iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.68665

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

SPEAK PDF

Ayush Kamble[1], Yash Patel[2], Bhagyashree Khaire[3], Dr. Renuka Deshpande[4]
Student, Shivajirao S Jondhale College of Engineering Dombivali East

Abstract: ***Our project is a comprehensive tool designed to convert PDF documents into audio format, integrating translation and summarization features, aimed at improving accessibility and promoting multilingualism. The proposed system utilizes state-of-the-art text-to-speech (TTS technology to convert text-based PDF documents into audio files, enabling individuals with visual impairments or learning disabilities to access content more conveniently. Moreover, this system incorporates machine translation algorithms to facilitate seamless conversion of PDFs into various languages, thus breaking down language barriers and fostering inclusivity. In the digital era, an overwhelming amount of information is shared through PDF documents, ranging from research papers and reports to legal contracts and business proposals. Manually extracting key insights from these documents can be time-consuming and challenging. Our approach leverages natural language processing (NLP) techniques to analyze text, identify crucial content, and generate human-like summaries that retain the original document's intent. By incorporating machine learning models, the system ensures that summaries are concise, accurate, and easy to understand. The goal is to enhance productivity by reducing the time required to comprehend lengthy documents while preserving their key messages. This solution can be valuable for professionals, researchers, and organizations dealing with extensive textual data, ultimately enabling smarter decision-making and improved information accessibility. We believe that this project has the potential to make a significant impact in the field of computer science and beyond.***

I. INTRODUCTION

In today's fast-paced digital world, professionals, students, and researchers frequently interact with extensive PDF documents, including research papers, reports, and legal documents. Manually reading and summarizing these documents can be time-consuming and cognitively demanding. To address this challenge, automated solutions that convert PDFs into audio format and generate concise summaries are gaining significant interest. This project focuses on developing an intelligent PDF-to-Audio Converter and Summarization System that enhances accessibility and productivity. By leveraging Natural Language Processing (NLP) and Text-to-Speech (TTS) technologies, our system extracts meaningful insights from PDFs, summarizes key information, and converts text into natural-sounding speech. This project focuses on developing an intelligent PDF-toAudio Converter and Summarization System that enhances accessibility and productivity. By leveraging Natural Language Processing (NLP) and Text-to-Speech (TTS) technologies, our system extracts meaningful insights from PDFs, summarizes key information, and converts text into natural-sounding speech. Incorporating insights from the papers [1], [2] this report underscores the critical role of advanced technologies in enhancing accessibility and communication. Baker and Tata[1] emphasize that Text-to-Speech (TTS) technology significantly improves the usability of PDF documents for visually impaired individuals, enabling them to access information that would otherwise remain inaccessible. This technology not only facilitates better engagement with digital content but also promotes inclusivity in information dissemination. Meanwhile, Saini and Sahula [2] highlight the challenges and advancements in machine translation for Indian languages, noting the importance of developing efficient systems that cater to the diverse linguistic landscape of India. They point out that addressing issues such as limited data availability and language-specific nuances is essential for creating robust translation solutions. Together, these perspectives illustrate the necessity of integrating TTS and machine translation technologies to create comprehensive systems that enhance accessibility, bridge communication gaps, and empower users from various backgrounds.

II. OBJECTIVE

The objectives of our project PDF to audio converter, translator, and summarizer project are to enhance accessibility and comprehension of textual content in diverse languages, addressing significant gaps identified in existing literature. Current research often highlights the challenges faced by individuals with visual impairments and non-native speakers when engaging with PDF documents. By developing a tool that not only converts text to audio but also translates it into various languages, we aim to bridge the accessibility divide. Additionally, incorporating a summarization feature will allow users to extract key information efficiently, catering to those with limited time or cognitive load. We have integrated all these functionalities into a single interface, streamlining the user experience and making it more efficient for document processing.

III. LITERATURE REVIEW

Şatır and H. Bulut introduces, "A Novel Hybrid Approach to Improve Neural Machine Translation Decoding using PhraseBased Statistical Machine Translation," presents a hybrid method to enhance neural machine translation (NMT) decoding, phrase-based statistical machine translation (PBSMT) with NMT, aiming to improve translation accuracy and fluency. The proposed technique leverages the phrasebased translation model during the decoding process to overcome the limitations of traditional NMT models, particularly in handling long-range dependencies and context.[3]

Ganesh et al. (2023) provide a comprehensive survey of various machine translation approaches, including rulebased, statistical, and neural methods. They analyze advancements, challenges, and future directions in the field, highlighting the impact of deep learning on translation quality. The study serves as a valuable resource for researchers exploring machine translation technologies.[4]

The review paper by S. Sharma, M. Diwakar, P. Singh, A. Tripathi, C. Arya, and S. Singh, titled "A Review of Neural Machine Translation based on Deep Learning Techniques," provides a comprehensive overview of neural machine translation (NMT) methods utilizing deep learning techniques. It examines the evolution of NMT, including architectures such as sequence-to-sequence models, attention mechanisms, and transformer networks. The review also highlights the advantages and challenges of NMT, particularly in handling complex language pairs and translation quality.[5]

The paper by S. Sarjun Beevi, Tayi Gopi Chand, Tamatam Hemanth Reddy, Tammana Rama Naga Sai Gokul, and Alamuru Harika, titled "PDF to Voice by Using Deep Learning," explores the use of deep learning techniques to convert PDF documents into voice. The proposed system employs deep learning models to extract text from PDFs and convert it into speech, making content accessible to visually impaired individuals. [6]

The paper "Image Text to Audio Conversion Using Raspberry Pi" by Gowri Ch1, Manikanta Y, Lohitha Y, Nagur Babu Sk, and Arun Kumar P discusses a system that extracts text from images using Optical Character Recognition (OCR) and converts it into speech using text-to-speech (TTS) technology. Implemented on a Raspberry Pi, the system aims to assist visually impaired individuals by providing an affordable and portable solution for reading printed or digital text aloud.[7]

Lee et al. (2021) present a novel approach to direct speechto-speech translation using discrete units, bypassing intermediate text representations. Their model learns discrete acoustic representations and translates speech directly, improving efficiency and robustness. The method demonstrates promising results on benchmark datasets, showcasing its potential for real-world applications in multilingual communication.[8]

The paper "Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation" by Melvin Johnson et al. introduces a single neural network that can translate multiple languages without needing separate models. It explores zero-shot translation, where the system translates between language pairs it has never explicitly seen during training. The study demonstrates how multilingual training enhances translation quality and enables efficient knowledge transfer across languages.[9]

The paper "Text Summarization with Pretrained Encoders" by Yang Liu and Mirella Lapata explores the use of pretrained transformer-based encoders, such as BERT, for text summarization. The study investigates how fine-tuning these models can improve both extractive and abstractive summarization tasks. The results show that leveraging pretrained representations significantly enhances summarization quality, outperforming traditional methods on benchmark datasets.[10]

Gaido et al. (2020) explore the use of knowledge distillation to improve end-to-end speech translation models. By transferring knowledge from a high-performing teacher model to a smaller student model, they enhance translation accuracy and efficiency. Their approach reduces complexity while maintaining performance, making speech translation more practical for real-world applications.[11]

The paper "Speech-to-Text and Text-to-Speech Recognition Using Deep Learning" by T. Vaishnavi Reddy and K. P. Kumar explores the application of deep learning models in speech-to-text (STT) and text-to-speech (TTS) systems. It discusses the effectiveness of neural networks in improving accuracy, naturalness, and efficiency in speech processing. The study highlights advancements in automatic speech recognition (ASR) and speech synthesis, demonstrating their impact on human-computer interaction and accessibility.[12]

A. Drawbacks

One significant challenge is the accuracy of Optical Character Recognition (OCR) technology[13], which can struggle with complex document layouts, varying fonts, and handwritten text, leading to errors in text extraction. Additionally, while Text-to-Speech (TTS) [14] engines have made strides in naturalness, they can still produce robotic-sounding voices that lack emotional nuances, potentially reducing user engagement.

In translation systems, particularly for less widely spoken languages, the limited availability of parallel corpora hampers the development of robust models, resulting in inaccuracies when handling idiomatic expressions and context-specific language. Furthermore, existing machine translation solutions often lack tailored evaluation metrics that accurately reflect translation quality, making it difficult to assess their effectiveness. Collectively, these limitations highlight the need for ongoing improvements to enhance the usability, accuracy, and accessibility of PDF to audio converters and translation systems.

IV. SYSTEM ARCHITECTURE

The above flowchart outlines a process for reading a PDF file using the PyPDF2 library. The sequence begins with the user uploading a PDF file. If the file is successfully uploaded, the program retrieves the pages and extracts the text. If no file is selected, it prompts the user to select one. Finally, the extracted content can be outputted as audio.

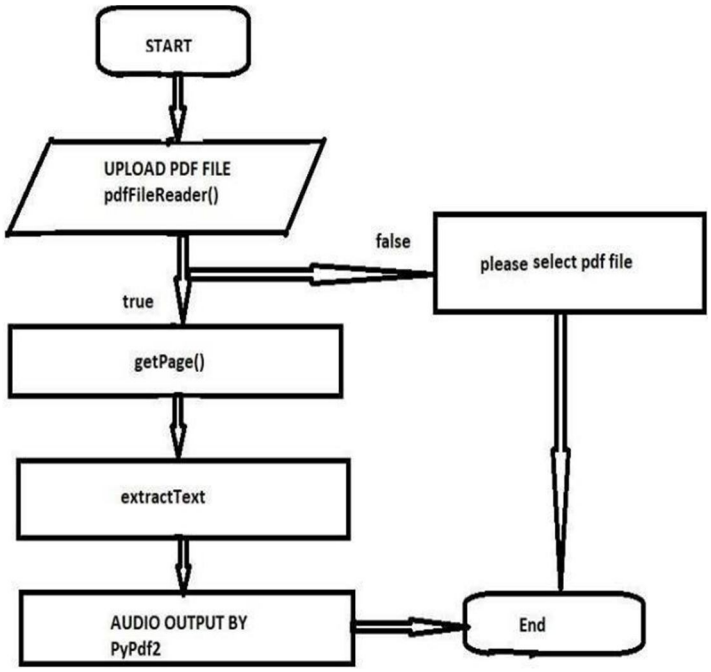


Fig 1 :- Pdf to Audio Converter

The following steps shows process of Pdf to Audio Converter-

1) Start: Initiate the PDF file reading process.
2) Upload PDF File: User uploads a PDF file using pdfFileReader().
3) Check Upload: If the upload is successful, proceed to extract text.
4) Extract Text: Use getPage() and extract Text() to retrieve text from the PDF.
5) No File Selected: If no file is selected, prompt the user to upload a PDF.
6) Audio Output: Convert the extracted text to audio format.
7) End: Complete the process.

V. METHODOLOGY

Developing our system that converts PDFs to audio and translates them into different languages involves various modules and technologies, primarily focusing on text extraction, translation, text-to-speech (TTS conversion, and user interaction. Below is a step-by-step methodology we used for creating a PDF to audio converter with a translation feature.

A. Input and Text Extraction

- Objective: We extracted raw text from the PDF file.
- Tools: Libraries like PyPDF2, PDFMiner, Apache PDFBox, or Tesseract[15] (for OCR on scanned PDFs.)
- Steps:
  - Load the PDF document. - Extract text content from the PDF pages.

B. Text Preprocessing

- Objective: We have clean and structure the text for better translation and audio generation.
- Tasks:
  - Remove any unnecessary characters (e.g., special symbols, extra spaces.
  - Handle sentence splitting, tokenization, and removing noise from the text.

C. Text Summarization

- Objective: To automatically generate a concise summary of the extracted text from the PDF to save time for the user and provide the essential content.
- Tasks:
  - The extracted text is passed to a summarization model (e.g., BART or T5). [16]
  - The system processes the text and outputs a summary with relevant and important information

D. Language Translation

- Objective: Translate the extracted text into a target language.
- Tools: APIs like Google Cloud Translate, Amazon Translate, or Microsoft Translator.
- Steps:
  - Input the pre-processed text into the translation API.
  - Specify the source language (either detect automatically or define manually).
  - Specify the target language for translation (e.g., English to Sanskrit). [17]
  - Perform translation and receive the translated text output.

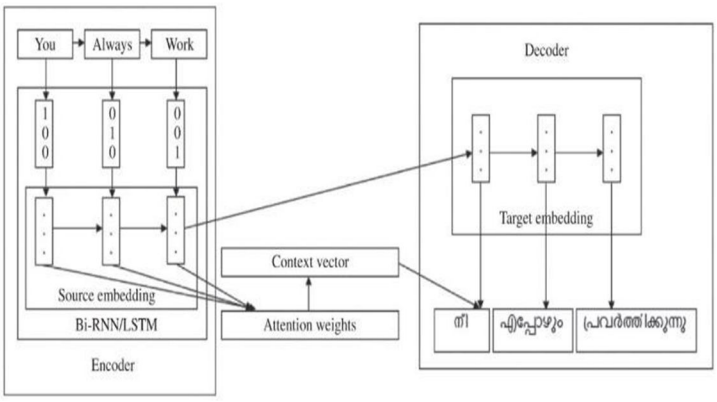


Fig .2 :- Language Translation

E. Text-to-Speech (TTS Conversion)

A. TTS Engine Selection

- Objective: Convert the translated text into audio in the desired language.
- Tools: TTS engines[17] such as Google Cloud Textto-Speech, Amazon Polly, Microsoft Azure Cognitive Services, or open-source TTS frameworks like espeak.
- Steps:
  - Feed the translated text into the TTS engine.
  - Generate the audio output in formats such as MP3 or WAV

F. Audio File Processing

o Objective: Ensure high-quality audio output. o Tasks:

- Apply noise reduction and audio normalization techniques.[18]
- Provide audio customization options, such as adjusting playback speed or selecting different voice styles.

G. Output and Results Delivery

o Objective: Provide users with the ability to listen to and download the generated audio and allows users to upload PDF files and select their preferred language and audio settings.

o Tasks:

- Integrated an audio player into the user interface for direct playback.
- Allow users to download the audio file in multiple formats (e.g., MP3, WAV.

VI. RESULTS

Our Speak Pdf Web Interface converts PDF documents into spoken audio while also providing concise summaries. It supports multiple languages, ensuring accessibility for a global audience. Users can listen to documents on the go and grasp key points quickly. Ideal for students, professionals, and visually impaired individuals. Below are rssults:-

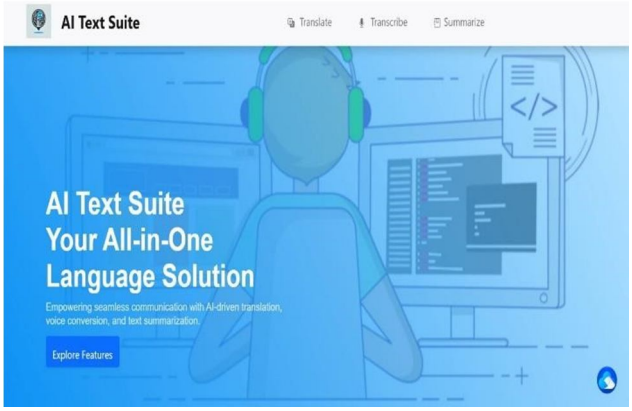


Fig 3 : Web Interface

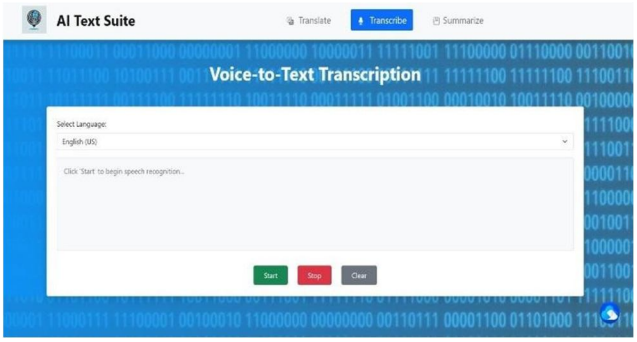


Fig 4: Voice to Text Translator

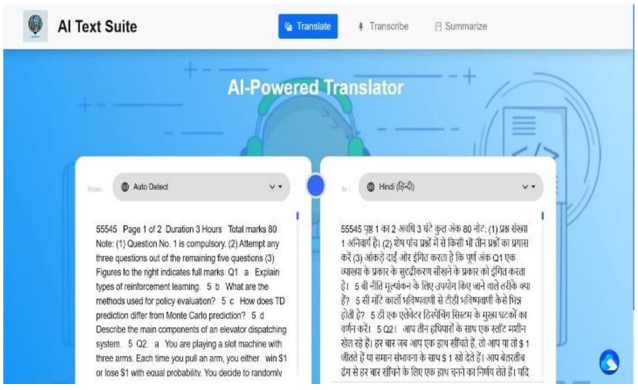


Fig 5 :- Text Translator

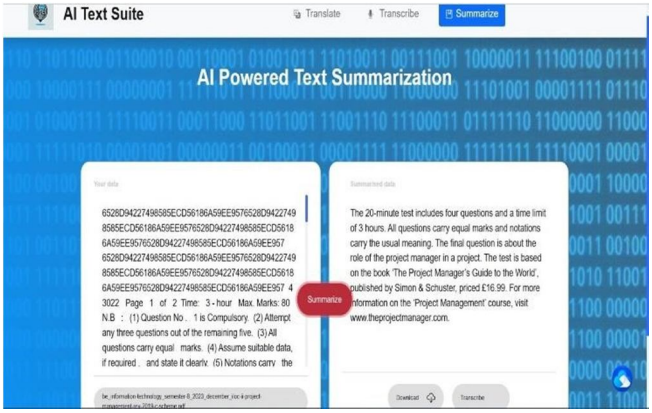


Fig 6:- Text Summarization

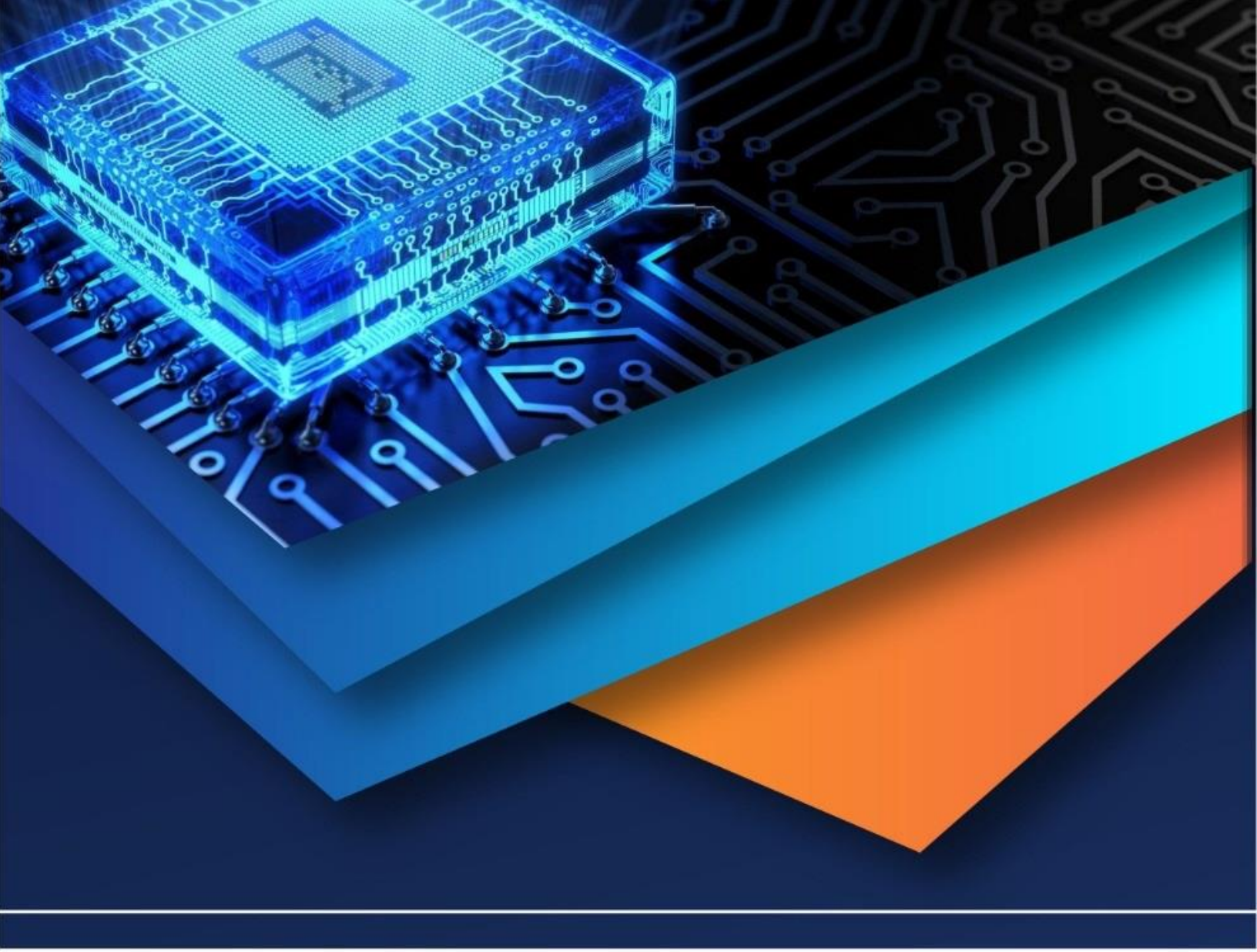
VII. CONCLUSION AND FUTURE SCOPE

The exploration of Online PDF to Audio Converter and Language Translator tools highlights their transformative impact on technology, linguistic accessibility, and inclusive communication. These tools have effectively addressed accessibility issues, particularly for individuals with visual impairments. The literature emphasizes the crucial role these tools play in fostering cross-cultural communication, connecting people across linguistic barriers, and contributing to a more interconnected global society. Technological advancements, especially in natural language processing and machine translation, have improved the effectiveness of these tools. However, challenges such as accuracy in language translation and ethical considerations like privacy protection require ongoing attention. Despite these challenges, the educational applications of these tools offer promising opportunities for enhancing language learning experiences and making educational materials more accessible to diverse learners. Future proposals include integrating artificial intelligence for context-aware translations. In conclusion, Online PDF to Audio Converter and language Translator tools are catalysts for positive change in digital communication, enabling inclusivity and understanding across linguistic and cultural boundaries.

REFERENCES

[1] Baker, D., & Tatar, D, “Improving Accessibility: The Role of Text-to-Speech Technology in PDF Document Conversion”, Journal of Assistive Technologies,2015
[2] Sandeep Saini, Vineet Sahula "Survey: Machine Translation for Indian Languages", IEEE International Conference on Computational Intelligence and Communication Technology,2015
[3] E. Şatır and H. Bulut, "A Novel Hybrid Approach to Improve Neural Machine Translation Decoding using Phrase-Based Statistical
[4] Machine Translation," 2021 International Conference on Innovations in Intelligent Systems and Applications (INISTA), Kocaeli, Turkey, 2021, pp. 1-5.
[5] S. Ganesh, V. Dhotre, P. Patil and D. Pawade, "A Comprehensive Survey of Machine Translation Approaches," 2023 6th International Conference on Advances in Science and Technology (ICAST), Mumbai, India, 2023, pp. 160-165
[6] S. Sharma, M. Diwakar, P. Singh, A. Tripathi, C. Arya and S. Singh, "A Review of Neural Machine Translation based on Deep learning techniques," 2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), Dehradun, India, 2021, pp. 1-5
[7] S.Sarjun Beevi, Tayi Gopi Chand, Tamatam Hemanth Reddy, Tammana Rama Naga Sai Gokul, Alamuru Harika,"Pdf to Voice by Using Deep Learning" ,International Journal of Innovative Science and Research Technology (IJISRT), 2015
[8] Gowri Ch1, Manikanta Y, Lohitha Y, Nagur Babu Sk, Arun Kumar P, “Image Text to Audio Conversion Using Raspberry Pi”, International Journal of Engineering Research & Technology Volume 13, Issue 03 (March 2024).
[9] Lee, Ann & Chen, Peng-Jen & Wang, Changhan & Gu, Jiatao & Ma, Xutai & Polyak, Adam & Adi, Yossi & He, Qing & Tang, Yun & Pino, Juan & Hsu, Wei-Ning, “Direct speech-to-speech translation with discrete units.”, 60th Annual Meeting of the Association for Computational Linguistics,2021, Volume 1,pages 3327 – 333.
[10] Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, and Jeffrey Dean, “Google’s Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation”, Transactions of the Association for Computational Linguistics, 2017,vol 5, pp.339–351.
[11] Yang Liu, Mirella Lapata ,“Text Summarization with Pretrained Encoders” , he 9th International Joint Conference on Natural Language Processing, Hong Kong, China, November 3–7, 2019, pp 3730–3740.
[12] Gaido, Marco & Di Gangi, Mattia & Negri, Matteo & Turchi, Marco. (2020). “End-to-End Speech-Translation with Knowledge Distillation” , 10.48550/arXiv.2006.02965.
[13] V. M. Reddy, T. Vaishnavi and K. P. Kumar, "Speechto-Text and Text-to-Speech Recognition Using Deep Learning," 2023 2nd International Conference on Edge Computing and Applications (ICECAA), Namakkal, India, 2023, pp.657-666
[14] J. Memon, M. Sami, R. A. Khan and M. Uddin, "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)," in IEEE Access, 2020,vol. 8, pp. 142642-142668.

[15] Ayushi Trivedi,Navya Pant, Pinal Shah,Simran Sonik and Supriya Agrawal, "Speech to text and text to speech recognition systems", 2018, Volume 20, Issue 2, pp. 36-43.
[16] K. Joshi and H. Arolkar, "Comparative Analysis of Outcomes of Tesseract OCR for Different Languages," 2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2024, pp. 95-100.
[17] Ramesh Nallapati, Feifei Zhai, Bowen Zhou, "SummaRuNNer: A Recurrent Neural Network based Sequence Model for Extractive Summarization of Documents", The Thirty-First AAAI Conference on Artificial Intelligence , 2017, arXiv:1611.04230 [cs.CL]
[18] Sitender, Bawa, S., Kumar, M. et al, "A comprehensive survey on machine translation for English, Hindi and Sanskrit languages", J Ambient Intelligence Human Computing , 2021, pp 3441–3474.
[19] Ye Jia , Michelle Tadmor Ramanovich ,Tal Remez, Roi Pomerantz , "Translatotron 2 - High-Quality Direct Speechto-Speech Translation with Voice Preservation", International Conference on Machine Learning, (2021), https://arxiv.org/abs/2107.08661



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY