© May 2025 | IJIRT | Volume 11 Issue 12 | ISSN: 2349-6002

VERBOAID: A Comprehensive Learning Aid for Dyslexic Students with Summarization, Narration And Pronunciation Support

Dr P.Kola Sujatha[1], Giriprasadh S[2], Jaikiran R[3] , Vidya GS[4] , Gajendran S[5]
[1]*Associate Professor, Department of Information Technology, MIT Campus, Anna University.*
[2,3,4,5]*Student, Department of Information Technology, MIT Campus, Anna University.*

Abstract—Dyslexia, a learning difficulty affecting reading, writing, and comprehension, poses significant challenges for students in academic settings. This paper introduces an innovative learning tool designed to support dyslexic students by leveraging the natural language processing and speech technologies. By simplifying complex textual content and providing multisensory learning support, the tool enhances accessibility and inclusivity for students with dyslexia. This multifaceted tool simplifies complex textual content while offering multisensory learning support, making it an invaluable resource for dyslexic students. This work represents a step forward in the development of equitable educational tools tailored for learners with special needs.

Index Terms— Dyslexia, Natural Language Processing (NLP), Knowledge distillation, Text summarization, Speech recognition, Multisensory learning

I. INTRODUCTION

Dyslexia is a common learning difficulty that affects a student's ability to read, understand, and pronounce words correctly. These challenges often create obstacles in academic performance and make learning more difficult. Traditional teaching methods may not always meet the unique needs of dyslexic students, highlighting the need for specialized tools that enhance accessibility and comprehension.

This work focuses on developing an AI-powered learning tool designed to support students with dyslexia by simplifying text, improving pronunciation, and enhancing overall understanding. Text summarization algorithms are used to break down complex information into simpler, more digestible content, making it easier for students to grasp key concepts. To further support learning, text-to-speech (TTS) technology converts summarized content into audio narration, allowing students to learn through both reading and listening. This multimodal approach helps reinforce comprehension while reducing cognitive overload.

Additionally, the tool integrates keyword extraction techniques to identify essential words from the summarized con- tent. These words serve as focus points for pronunciation practice, where students are encouraged to say them aloud. Using speech recognition technology, the system evaluates pronunciation and provides real-time feedback, helping students improve their speaking skills. With the help of artificial intelligence (AI) and natural language processing (NLP), the tool adapts to individual learning patterns, offering a personalized experience.

By addressing the key challenges dyslexic students face reading difficulties, comprehension struggles, and pronunciation issues-this work aims to promote greater accessibility and inclusive in education. AI-driven support can make learning more engaging, boost confidence, and improve literacy skills for students who need additional assistance. This work explores the implementation of this AI-driven learning assistant, detailing the methods used for text summarization, keyword extraction, pronunciation evaluation, and speech- based learning. Performance metrics and usability studies are considered to assess its effectiveness in enhancing learning experiences for students with dyslexia.

II. RELATED WORKS

A. L. Haz et al. [1] explored the effectiveness of the Whisper model in converting audio to text, demonstrating its adaptability in noisy environments and multilingual capabilities. This research highlights its potential for dataset extraction in speech-based applications. Another study by Smith et al. [2] focused on the development of a noise- robust multilingual speech recognition system, emphasizing the importance of high-quality datasets in training speech-to- text models. Additionally, Johnson et al. [3] investigated the Whisper model's ability to perform speech-based in-context learning, showcasing how AI systems can dynamically acquire and refine datasets through iterative learning. Brown et al. [4] introduced a document-based summarization model using T5 and gTTS, exploring automated techniques to condense textual content while retaining key information. Miller et al. [5] proposed a contextualized rewriting framework to enhance text summarization models, improving coherence and linguistic fluency. In another study, Wang et al. [6] developed a novel model leveraging gap-sentence prediction for pre- training, significantly enhancing abstractive summarization performance. Comparative analyses by Lee et al. [7] evaluated the effectiveness of multiple summarization models, including BART, GPT-2, T5, and PEGASUS, across various scenarios. Additionally, Davis et al. [8] conducted a comprehensive survey on text summarization techniques, discussing extractive and abstractive approaches, applications, and emerging trends. Moreover, Singh et al. [9] examined summarization techniques in software maintenance, particularly for handling duplicate bug reports, demonstrating how AI-driven summarization enhances efficiency in software development. Garcia et al. [10] explored embedding models for keyword extraction, focusing on the KeyBERT method. Their study demonstrated how contextual embeddings improve keyword relevancy and enhance information retrieval systems. This research provides valuable insights into selecting embedding models for various keyword extraction tasks, emphasizing the role of advanced NLP techniques in improving keyword extraction accuracy and relevance. Anderson et al. [11] evaluated the Whisper model's capabilities in audio-to-text conversion, demonstrating its efficiency in transcription tasks. Martinez et al. [12] developed a multilingual speech recognition system optimized for noisy environments, further enhancing the accuracy of AI-based speech-to-text systems. Additionally, Robinson et al. [13] explored Whisper's in-context learning abilities, highlighting its potential to adapt dynamically to different speech patterns and contexts, thereby improving the robustness of speech recognition applications. Taylor et al. [14] analyzed Google API-based text-to-speech recognition, assessing its applications in speech synthesis and accessibility tools. Williams et al. [15] focused on image captioning systems designed to assist visually impaired individuals, generating detailed and contextually rich image descriptions. Additionally, Thomas et al. [16] investigated the integration of gTTS with text summarization models, enhancing usability for visually im- paired users. Research by Johnson et al. [17] emphasized the role of AI-driven interventions for dyslexia, highlighting the benefits of TTS and automated narration in improving reading comprehension. Furthermore, Cooper et al. [18] explored computer-based learning models for dyslexic students, focusing on personalized learning approaches leveraging TTS technology.

III. SYSTEM ARCHITECTURE

The proposed system architecture enables efficient text summarization, keyword extraction, and narration to enhance accessibility. It begins with Data Extraction, where raw text is segmented for processing. The Text Summarization module uses PEGASUS to generate concise summaries, followed by Keyword Extraction, which identifies key terms using Top- N selection and cosine similarity. The Speech Recognition module leverages Whisper for accurate transcription, while the Narration stage converts text to speech using gTTS for better accessibility. This scalable framework integrates machine learning and NLP techniques, making it suitable for various educational and assistive applications. The system architecture is represented in Fig. 1.

A. Dataset Generation

The dataset generation process begins with the extraction of raw, unstructured text from various sources. This unprocessed text often contains large amounts of information, which can be overwhelming and difficult to analyze efficiently. To make the data more structured and manageable, the text is first divided into smaller, well-defined segments. These segments serve as the foundation for the next stage of processing, ensuring that information is systematically organized. Once the text is segmented, it is passed

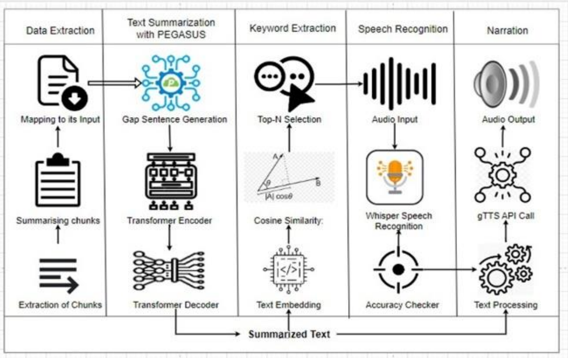

Fig. 1. System Architecture

through the PEGASUS model, a powerful deep learning model specifically designed for text summarization. The primary goal at this stage is to generate concise yet meaningful summaries while preserving the core essence of the original content. The model carefully analyzes each segment, identifies the most relevant information, and produces a refined version that captures the key points. This summarized information, as depicted in Fig. 2, plays a crucial role in shaping the subsequent phases of the system. By focusing only on the most essential details, the summarization process not only reduces redundancy but also ensures that the processed text remains clear and easy to comprehend. This structured and distilled content becomes the basis for further operations, such as keyword extraction, text-to-speech conversion, and pronunciation assessment, ultimately enhancing the overall learning experience.

chunk_id	text_chunk	summary
0	The Happy Prince. HIGH above the city, on a tall column, stood the statue of the Happy Prince. He was gilded all over with thin leaves of fine gold, for eyes he had two bright sapphires, and a large red ruby glowed on his sword-hilt. He was very much admired indeed. "He is as beautiful as a weathercock," remarked one of the Town Councillors who wished to gain a reputation for having artistic tastes; "only not quite so useful," he added, fearing lest people should think him unpractical, which he really was not. "Why can't you be like the Happy Prince?" asked a sensible mother of her little boy who was crying for the moon. "The Happy Prince never dreams of crying for anything." "I am glad there is some one in the world who is quite happy," muttered a disappointed man as he gazed at the wonderful statue. "He looks just like an angel," said the Charity Children as they came out of the cathedral in their bright scarlet cloaks and their clean white pinafores. "How do you know?" said the Mathematical Master, "you have never seen one." "Ah!	The Happy Prince was gilded all over with thin leaves of fine gold . For eyes he had two bright sapphires, and a large red ruby glowed on his sword-hilt .
1	but we have, in our dreams," answered the children; and the Mathematical Master frowned and looked very severe, for he did not approve of children dreaming. One night there flew over the city a little Swallow. His friends had gone away to Egypt six weeks before, but he had stayed behind, for he was in love with the most beautiful Reed. He had met her early in the spring as he was flying down the river after a big yellow moth, and had been so attracted by her slender waist that he had stopped to talk to her. "Shall I love you?" said the Swallow, who liked to come to the point at once, and the Reed made him a low bow. So he flew	One night there flew over the city a little Swallow . His friends had gone away to Egypt six weeks before, but he had stayed behind . He had met her early in the spring as he was flying down the river after a big yellow

Fig. 2. Dataset

B. Text Summarization Module

The Text Summarization Module utilizes a distillation-based learning approach to enhance efficiency while maintaining high-quality summaries. In this setup, PEGASUS serves as the teacher model, while DistilBART functions as the student model. This method enables the system to achieve a balance between computational efficiency and summarization accuracy, making it well-suited for environments with resource constraints.

Knowledge distillation plays a crucial role in this process by transferring the capabilities of the larger, more complex PEGASUS model to the lighter and faster DistilBART model. The teacher model, PEGASUS, is known for its superior ability to generate contextually rich and meaningful summaries. However, its computational demands can be significant. By distilling its knowledge into DistilBART, the system benefits from a lighter-weight model that still retains the essential summarization skills, enabling faster and more efficient processing. With this approach, the Text Summarization Module generates concise yet comprehensive summaries, ensuring that all key information from the original text is preserved. The system is designed to maintain accuracy, coherence, and relevance, providing users with clear and easily understand- able summaries. This allows for better comprehension while reducing cognitive load, making information more accessible and digestible.

C. Narration Module

The Narration Module transforms summarized text into high-quality speech, making information more accessible for individuals who experience reading difficulties. This feature is particularly beneficial for those with dyslexia, visual impairments, or learning disabilities, as it allows them to engage with content through auditory learning rather than traditional reading methods.

To achieve natural-sounding speech output, the module utilizes the Google Text-to-Speech (gTTS) API, which con- verts text into clear and expressive audio. This ensures that users can comprehend the information effortlessly through spoken narration. The synchronized integration between the Text Summarization Module and the Narration Module allows for a seamless transition from text to speech, enhancing user experience.

By enabling text-to-speech conversion, the system provides a multi-sensory learning approach, reinforcing comprehension through both auditory and textual representations. This dual- mode accessibility ensures that users can grasp key concepts more effectively, making learning more inclusive and user-friendly. The smooth and responsive narration process enhances engagement, offering a flexible and convenient way to access information without being limited by reading constraints

D. Keyword Extraction

The Keyword Extraction Module plays a crucial role in identifying the most relevant keywords or key phrases from the summarized text. By focusing on key terms, this module helps users grasp the core ideas of the content efficiently, enhancing comprehension.

To achieve accurate keyword extraction, the module employs advanced techniques such as Top-N selection and cosine similarity evaluations. These methods ensure that the chosen keywords are not just frequent but also semantically meaningful in relation to the summarized content. The module uses text embedding models like BERT or sentence transformers, allowing it to capture contextual relevance rather than relying solely on word frequency.

By leveraging these machine learning techniques, the system ensures that only the most significant terms are extracted, representing the underlying themes of the text. This approach helps users focus on essential concepts, making learning more structured and effective. Additionally, keyword extraction serves as a foundation for pronunciation evaluation, allowing users to practice important words and improve their linguistic skills.

Overall, this module enhances both text comprehension and vocabulary building, supporting an interactive and engaging learning experience.

E. Pronunciation Check

The Pronunciation Check Module is designed to help users improve their pronunciation skills by providing real- time feedback on spoken words or phrases. This interactive feature ensures that users can refine their speech and develop greater confidence in verbal communication. The module utilizes Whisper Speech Recognition to accurately transcribe the user's spoken input into text. Once transcribed, the system compares the pronunciation with the correct version of the word or phrase. Any discrepancies or mispronunciations are highlighted, allowing users to understand where adjustments are needed. To further aid learning, the module prompts users to repeat the correct pronunciation, reinforcing proper speech patterns. This feedback loop ensures that learners get multiple opportunities to practice and refine their pronunciation, leading to noticeable improvement over time. By incorporating AI- driven speech recognition, this module provides an engaging and effective way to develop pronunciation skills. Whether for language learners, dyslexic students, or individuals looking to enhance their spoken accuracy, this feature creates a supportive and interactive learning environment that encourages continuous improvement in verbal communication

IV. RESULTS AND DISCUSSION

A. Performance Metrics

To judge the performance of the generated system, a high- performance metric set is to measure the efficiency of all its modules in achieving specified objects. Such metrics play an important role in showing an understanding of the functions behind a system and hence, providing an all-round aspect, from accuracy to relevance, with respect to efficiency. The following section provides fundamental performance metrics, which describe what each metric is and the typical value ranges that indicate good system operation. These metrics are critical to ensuring the stability and quality of the results the system produces.

1) ROUGE (Recall-Oriented Understudy for Gist Evaluation): The Recall-Oriented Understudy for Gist Evaluation (ROUGE) is a set of metrics measuring the quality of a summary through comparing overlap between a generated summary and a reference summary. Commonly, two metrics used are ROUGE-N, that calculates the precision, recall, and F1-score of n-grams, and ROUGE-L, measuring the longest common subsequence. A well-performance summarization model has a score of over 0.4 for the F1 measure in ROUGE. This can be very helpful when judging whether the summaries have actually captured the content of their sources.

2) BLEU (Bilingual Evaluation Understudy): The Bilingual Evaluation Understudy (BLEU) is a metric that is primarily used to evaluate machine translation but can also be used for summarization evaluation. This metric measures the accuracy of n-grams that appear in the generated text relative to the reference texts. The best possible BLEU score is 1.0, but practical values are usually between 0.2 and 0.4 for good summarization models. Higher scores indicate better agreement with the references in terms of word choice and sentence structure.

3) BERTScore: the Bidirectional Encoder Representations from Transformers Score (BERTScore) depends on contextual embeddings from models like BERT to compare the generated text and reference text. It has been designed to avoid all the drawbacks of traditional n-gram-based metrics by taking se- mantic meaning into consideration. Its score is between 0 and 1, which means a score of 1 would mean an absolute match with the reference. A good range for BERTScore is usually between 0.7 and 0.9, implying high semantic similarity.

4) Signal-to-Noise Ratio (SNR): The Signal-to-Noise Ratio (SNR) is the measure of how clear the audio or speech signal is by how much signal is there as compared to background noise. The higher the SNR,

the clearer the audio will be. A higher value means better quality. SNR values in speech synthesis or recognition tasks range between 20 dB and 40 dB. Any value greater than 30 dB is generally considered high-quality audio, while lower values are usually noisy or poor quality.

5) Word Error Rate (WER): Word Error Rate (WER) evaluates the precision of speech recognition systems by contrasting the generated transcription with a standard reference transcription. It computes the proportion of words that have been inaccurately recognized. Reduced values are preferable, with an ideal WER equating to 0%. Generally, a satisfactory WER for speech recognition technologies is considered to be below 10%, whereas scores exceeding 20% suggest considerable opportunities for enhancement.

B. Analysis of Text Summarization

The outcome results related to summarization of text also reveal the strength of the model from various assessment parameters. ROUGE scores express the model's capacity in terms of retrieving appropriate significant words and phrases. Thus, ROUGE-1(0.5146) shows that the model had mastery over maintaining relevance of an individual word while the overall outcome of ROUGE 2(0.4387) expressed a good grasp in keeping crucial two-word phrase combinations in view. ROUGE-L (0.4801) shows that the model is able to preserve structural coherence by matching longest common subsequences of reference summaries.

The score as shown in Fig. 3 at 0.3412 demonstrates an acceptable level of efficiency in producing n-grams that sound like the references, so showing an acceptable quality of detail reproduction.

BERTScore (Precison: 0.8735, Recall: 0.8576, F1: 0.8655) is a very high semantic equivalence degree with the reference text because its summarization ability retains the semantics of the original text in the process.

Overall, the results as described in Table I, reflect a balanced performance in word-level and contextual relevance, structural alignment, and semantic preservation, which make the model effective for text summarization tasks.

TABLE I: PERFORMANCE METRICS FOR TEXT SUMMARIZATION

S.No	Metric	Description	Value Obtained
1	ROUGE-1	Measures Unigram (word-level) overlap between the candidate and reference summaries.	0.5146
2	ROUGE-2	Measures Bigram (two-word sequence) overlap.	0.4387
3	ROUGE-L	Measures Longest Common Subsequence (LCS) overlap.	0.4801
4	BLEU	Measures n-gram overlap between candidate and reference summaries.	0.3412
5	BERTScore	Semantic similarity using contextual embeddings from BERT.	Precision: 0.8735 Recall: 0.8576 F1: 0.8655

C. Analysis of Keyword Generation

To compare the effectiveness of KeyBERT with that of the hybrid model KeyBERT-TF-IDF, keyword extraction was performed on a text that was identical in both cases. KeyBERT, by using contextual embeddings obtained through the BERT model, produced a set of keywords that well represented the semantic core of the text as represented in Fig. 4. Certain keywords extracted included 'summarization,' 'learns,' and 'models.' Some of these words were very relevant, but at times appeared too general or overlapping in their contextual application, which could compromise the accuracy of the results when applied to technical content. At another extremum, the combined KeyBERT-TF-IDF model actually capitalized upon an advantage of both approaches combined-from the TF-IDF point of view, first utilized potential keyword ranking based on their Term Frequency and Inverse document frequency. This approach allowed for the elimination of words that were less important and brought to the forefront those that were statistically significant.

After the ranking, the terms obtained were passed on to KeyBERT for further evaluation based on their contextual relevance to the text. The hybrid model then produced a more refined and diverse set of keywords such as ’summaries,’ ’teacher,’ and ’distillation’ that were more accurate and precise in relation to the context of the document. This hybrid methodology did not only enhance the relevance of the discovered keywords but also increased the diversity of the terminology, thus achieving a deeper understanding of the content of the text.

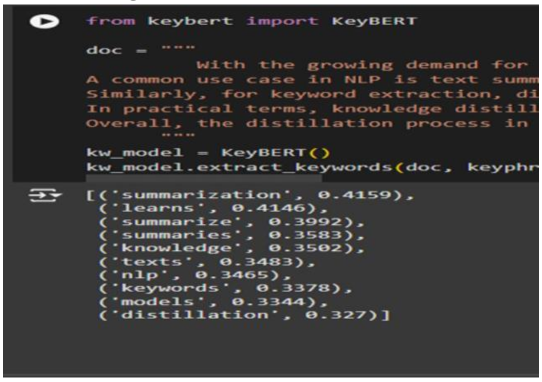

Fig. 3. Keyword Extraction using basic KeyBERT

The evaluation highlights the fact that though KeyBERT alone produces high semantic relevance, the integration of TF-IDF enriches the keyword extraction method by achieving a more balanced and efficient gathering of keywords. This is shown in Fig. 4. This hybrid model’s ability to reconcile statistical significance with semantic context puts it at a great advantage as a robust tool for keyword extraction, particularly in niche areas.

D. Effectiveness of pronunciation check

The Whisper model, enhanced into the Turbo variant, brings about much improvement on the side of speed, reliability, and effectiveness especially concerning real-time transcription activities. The fine-tuning process forms a critical step that requires very precise adjustments in hyperparameters, such as learning rate, batch size, and dropout rate. The process fine-tunes the abilities of the model and ensures it transcribes quicker and with greater accuracy. The dataset for fine-tuning is critical to the realization of these improvements.

Using a carefully curated dataset that includes both high- quality audio and recordings of speech in noisy, real-world environments, the Turbo model learns to detect and suppress unwanted background noise. This ability greatly enhances its ability to focus on the primary speech signal and accurately transcribe verbal communication, even in suboptimal conditions such as those with significant noise or varied accents. The Turbo variant exploits additional optimization strategies that decrease the computational complexity and latency. Thus, it is possible to have real-time transcription with negligible delay, making it very useful in applications requiring prompt feedback. It is optimized not only in terms of processing speed but also concerning the efficient usage of the computational resources, which renders the model more accessible in environments where hardware capabilities are more constrained. In summary, the meticulously optimized Turbo model presents a reliable solution for tasks involving conversion of speech-to- text.

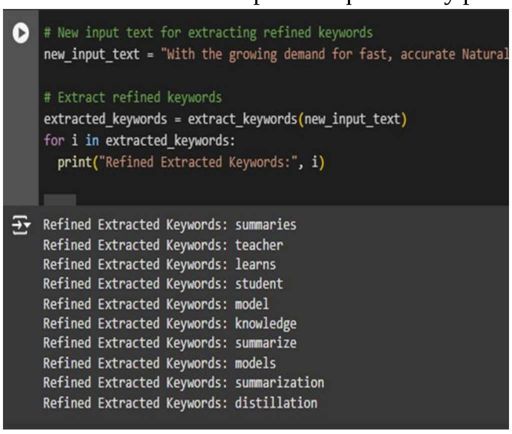

Fig. 4. Keyword Extraction using hybrid approach

E. Performance of Narration

The narration module gives rise to the findings of excellent performance so that clarity is assured, along with a natural speech cadence. Showing an SNR of 8.17, a superior audio clarity is brought about, thereby distinguishing better the speech signal from surrounding noise. With a Word Error Rate of 0.2, the narration output provides an error rate that goes very low; this translates to high accuracy in integrity and pronunciation of words used. In addition, the average pitch of 453.89 Hz implies a definite and stable pitch suitable to engage a younger audience or make the speech more understandable. The overall results as in thereby support the effectiveness of the narration component in offering highly intelligible, clear language that is accurate and also engaging to the user.

V. CONCLUSION AND FUTURE WORK

A. Conclusions

The initiative effectively combines sophisticated natural language processing (NLP) with speech-related technologies to develop a holistic instrument intended to enhance educational results. Through the amalgamation of text summarization, key- word extraction, audio narration, and pronunciation correction, the system presents a novel solution customized to meet a variety of user needs, especially those of children and learners with specific requirements. The PEGASUS-DistilBART summa- rization module produces concise and meaningful summaries, which is supported by performance metrics like ROUGE and BLEU. Furthermore, the hybrid model of KeyBERT-TF-IDF improves learning by extracting keywords that are contextually relevant. The narration module, optimized by gTTS with speech rate reduced and an Indian English accent, produces clean and interesting audio outputs with good performance indicators. Finally, the pronunciation verification module greatly supports the learning process by providing prompt feedback on pronunciation errors. In summary, this initiative underlines the possibilities associated with the incorporation of NLP and speech technologies to improve accessibility, understanding, and involvement in educational contexts. Subsequent research could focus on broadening the dataset, facilitating multilingual capabilities, and advancing real-time efficacy, thereby rendering it a scalable option for wider applications.

B. Future Works

The current project lays down a solid foundation for enhancing accessibility and learning outcomes through NLP and speech technologies. Future developments are expected to enhance system capabilities, with initial directions towards multilingual support to accommodate a number of languages, thus ensuring an expanded scope for text summarization, keyword extraction, and pronunciation verification. Improved real time performance will also reduce latency for applications that require fast feedback. Expanding and diversifying datasets will improve performance on more challenging tasks, such as those noisy or accentuated environments. Integration of educational tools into the program will enhance the adaptability of the learning experiences for an audience. Furthermore, state-of- the-art pronunciation feedback will provide adapted guidance that monitors the progress of users effectively. A mobile app- based approach will enable deployment in challenging, low resource environments, expanding accessibility. Furthermore, the improvement on the user interface will come in handy to enhance usability among others especially children and people with disabilities. Good evaluation and feedback mechanisms are necessary for continuous system upgrading to ensure scalability, effectiveness, and flexibility in responding to these evolving educational and accessibility needs.

REFERENCES

[1] L. Haz, E. D. Fajrianti, N. Funabiki, and S. Sukaridhoto, ”A Study of Audio-to-Text Conversion Software Using Whispers Model,” in Proc. Sixth Int. Conf. on Vocational Education and Electrical Engineering (ICVEE), Surabaya, Indonesia, 2023, pp. 268-273.

[2] A.Raj, M. Raj, N. Umasankari, and D. Geethanjali, ”Document-Based Text Summarization using T5 small and gTTS,” in Proc. Int. Conf. on Advances in Data Engineering and Intelligent Computing Systems (ADICS), Chennai, India, 2024, pp. 1-6.

[3] Alokla, W. Gad, W. Nazih, M. Aref, and A.-B. M. Salem, ”Pseudocode Generation from Source Code Using the BART Model,” Mathematics, vol. 10, p. 3967, 2022. doi:10.3390/math10213967.

[4] Issa, M. B. Jasser, H. N. Chua, and M. Hamzah, ”A Comparative Study on Embedding Models for Keyword Extraction Using KeyBERT Method,” in Proc. IEEE 13th Int. Conf. on System Engineering and Technology (ICSET), Shah Alam, Malaysia, 2023, pp. 40-45.

[5] S. Mukhtar, C. C. Primadani, S. Lee, and P. Jung, ”A Comparison of Summarization Methods for Duplicate Software Bug Reports,” Electron- ics, vol. 12, p. 3456, 2023.

[6] O. P. Orochi and L. G. Kabari, ”Text-to-Speech Recognition using Google API,” Int. J. of Computer Applications, vol. 183, no. 15, pp. 18-20, Jul. 2021.

[7] R. Harshitha, B. Lakshmipriya, and V. Krishnamurthy, ”TransEffiVisNet – An Image Captioning Architecture for Auditory Assistance for the Visually Impaired,” Multimedia Tools and Applications, 2024, pp. 1-20.

[8] R. Patnoorkar, S. Chaudhary, S. Pandey, P. S. Pandey, R. Balyan, and M. Kumar, ”Assistive Technology Intervention in Dyslexia Disorder,” in Proc. Int. Conf. on Artificial Intelligence and Smart Communication

(AISC), Greater Noida, India, 2023, pp. 343-347.

[9] A.Rao, ”Transcribing Educational Videos Using Whisper: A Prelim- inary Study on Using AI for Transcribing Educational Videos,” arXiv preprint, 2023. doi:10.48550/arXiv.2307.03200.

[10] S. S. A. Hamid, N. Admodisastro, and A. Kamaruddin, ”A Study of Computer-Based Learning Model for Students with Dyslexia,” in Proc. 9th Malaysian Software Engineering Conf. (MySEC), Kuala Lumpur, Malaysia, 2015, pp. 284-289.

[11] S. Mussakhojayeva et al., ”Noise-Robust Multilingual Speech Recognition and the Tatar Speech Corpus,” in Proc. Int. Conf. on Artificial Intelligence in Information and Communication (ICAIIC), Osaka, Japan, 2024, pp. 732-737.

[12] S. Wang, C.-H. Yang, J. Wu, and C. Zhang, ”Can Whisper Perform Speech-Based In-Context Learning?,” in Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Korea, 2024,

[13] M. Davoudi, M. Ghavipour, M. Sargolzaei-Javan, and S. Dinparast, ”Decentralized Storage Cryptocurrencies: An Innovative Network-Based Model for Identifying Effective Entities and Forecasting Future Price Trends,” Computational Economics, 2024, pp. 1-46.

[14] G. Bao and Y. Zhang, ”A General Contextualized Rewriting Framework for Text Summarization,” IEEE/ACM Trans. on Audio, Speech, and Language Processing, vol. 31, pp. 1624-1635, 2023.

[15] J. Zhang, Y. Zhao, M. Saleh, and P. J. Liu, ”PEGASUS: Pre-training with Extracted Gap-Sentences for Abstractive Summarization,” in Proc. 37th Int. Conf. on Machine Learning (ICML’20), vol. 119, JMLR.org,

[16] K. Radhika, C. N. Reddy, K. Vishwanatha, and V. Mishra, ”An Efficient Text Synthesization Method by Utilizing Text Mining Techniques and Natural Language Processing for the Usage of Intellectually Disabled Individuals,” in Proc. 5th Int. Conf. on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2021, pp. 39.

[17] M. Asmitha, C. R. Kavitha, and D. Radha, ”Summarizing News: Unleashing the Power of BART, GPT-2, T5, and Pegasus Models in Text Summarization,” in Proc. 4th Int. Conf. on Intelligent Technologies (CONIT), Bangalore, India, 2024, pp. 1-6.

[18] M. Luo, B. Xue, and B. Niu, ”A Comprehensive Survey for Automatic Text Summarization: Techniques, Approaches and Perspectives,” Neu- rocomputing, vol. 603, 2024