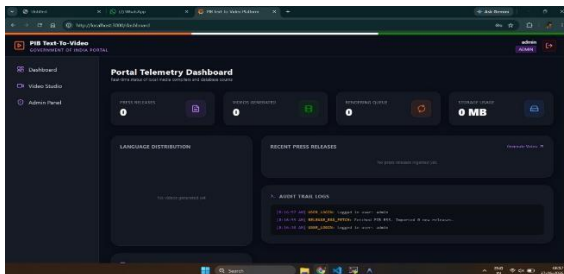


Fig 1.5: Dashboard page

The Dashboard serves as the main workspace of the PIB Text-to-Video Generator after successful user login.

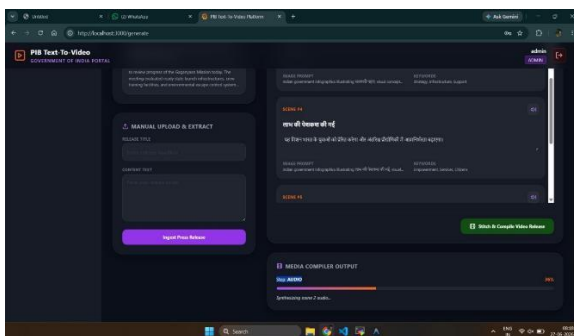


Fig 1.6: Video Generation page

The Video Generation Page is the core module of the PIB Text-to-Video Generator

application, where users convert PIB press releases into AI-generated videos.

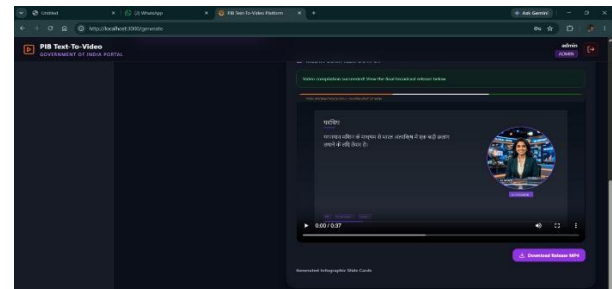


Fig 1.7: Generated video in Hindi language

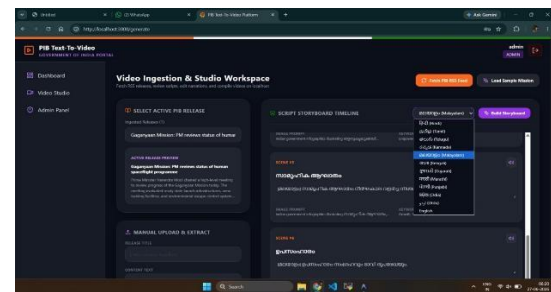


Fig 1.8: Availability of 12 regional languages

CONCLUSION

The "Text to Video of Various PIB Press Releases Using Artificial Intelligence in English and Regional Languages" project demonstrates the effective use of Artificial Intelligence to convert textual government press releases into informative multilingual videos. By combining text processing, language translation, voice narration, and visual content generation, the system improves accessibility and enhances the communication of official information. The technology reduces manual effort, saves time, and makes government



announcements easier to understand for people speaking different languages. Overall, the project highlights the importance of Artificial Intelligence in delivering information through engaging and accessible multimedia content.

FUTURE SCOPE

The field of Artificial Intelligence-based text-to-video generation has great potential for future advancements. The technology can be extended to convert news articles, educational content, healthcare information, and government documents into multilingual videos automatically. Future improvements may include support for more regional languages, realistic visual generation, natural voice narration, and real-time video creation. These advancements will enhance accessibility, improve digital communication, and enable faster dissemination of information across diverse audiences.

REFERENCES

- [1] D. P. Harini, "Electronic Health Record Stage Identification using Principal Component Analysis," *Machine Intelligence Research*, vol. 18, no. 01, pp. 864- 873, 17 8 2024.
- [2] D. P. Harini, "Analyze and Predict of Human Cyber Attackers using Artificial Neural Network," *Journal of Basic Science and Engineering*, vol. 21, no. 1, pp. 1529 - 1536, 8 7 2024.
- [3] D. P. Harini, "Scalable and Secure Sharing of Client Medical Records in Cloud," *International Journal Of Computer Sc*, vol. 4, no. 4, pp. 161-163, 2013.
- [4] J. Ho, T. Salimans, W. Chan, *et al.*, "Imagen Video: High Definition Video Generation with Diffusion Models," Google Research, 2022.

<https://arxiv.org/abs/2210.02303>

- [5] J. Kondratyuk, D. Ross, *et al.*, "CogVideo: Large-scale Pretraining for Text-to-Video Generation via Transformers," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2023.

<https://arxiv.org/abs/2205.15868>

- [6] OpenAI, "GPT-4 Technical Report," 2023. <https://arxiv.org/abs/2303.08774>

- [7] H. Touvron, L. Martin, K. Stone, *et al.*, "Llama 2: Open Foundation and Fine-Tuned Chat Models," Meta AI, 2023.

<https://arxiv.org/abs/2307.09288>

- [8] Google DeepMind, "Gemini: A Family of Highly Capable Multimodal Models," 2023.

<https://arxiv.org/abs/2312.11805>

- [9] T. B. Brown, B. Mann, N. Ryder, *et al.*, "Language Models are Few-Shot



Learners," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020.

<https://arxiv.org/abs/2005.14165>

[10] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proc. NAACL-HLT*, 2019.

<https://arxiv.org/abs/1810.04805>

[11] Hugging Face Team, "Transformers Documentation," 2024.

<https://huggingface.co/docs/transformers>

[12] Google Cloud, "Cloud Translation API Documentation," 2024.

<https://cloud.google.com/translate/docs>

[13] Google Cloud, "Cloud Text-to-Speech API Documentation," 2024.

<https://cloud.google.com/text-to-speech/docs>

[14] OpenCV Team, "OpenCV: Open Source Computer Vision Library Documentation," 2024. <https://opencv.org/>

[15] Press Information Bureau, Government of India, "Official Press Information Bureau Portal," 2025.

<https://pib.gov.in/>

[16] Shashank, A. (2025). Self-Healing Data Pipelines for Enhanced Reliability: A Paradigm Shift in Enterprise Data Management. *Journal of Computer Science and Technology Studies*, 7(8), 1097-1104.