International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

AI-Based Multilingual Story Generation and Speech Synthesis System

J.Soumya[1], G.Rajini[2]
[1]MCA Student , Dept of MCA , Ellenki College of Engineering and Technology , Hyderabad
[2]Assistant Professor , Dept of MCA , Ellenki College of Engineering and Technology , Hyderabad

***Abstract*— Storytelling plays an important role in education, creativity, and entertainment, but writing stories manually requires time, imagination, and language proficiency. This project presents an AI-Powered Multilingual Story Generation and Speech Synthesis System that automatically generates meaningful stories from user-provided keywords. The system allows users to register, log in, select their preferred language, and generate stories in languages such as English, Hindi, Odia, and Telugu. It is developed using Python, Django, and MySQL, while the Gemini AI model is used to understand the input keywords and produce coherent, grammatically correct, and context-aware stories. In addition to text generation, the system converts the generated stories into speech, providing a more engaging and accessible experience for users. The proposed system reduces manual effort, supports multilingual communication, and enhances storytelling through artificial intelligence. It offers a simple, interactive, and efficient platform suitable for educational, learning, and entertainment applications.**

***Keywords* — Artificial Intelligence (AI), Natural Language Processing (NLP), Multilingual Story Generation, Gemini AI, Transformer Model, Text-to-Speech (TTS), Speech Synthesis, Django, Python, MySQL.**

I. INTRODUCTION

Recent advancements in Artificial Intelligence (AI) and Natural Language Processing (NLP) have significantly transformed the way digital content is created and consumed. Story generation systems have evolved from simple template-based approaches to intelligent models capable of understanding user input and producing meaningful, coherent, and context-aware stories. At the same time, multilingual processing has become increasingly important because users prefer to interact with applications in their native languages. Speech synthesis technology further enhances accessibility by enabling generated text to be converted into natural speech, making storytelling more interactive and beneficial for children, language learners, and visually impaired users. These developments have encouraged the integration of multilingual language processing and speech technologies into modern AI applications, providing a richer and more inclusive user experience [1], [2], [3].

Although several story generation applications are available, many existing systems provide limited functionality. Most platforms either support only a single language or require users to manually create and edit stories. Such systems often fail to understand the context of user-provided keywords and generate outputs that lack creativity and consistency. In addition, the absence of integrated speech synthesis limits accessibility for users who prefer listening instead of reading. Recent research in multilingual neural networks and speech representation learning has demonstrated that advanced deep learning models can effectively capture semantic information across multiple languages, thereby improving the quality and consistency of generated content and multilingual speech processing [4], [5], [6], [7].

To address these limitations, this work presents an AI-Powered Multilingual Story Generation and Speech Synthesis System that automatically generates meaningful stories from user-provided keywords and converts them into speech. The proposed system is developed using Python, Django, and MySQL, while the Gemini Transformer model is employed to understand the semantic relationship between keywords and produce grammatically correct and contextually relevant stories. Users can register securely, choose their preferred language, enter story keywords, and obtain creative stories in multiple languages such as English, Hindi, Odia, and Telugu. The generated stories can also be narrated through a speech synthesis module, providing an engaging and user-friendly storytelling experience [8], [9], [10], [11].

The integration of multilingual AI models, Transformer-based language understanding, and speech synthesis makes the proposed system suitable for educational, entertainment, and

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

language-learning applications. By combining intelligent story generation with multilingual support and voice narration, the system reduces manual effort while improving accessibility and user engagement. Furthermore, the use of modern deep learning techniques ensures better language understanding, coherent story generation, and scalable performance across different languages. This work demonstrates how recent advances in multilingual language modeling and end-to-end speech technologies can be effectively utilized to build an efficient and interactive storytelling platform for diverse users [12], [13], [14], [15].

II. RELATED WORK

"Network Architectures for Multilingual Speech Representation Learning," T. Sercu, G. Saon, J. Cui, B. Ramabhadran, B. Kingsbury, and A. Sethy [4]

This study proposed advanced deep neural network architectures for learning language-independent speech representations across multiple languages. The authors focused on developing shared feature extraction techniques that improve multilingual speech recognition while reducing the need for language-specific models. By training the network on speech data from different languages, the proposed architecture effectively captured common acoustic and linguistic patterns, resulting in improved recognition accuracy and better generalization. The research demonstrated that multilingual representation learning enables knowledge sharing among languages, making speech recognition systems more scalable and efficient. The findings highlighted the importance of deep learning for multilingual applications and inspired the development of intelligent systems capable of processing multiple languages. This work provides a strong foundation for AI-based multilingual applications, including multilingual story generation and speech synthesis, where language-independent representations improve both text understanding and speech generation quality.

"Knowledge Distillation Across Ensembles of Multilingual Models for Low-Resource Languages," J. Cui, B. Kingsbury, B. Ramabhadran, et al. [5]

This paper presented a knowledge distillation approach for improving multilingual speech recognition in low-resource languages. The authors proposed transferring knowledge from large ensemble models to smaller and computationally efficient student models without significantly reducing performance. The approach enabled the smaller models to learn generalized multilingual representations while requiring fewer computational resources during inference. Experimental evaluations demonstrated noticeable improvements in recognition accuracy for languages with limited training data. The study emphasized that knowledge sharing among multilingual models enhances learning efficiency and reduces dependency on large annotated datasets. The proposed framework contributes to the development of scalable multilingual Artificial Intelligence systems capable of supporting diverse languages. The concepts presented in this work are valuable for multilingual content generation systems, where efficient AI models can generate meaningful stories and speech outputs while maintaining high performance across different languages.

"Language Independent End-to-End Architecture for Joint Language Identification and Speech Recognition," S. Watanabe, T. Hori, and J. Hershey [6]

This research introduced a language-independent end-to-end architecture capable of performing both language identification and speech recognition within a unified framework. Unlike traditional systems that require separate modules for language detection and recognition, the proposed approach integrates both tasks into a single deep learning model. The architecture automatically identifies the spoken language and generates accurate speech recognition results without additional preprocessing. Experimental results demonstrated improved recognition accuracy, reduced computational complexity, and better adaptability across multilingual environments. The study highlighted the effectiveness of end-to-end learning in simplifying multilingual speech processing while maintaining robust performance. This research significantly influenced the development of modern multilingual AI systems by demonstrating how integrated architectures improve efficiency. Its concepts are directly applicable to multilingual story generation and speech synthesis systems that require seamless processing across multiple languages.

"Analysis of Multilingual Sequence-to-Sequence Speech Recognition Systems," M. Karafiát, M. Baskar, S. Watanabe, T. Hori, and M. Wiesner [7]

This paper analyzed multilingual sequence-to-sequence deep learning models for speech recognition across multiple languages. The authors investigated how encoder-decoder architectures learn common linguistic representations while preserving language-specific characteristics. Their experiments showed that multilingual training improves recognition performance by enabling the model to share information across related languages.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

The proposed sequence-to-sequence framework demonstrated better generalization, robustness, and adaptability than conventional speech recognition systems. The study also emphasized the importance of attention mechanisms in capturing long-range contextual dependencies during speech processing. The findings highlighted that multilingual deep learning models can efficiently support diverse language applications with minimal performance degradation. This work provides valuable insights for multilingual AI systems, including story generation and speech synthesis applications, where understanding contextual information and generating coherent multilingual outputs are essential for delivering high-quality user experiences.

"Large-Scale Multilingual Speech Recognition with a Streaming End-to-End Model," A. Kannan, A. Datta, T. N. Sainath, E. Weinstein, B. Ramabhadran, Y. Wu, A. Bapna, Z. Chen, and S. Lee [12]

This study proposed a large-scale streaming end-to-end multilingual speech recognition model capable of recognizing multiple languages within a single unified architecture. The model was designed to perform real-time speech recognition while maintaining high accuracy and low latency, making it suitable for practical applications. The researchers trained the system on extensive multilingual datasets and demonstrated that shared learning across languages significantly improves recognition performance and scalability. The proposed architecture reduced the need for maintaining separate models for individual languages, thereby simplifying deployment and maintenance. Experimental results confirmed that the streaming model achieved competitive performance across numerous languages while supporting efficient real-time processing. This research represents a major advancement in multilingual Artificial Intelligence and speech technology. Its contributions are highly relevant to multilingual story generation and speech synthesis systems, where fast, accurate, and scalable multilingual processing is essential for providing an interactive user experience.

III. DATASET DETAILS

The dataset used in this study consists of user-provided keywords that are utilized to generate multilingual stories through the Gemini Artificial Intelligence model. Unlike conventional datasets containing predefined text samples, this system accepts dynamic keyword inputs entered by users and processes them to create meaningful and context-aware stories. The generated stories are produced in multiple languages, including English, Hindi, Odia, and Telugu, based on the user's language selection. Each input consists of one or more descriptive keywords that define the theme, characters, or setting of the story. The Gemini Transformer model analyses the semantic relationships among the keywords and generates grammatically correct, coherent, and creative stories while maintaining a proper narrative structure. The dataset therefore represents diverse user inputs covering educational, moral, fantasy, adventure, and general storytelling topics. This variety enables the AI model to generate stories across different domains while preserving contextual relevance and linguistic consistency. The multilingual nature of the dataset makes the proposed system suitable for users from different language backgrounds and supports the development of an intelligent, inclusive, and interactive storytelling platform.

Before story generation, the input keywords undergo preprocessing to improve the quality and relevance of the generated content. The system validates user inputs by removing unnecessary characters, handling empty or invalid entries, and ensuring that the provided keywords are suitable for children's moral stories. The processed keywords are then supplied to the Gemini Transformer model, which performs contextual understanding and generates stories in the selected language. The generated outputs are further converted into speech using a text-to-speech module, providing an enhanced storytelling experience. During system evaluation, different combinations of keywords and multiple language selections are tested to verify the consistency, readability, grammatical correctness, and contextual accuracy of the generated stories. The dataset is utilized throughout the training and evaluation process to assess the effectiveness of the multilingual story generation and speech synthesis system. This approach enables the proposed framework to produce high-quality stories while supporting multilingual communication and improving accessibility for users through both text and speech outputs.

IV. PROPOSED METHODOLOGY

The proposed methodology develops an AI-Powered Multilingual Story Generation and Speech Synthesis System that automatically generates meaningful stories from user-provided keywords using the Gemini Transformer model. Initially, the required software environment is configured by installing Python 3.10.11 along with all the necessary libraries specified in the requirements.txt file. After successful installation, the Django web

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

server is started by executing the runWebserver.bat file. Once the server is running, the user accesses the application through a web browser using the URL http://127.0.0.1:8000/index.html. The system first displays the home page, where new users can register by providing their personal details such as username, email, password, contact number, and address. The registration information is securely stored in the MySQL database. Existing users can log in using their registered credentials to access the multilingual story generation module.

After successful authentication, users are redirected to the story generation interface, where they can enter one or more keywords describing the desired story theme. The system also provides options to select the preferred output language, including English, Hindi, Odia, and Telugu, along with an optional speech generation feature. Once the user submits the keywords, the Gemini AI model analyzes their semantic meaning and contextual relationships using Transformer-based Natural Language Processing techniques. Based on the provided keywords and selected language, the model generates a coherent, grammatically correct, and meaningful children's moral story while maintaining proper story structure and readability. If the user enables the speech option, the generated story is automatically converted into speech through the integrated Text-to-Speech module, allowing users to listen to the narrated story. The generated story and speech output are displayed through a simple and user-friendly web interface, enabling users to generate multiple stories in different languages with ease.

The overall system architecture is shown in Figure 1, which illustrates the complete workflow of the proposed multilingual story generation and speech synthesis system. The architecture consists of user registration, user authentication, keyword input, language selection, Gemini AI-based story generation, multilingual text output, optional speech synthesis, and result visualization through the Django web interface.

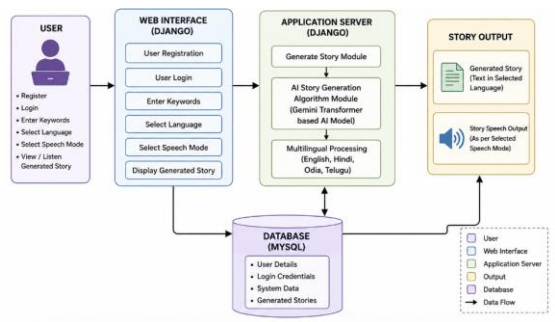

Figure 1. System Architecture for AI-Powered Multilingual Story Generation and Speech Synthesis System

The proposed methodology provides an intelligent and interactive approach for multilingual storytelling using Artificial Intelligence and Natural Language Processing. Initially, users register and log in to the application before entering keywords related to the desired story. The Gemini Transformer model processes these keywords, understands their context, and generates meaningful stories in the selected language. The generated stories maintain grammatical correctness, contextual relevance, and a clear narrative flow. Users can also enable the speech synthesis option to convert the generated stories into natural voice output for improved accessibility and user engagement. The system stores user information securely using MySQL while Django manages all application functionalities. Finally, the proposed system enables users to generate multilingual stories repeatedly with different keywords and languages, providing a scalable, user-friendly, and efficient platform suitable for educational, entertainment, and language-learning applications.

V. RESULT AND DISCUSSION

The experimental results demonstrate the effectiveness of the proposed AI-Powered Multilingual Story Generation and Speech Synthesis System for automatically generating meaningful stories from user-provided keywords. The system was developed using Python, Django, MySQL, and the Gemini Transformer model to provide an intelligent multilingual storytelling platform. Initially, users register by providing their personal information, which is securely stored in the database. After successful registration, users log in using their credentials to access the story generation module. The authenticated users enter one or more keywords, select their preferred language, and optionally enable speech synthesis before generating the story. The Gemini AI model analyzes the semantic relationship between the provided keywords and produces context-aware, grammatically correct, and creative stories in the selected language. Experimental observations showed that the system successfully generated stories in multiple languages such as English, Hindi, Odia, and Telugu while maintaining readability, proper story structure, and contextual consistency. The integrated speech synthesis module further enhanced user interaction by converting the

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

generated text into natural voice output whenever the speech option was enabled.

The developed system was tested using different combinations of keywords and multiple language selections to evaluate the consistency and quality of the generated stories. The registration and login modules successfully authenticated users and ensured secure access to the application. The multilingual story generation module effectively produced unique stories based on different keyword combinations without requiring manual story writing. The generated content maintained grammatical correctness and included meaningful moral lessons suitable for children. When speech mode was selected, the generated stories were successfully converted into speech, improving accessibility and providing an interactive storytelling experience. The overall performance demonstrated that the proposed system provides a simple, efficient, and user-friendly solution for multilingual story creation. The integration of Artificial Intelligence, Natural Language Processing, and Text-to-Speech technology significantly improved user experience and reduced the effort required to generate high-quality stories. These results confirm that the proposed system is suitable for educational, entertainment, and language-learning applications by supporting multilingual communication and intelligent automated storytelling.

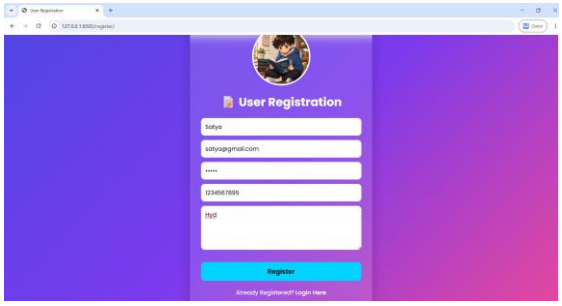

Figure [2] User Registration Module

Figure 2 illustrates the User Registration module of the proposed system. New users provide details such as username, email address, password, contact number, and address to create an account. After successful validation, the user information is securely stored in the MySQL database, allowing authorized access to the multilingual story generation system.

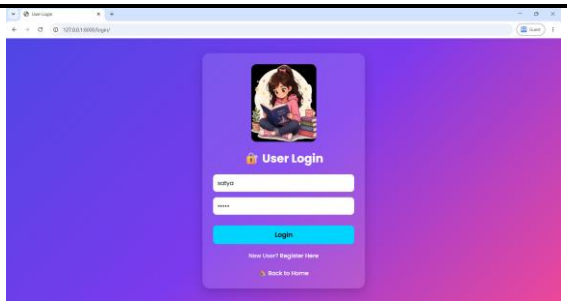

Figure 3: User Login Module

Figure 3 presents the User Login module. Registered users enter their username and password to authenticate themselves. Upon successful verification, the system redirects users to the story generation interface, ensuring secure access to all application functionalities and preventing unauthorized usage.

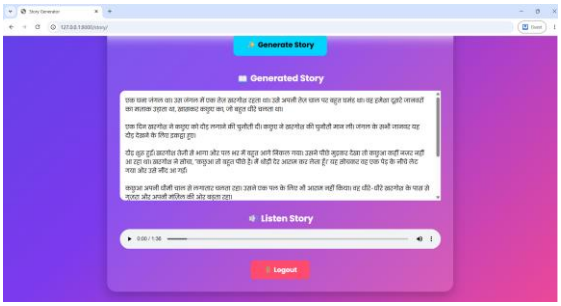

Figure 4: Generate Story Module

Figure 4 illustrates the Generate Story module, which represents the core functionality of the proposed system. Users enter story keywords, select the desired language such as English, Hindi, Odia, or Telugu, and optionally enable the speech synthesis feature. The Gemini Transformer model processes the input keywords, understands their contextual meaning, and generates a meaningful, grammatically correct, and creative story in the selected language. When speech mode is enabled, the generated story is also converted into natural voice output, providing an interactive storytelling experience.

DISCUSSION

The experimental findings demonstrate the effectiveness of the proposed AI-Powered Multilingual Story Generation and Speech Synthesis System in generating meaningful stories across multiple languages. The integration of the Gemini Transformer model enables the system to understand user-provided keywords and generate context-aware stories with proper grammar, logical flow, and readability. Unlike conventional story generation

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

methods that require manual writing, the proposed system automatically creates unique stories while significantly reducing user effort. The multilingual capability allows users to generate stories in different regional languages, making the system accessible to a wider audience.

The registration and login modules provide secure user authentication and efficient management of user information, ensuring reliable access to the application. The generated stories successfully maintained narrative consistency and included appropriate moral lessons suitable for children. Furthermore, the speech synthesis functionality enhanced user engagement by converting generated stories into spoken narration, making the platform more useful for children, visually impaired users, and language learners. Overall, the proposed system combines Artificial Intelligence, Natural Language Processing, and Text-to-Speech technology to deliver an intelligent, interactive, and user-friendly storytelling platform suitable for educational, entertainment, and multilingual learning environments.

VI. CONCLUSION

The proposed "Multilingual Story Generation and Speech System" provides an effective platform for automatically generating meaningful and creative stories using Artificial Intelligence. The system enables users to register, log in, enter keywords, select a preferred language, and generate stories according to their requirements. It integrates functionalities such as user registration, user authentication, multilingual story generation, language selection, speech mode, and AI-based story generation within a single environment. The Gemini Transformer-based Artificial Intelligence model analyzes user-provided keywords, understands their context, and generates coherent, creative, and readable stories while maintaining proper grammar and story structure. The generated stories are produced in multiple languages including English, Hindi, Odia, and Telugu, allowing users to access content according to their linguistic preferences. The project is implemented using Python, Django, MySQL, Google Gemini AI, and the Google Generative AI library to ensure efficient processing, secure user management, and intelligent story generation. Overall, the developed system provides a reliable, efficient, and scalable solution for multilingual story generation and speech-based storytelling by integrating Artificial Intelligence, multilingual language support, user authentication, and automated story creation within a user-friendly platform.

REFERENCES

[1] C. Fugen, S. Stuker, H. Soltau, F. Metze, and T. Schultz, "Efficient Handling of Multilingual Language Models," in Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), 2003.

[2] S. Thomas, S. Ganapathy, and H. Hermansky, "Multilingual MLP Features for Low Resource LVCSR Systems," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012.

[3] Z. Tuske, J. Pinto, D. Willett, and R. Schluter, "Investigation on Cross and Multilingual MLP Features under Matched and Mismatched Acoustical Conditions," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013.

[4] T. Sercu, G. Saon, J. Cui, B. Ramabhadran, B. Kingsbury, and A. Sethy, "Network Architectures for Multilingual Speech Representation Learning," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017.

[5] J. Cui, B. Kingsbury, B. Ramabhadran, et al., "Knowledge Distillation Across Ensembles of Multilingual Models for Low-Resource Languages," in Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), 2017.

[6] S. Watanabe, T. Hori, and J. Hershey, "Language Independent End-to-End Architecture for Joint Language Identification and Speech Recognition," in Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), 2017.

[7] M. Karafiát, M. Baskar, S. Watanabe, T. Hori, and M. Wiesner, "Analysis of Multilingual Sequence-to-Sequence Speech Recognition Systems," arXiv preprint arXiv:1811.03451, 2018.

[8] A. Rosenberg, K. Audhkhasi, A. Sethy, B. Ramabhadran, and M. Picheny, "End-to-End Speech Recognition and Keyword Search on Low-Resource Languages," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

[9] B. Ma, C. Guan, H. Li, and C.-H. Lee, "Multilingual Speech Recognition with Language Identification," in Proceedings of the International Conference on Spoken Language Processing (ICSLP), 2002.

[10] A. Cutler, Y. Zhang, E. Chuangsuwanich, and J. R. Glass, "Language ID-Based Training of Multilingual Stacked Bottleneck Features," in Proceedings of Interspeech, 2014.

[11] S. Watanabe, T. Hori, and J. R. Hershey, "Language Independent End-to-End Architecture for Joint Language Identification and Speech Recognition," in Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), 2017.

[12] A. Kannan, A. Datta, T. N. Sainath, E. Weinstein, B. Ramabhadran, Y. Wu, A. Bapna, Z. Chen, and S. Lee, "Large-Scale Multilingual Speech Recognition with a Streaming End-to-End Model," arXiv preprint arXiv:1909.05330, 2019.

[13] H. Seki, S. Watanabe, T. Hori, J. Le Roux, and J. R. Hershey, "An End-to-End Language-Tracking Speech Recognizer for Mixed-Language Speech," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018.

[14] A. Waters, N. Gaur, P. Haghani, P. Moreno, and Z. Qu, "Leveraging Language ID in Multilingual End-to-End Speech Recognition," in Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), 2019.

[15] N. T. Vu, D. C. Lyu, J. Weiner, D. Telaar, T. Schlippe, F. Blaicher, E. S. Chng, T. Schultz, and H. Li, "A First Speech Recognition System for Mandarin-English Code-Switch Conversational Speech," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012, pp. 4889–4892.

[16] J. Emond, B. Ramabhadran, B. Roark, P. Moreno, and M. Ma, "Transliteration Based Approaches to Improve Code-Switched Speech Recognition Performance," in Proceedings of the IEEE Spoken Language Technology Workshop (SLT), 2018.

[17] B. M. L. Srivastava and S. Sitaram, "Homophone Identification and Merging for Code-Switched Speech Recognition," in Proceedings of Interspeech, 2018, pp. 1943–1947.

[18] A. Ragni, M. J. F. Gales, and K. M. Knill, "A Language Space Representation for Speech Recognition," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015.

[19] P. Swietojanski and S. Renals, "Learning Hidden Unit Contributions for Unsupervised Speaker Adaptation of Neural Network Acoustic Models," in Proceedings of the IEEE Spoken Language Technology Workshop (SLT), 2014.

[20] S. Tong, P. N. Garner, and H. Bourlard, "Multilingual Training and Cross-Lingual Adaptation on CTC-Based Acoustic Models," arXiv preprint arXiv:1711.10025, 2017.

[21] M. Karafiát, F. Grézl, M. Hannemann, and J. H. Černockỳ, "BUT Neural Network Features for Spontaneous Vietnamese in BABEL," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014.

[22] L. Hellsten, B. Roark, P. Goyal, C. Allauzen, F. Beaufays, T. Ouyang, M. Riley, and D. Rybach, "Transliterated Mobile Keyboard Input via Weighted Finite-State Transducers," in Proceedings of FSMNLP, 2017, pp. 10–19.

[23] M. Bisani and H. Ney, "Joint-Sequence Models for Grapheme-to-Phoneme Conversion," Speech Communication, vol. 50, no. 5, pp. 434–451, 2008.

[24] C. Kim, A. Misra, K. Chin, T. Hughes, A. Narayanan, T. N. Sainath, and M. Bacchiani, "Generation of Large-Scale Simulated Utterances in Virtual Rooms to Train Deep Neural Networks for Far-Field Speech Recognition in Google Home," in Proceedings of Interspeech, 2017.

[25] A. Graves, "Sequence Transduction with Recurrent Neural Networks," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012.

[26] Y. He, T. N. Sainath, R. Prabhavalkar, I. McGraw, R. Alvarez, D. Zhao, D. Rybach, A. Kannan, Y. Wu, R. Pang, et al., "Streaming End-to-End Speech Recognition for Mobile Devices," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019.

[27] J. Shen, P. Nguyen, Y. Wu, Z. Chen, et al., "Lingvo: A Modular and Scalable Framework for

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Sequence-to-Sequence Modeling," arXiv preprint arXiv:1902.08295, 2019.

[28] N. Jouppi, C. Young, and N. Patil, "In-Datacenter Performance Analysis of a Tensor Processing Unit," CoRR abs/1704.04760, 2017.

[29] J. Emond, B. Ramabhadran, B. Roark, P. Moreno, and M. Ma, "Transliteration Based Approaches to Improve Code-Switched Speech Recognition Performance," in Proceedings of the IEEE Spoken Language Technology Workshop (SLT), 2018.

[30] A. Datta and A. Kannan, "Large-Scale Multilingual Speech Recognition with a Streaming End-to-End Model," Google AI Blog, Sept. 2019.