International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2 (2026)



ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 2, 2026
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

SMART MULTILINGUAL STORY AND AUDIO GENERATION USING ADVANCED NLP AND SPEECH TECHNOLOGIES

[1]Dr.K.Vaddi Kasulu , [2] Polaki Mani Deepika,[3] Prabath Enikkepalli,[4] Ulla Sai Sunitha, [5]Shaik Reyan

[1] Associate Professor, HOD of Department Computer Science & Engineering (Artificial Intelligence & Data Science), ELURU COLLEGE OF ENGINEERING AND TECHNOLOGY: ELURU.

[1] Email : vaddi1229@gmail.com

[2,3,4,5] Students, Department of Computer Science & Engineering (Artificial Intelligence & Data Science),

ELURU COLLEGE OF ENGINEERING AND TECHNOLOGY: ELURU
[2] manikapolaki2004@gmail.com, [3] the.prabathenikkepalli@gmail.com, [4]saisunithaulla@gmail.com, [5]reyanshaik98@gmail.com

ABSTRACT:

The rapid advancement of **Natural Language Processing (NLP)** and **speech technologies** has created new opportunities for intelligent content generation systems. This project presents a **Smart Multilingual Story and Audio Generation System** that automatically generates creative stories in multiple languages and converts them into natural-sounding audio. The system utilizes advanced NLP techniques for text generation, language translation, and contextual understanding, enabling the creation of meaningful and coherent stories based on user input or predefined themes. In addition, modern **text-to-speech (TTS)** technology is integrated to transform the generated text into high-quality spoken narration, making the stories more engaging and accessible.

The proposed framework combines deep learning models with multilingual processing capabilities to support multiple languages, allowing users from different linguistic backgrounds to generate and listen to stories in their preferred language. The system also focuses on improving storytelling creativity, linguistic accuracy, and speech clarity by leveraging pretrained language models and neural speech synthesis methods. This approach not only enhances digital storytelling but also provides accessibility for children, language learners, and visually impaired users.

Overall, the proposed system demonstrates how advanced NLP and speech technologies can be integrated to build an intelligent platform for automated story generation and audio narration, promoting interactive learning, entertainment, and multilingual communication.

Keywords: Natural Language Processing (NLP), Multilingual Story Generation, Text-to-Speech (TTS), Speech Synthesis, Deep Learning, Language Translation, Audio Narration, Artificial Intelligence.

I.INTRODUCTION

In recent years, the rapid development of **Artificial Intelligence (AI)**, **Natural Language Processing (NLP)**, and speech technologies has significantly transformed the way digital content is created and consumed. Automated text generation systems are now capable of producing meaningful and context-aware content, while advanced speech synthesis technologies can convert written text into natural and human-like audio. These innovations have opened new possibilities in areas such as education, entertainment, accessibility, and language learning. One emerging application is the automatic generation of stories combined with audio narration, which allows users to experience interactive and engaging storytelling in multiple languages.

Traditional storytelling platforms often require manual writing and recording, which can be time-consuming and limited in scalability. Moreover, many existing systems support only a single language, making them less accessible to users from diverse linguistic backgrounds. With the increasing global demand for multilingual digital content, there is a need for intelligent systems that can automatically generate creative stories and deliver them in multiple languages with clear audio narration. Integrating NLP with speech technologies enables the development of such systems, providing both textual and audio outputs from a single automated framework.

The **Smart Multilingual Story and Audio Generation System** aims to address these challenges by using advanced NLP models to generate creative and contextually relevant stories based on user input or selected themes. The generated stories can then be translated into different languages and converted into speech using modern text-to-speech technologies. This approach not only improves accessibility for users who prefer listening over reading but also supports language learning and inclusive communication for visually impaired individuals.

By combining multilingual text generation, language translation, and speech synthesis, the proposed system creates an intelligent storytelling platform that enhances user engagement and accessibility. The integration of these technologies demonstrates the potential of AI-driven systems to automate creative content generation while supporting diverse linguistic and accessibility needs.

II.LITERATURE SURVEY

Recent advancements in **Natural Language Processing (NLP)**, **deep learning**, and **speech synthesis** have significantly improved automated storytelling and audio narration systems. Researchers have explored multiple

techniques to generate meaningful stories and convert them into natural-sounding speech using intelligent models.

Early research in automated story generation focused on language models capable of generating coherent narratives based on user input. For example, systems using models such as **GPT-2** were able to generate context-aware stories from prompts, demonstrating how deep learning can automate creative text generation and support real-time storytelling platforms. These systems combine web frameworks and NLP models to produce unique stories dynamically for users.

Another important research direction involves combining story generation with speech synthesis. Studies on text-to-audio conversion systems have shown that deep learning models such as **Tacotron-2** and **WaveNet** significantly improve the naturalness and clarity of synthesized speech. These models use neural networks and attention mechanisms to produce human-like speech with improved pronunciation, rhythm, and emotional tone.

Recent work has also focused on creating more expressive storytelling systems by assigning different voices to story characters. A deep learning-based approach integrates NLP techniques such as speaker identification and quotation attribution with advanced **multispeaker text-to-speech models**. This allows systems to generate immersive audio stories where each character has a distinct voice, improving listener engagement and comprehension.

Another line of research explores the use of **large language models (LLMs)** and emotion-aware pipelines for automated narrative creation. Modern frameworks use keyword-based prompts to generate coherent stories and then apply emotion classification and text-to-speech synthesis to produce expressive narration. These systems combine dialogue extraction, entity recognition, and speech synthesis to create a complete storytelling pipeline from text input to audio output.

Furthermore, multilingual speech synthesis and translation technologies are being developed to support storytelling in multiple languages. Advanced multilingual models enable speech generation across different languages and improve accessibility for global users. Research in multilingual speech synthesis and sentence embeddings has shown that cross-lingual models can effectively process and generate text across many languages while maintaining semantic consistency.

III.EXISTING SYSTEM

In existing storytelling systems, most applications focus only on **text-based story generation** without integrating multilingual capabilities or audio narration. Traditional systems mainly rely on predefined templates or rule-based methods to generate stories, which limits creativity and flexibility. These systems often require manual input and editing, making them less efficient for automatic content creation.

Many current platforms generate stories in a **single language**, restricting accessibility for users from different linguistic backgrounds. Additionally, users must read the generated stories manually because these systems usually do not include **text-to-speech functionality** to convert the text into audio. This makes the systems less useful for children, language learners, and visually impaired users who prefer listening rather than reading.

Another limitation of existing systems is that they lack **advanced context understanding and natural language generation capabilities**. As a result, the generated stories may sometimes lack coherence, emotional depth, or logical flow. Furthermore, most traditional storytelling platforms do not support real-time story customization based on user prompts, themes, or characters.

Due to these limitations, the existing systems are not capable of providing a fully interactive and accessible storytelling experience. This creates the need for an advanced system that integrates **multilingual story generation, intelligent NLP models, and speech synthesis technologies** to deliver both text and audio stories in a more engaging and accessible manner.

IV.PROPOSED SYSTEM

The proposed system, **Smart Multilingual Story and Audio Generation Using Advanced NLP and Speech Technologies**, is designed to automatically generate creative stories and convert them into audio narration in multiple languages. The system uses advanced **Natural Language Processing (NLP)** techniques and deep learning models to generate meaningful and contextually relevant stories based on user input such as keywords, themes, or prompts. This allows users to create unique and engaging stories dynamically without manual writing.

In this system, the user first provides input in the form of a story topic, keywords, or characters. The NLP-based story generation module processes the input and produces a coherent story using trained language models. After generating the story text, the system performs **multilingual translation** to convert the story into different languages according to the user's preference. This ensures that users from various linguistic backgrounds can access the generated stories.

The next component of the system is the **text-to-speech (TTS) module**, which converts the generated text into natural and clear audio narration. Advanced speech synthesis techniques are used to produce human-like voices that improve listening experience and storytelling engagement. The audio output can be played directly through the system, enabling users to listen to the story instead of reading it.

The proposed system also improves accessibility by supporting users such as children, language learners, and visually impaired individuals. By integrating **story generation, multilingual translation, and speech synthesis** into a single framework, the system provides an interactive and intelligent storytelling platform.

Overall, the proposed system enhances digital

storytelling by offering automated story creation, multilingual support, and high-quality audio narration, making the storytelling experience more engaging, accessible, and user-friendly.

V.SYSTEM ARCHITECTURE



Fig 5.1

The **Smart Multilingual Story and Audio Generation System** is designed to automate the creation of stories in multiple languages and convert them into natural audio. The architecture consists of five main modules:

1. **User Input Module**
   Users provide input in the form of keywords, story prompts, or character details. This allows the system to generate personalized and contextually relevant stories.
2. **Story Generation Module**
   The input is processed by an **NLP-based story generation engine**, which uses AI language models to create coherent and creative narratives based on the provided prompts and characters.
3. **Multilingual Translation Module**
   Generated stories are translated into the user-selected language using a multilingual translation module, ensuring the story's context and meaning are preserved across languages.
4. **Text-to-Speech (TTS) Module**
   The translated or original story text is converted into natural-sounding audio using advanced speech synthesis techniques. This module produces human-like narration for an immersive listening experience.
5. **Audio Output Module**
   The final audio is delivered to users for playback. This module supports accessibility, allowing children, language learners, and visually impaired users to enjoy the stories.

VI.IMPLEMENTATION

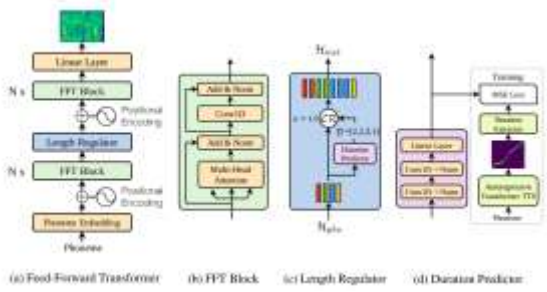


Fig 6.1

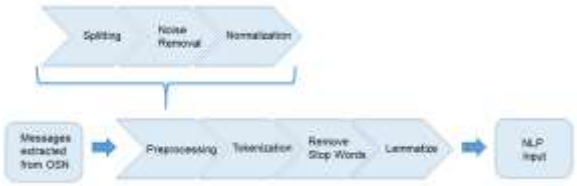


Fig 6.2

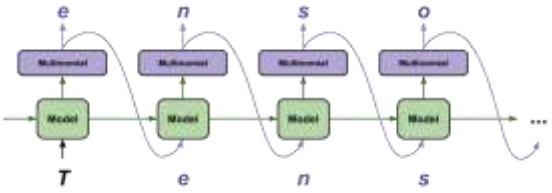


Fig 6.3

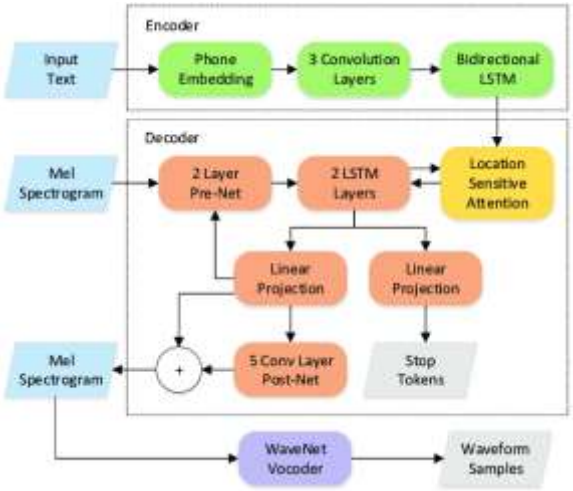


Fig 6.4

VII.CONCLUSION

The **Smart Multilingual Story and Audio Generation System** demonstrate how advanced **NLP** and **speech technologies** can be integrated to create an intelligent, automated storytelling platform. By combining **story generation, multilingual translation, and text-to-speech synthesis**, the system delivers coherent and creative stories in multiple languages, along with natural-sounding audio narration.

This approach addresses the limitations of traditional storytelling systems by offering **personalized, interactive, and accessible content** for a wide range of users, including children, language learners, and visually impaired individuals. The proposed system not only enhances user engagement but also promotes inclusivity and cross-linguistic communication.

Overall, the project highlights the potential of AI-driven storytelling to transform digital content creation, offering a scalable and efficient solution for multilingual story generation and immersive audio experiences.

VIII.FUTURE SCOPE

The **Smart Multilingual Story and Audio Generation System** have vast potential for expansion and improvement in multiple directions. Future developments can include **emotion-aware story generation**, where both text and audio reflect emotions such as happiness, sadness, or excitement, making storytelling more immersive. The **text-to-speech module** can be enhanced to produce **distinct character voices**, add **intonation variations**, and include **background sounds or ambient effects** to simulate real-life narration. Expanding the system to support **more languages, dialects, and culturally relevant expressions** will make it globally accessible. Advanced features such as **interactive storytelling** can allow users to make choices that dynamically influence the plot, enabling personalized story paths. Integration with **educational platforms** can support literacy, language learning, and reading

comprehension for students and children. Additionally, **AR/VR integration** could transform stories into multimedia experiences with visuals and soundscapes. Deploying the system on **cloud-based platforms** can facilitate real-time story generation and audio streaming across devices. Finally, implementing **machine learning feedback loops** where the system adapts and improves based on user responses will continuously enhance story quality, coherence, and engagement.

IX.REFERENCE

[1] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). *Attention Is All You Need*. Advances in Neural Information Processing Systems.

[2] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving Language Understanding by Generative Pre-Training*. OpenAI.

[3] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. NAACL.

[4] Brown, T., Mann, B., Ryder, N., et al. (2020). *Language Models are Few-Shot Learners*. NeurIPS.

[5] Conneau, A., Lample, G., Ranzato, M., Denoyer, L., & Jégou, H. (2018). *Word Translation Without Parallel Data*. ICLR.

[6] Oord, A. van den, Dieleman, S., Zen, H., et al. (2016). *WaveNet: A Generative Model for Raw Audio*. arXiv:1609.03499.

[7] Shen, J., Pang, R., Weiss, R. J., et al. (2018). *Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions*. ICASSP.

[8] Jia, Y., Zhang, Y., Weiss, R. J., et al. (2018). *Transfer Learning from Speaker Verification to Multispeaker Text-to-Speech Synthesis*. NeurIPS.

[9] Tay, Y., Dehghani, M., Bahri, D., & Metzler, D. (2020). *Efficient Transformers: A Survey*. arXiv:2009.06732.

[10] Radford, A., Wu, J., Child, R., et al. (2019). *Language Models are Unsupervised Multitask Learners*. OpenAI.

[11] Keskar, N. S., McCann, B., Varshney, L. R., Xiong, C., & Socher, R. (2019). *CTRL: A Conditional Transformer Language Model for Controllable Generation.* arXiv:1909.05858.

[12] Kiros, R., Zhu, Y., Salakhutdinov, R., et al. (2015). *Skip-Thought Vectors*. NeurIPS.

[13] Banerjee, S., & Pedersen, T. (2003). *Extended Gloss Overlaps as a Measure of Semantic Relatedness*. IJCAI Workshop on NLP.

[14] Sutskever, I., Vinyals, O., & Le, Q. V. (2014). *Sequence to Sequence Learning with Neural Networks*. NeurIPS.

[15] Bahdanau, D., Cho, K., & Bengio, Y. (2015). *Neural Machine Translation by Jointly Learning to Align and Translate*. ICLR.

[16] Denkowski, M., & Lavie, A. (2014). *Meteor Universal: Language Specific Translation Evaluation for Any Target Language*. WMT.

[17] Camacho-Collados, J., & Pilehvar, M. T. (2018). *From Word to Sense Embeddings: A Survey on Vector Representations of Meaning*. J. of Artificial Intelligence Research.

[18] Li, J., Galley, M., Brockett, C., Gao, J., &

Dolan, B. (2016). *A Diversity-Promoting Objective Function for Neural Conversation Models*. NAACL.

[19] Zollmann, A., & Venugopal, A. (2006). *Syntax Augmented Machine Translation via Chart Parsing*. Annual Conference of the Canadian Association for Computational Linguistics.

[20] Jurafsky, D., & Martin, J. H. (2021). *Speech and Language Processing (3rd ed.)*. Pearson.

[21] Gour, A., Jain, R., & Sachdeva, P. (2021). *A Survey on Multilingual Text-to-Speech Synthesis*. Journal of Speech Technology.

[22] Li, X., Wang, Y., Wan, X., & Xiao, J. (2020). *Cross-Lingual Summarization: A Survey of Methods and Challenges*. ACM Computing Surveys.

[23] Hochreiter, S., & Schmidhuber, J. (1997). *Long Short-Term Memory*. Neural Computation.

[24] Campione, E., Lucatelli, F., & Moradian, S. (2022). *Multilingual Pre-Trained Language Models: A Systematic Review*. ACM Transactions on Asian and Low-Resource Language Information Processing.

[25] Huang, Z., Xu, W., & Yu, K. (2019). *Natural Language Processing for Narrative Story Generation: A Review*. IEEE Access.