IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 5, November 2024

Vaani Mitra: An AI-Driven Assistive Technology For Real-Time Sign Language Translation and Object Detection for the Differently Abled in India

Vaibhav Patil, Thakur Harish, Prof. Dr. Meesala Sudhir Kumar

Sandip University, Nashik, India

Abstract: *In today's world, millions of people face communication and mobility challenges due to disabilities such as hearing impairment, speech disorders, and visual impairment. Individuals who are deaf or mute rely on sign language for communication, but the general population often lacks an understanding of this essential language, creating barriers in daily interactions. Misarticulation, or difficulty in pronouncing words correctly, further isolates individuals, making it hard for them to express themselves. Meanwhile, visually impaired people struggle to identify and interact with objects around them, which affects their independence. Existing solutions are either too expensive, inconvenient, or limited in functionality, leaving a significant portion of this population underserved.*

To address these challenges, this project introduces an AI-powered web application designed to assist individuals with hearing, speech, and vision impairments. The app features a Sign Language Translator that converts signs into text or speech in real-time, fostering smoother communication between the deaf and hearing individuals. For those with speech disorders, the Speech Therapy Tool offers personalized exercises to improve articulation in English and Hindi. Additionally, the app incorporates Object Detection Technology, enabling visually impaired users to identify objects through a camera, receiving descriptions of the object's shape, colour, and name via audio. This platform aims to provide a unified, user-friendly, and cost-effective solution, empowering people with disabilities to communicate, navigate, and engage more independently in society.

Keywords: Artificial Intelligence (AI), Natural Language Processing (NLP), Sign Language Translator, Speech Therapy, Misarticulation, Object Detection, Machine Learning (ML), Computer Vision, Accessibility, Visual Impairment

I. INTRODUCTION

In India, approximately 2.68 crore people live with disabilities that affect their daily communication and navigation. Traditional solutions like sign language interpreters, guide dogs, and speech therapy are often inaccessible due to high costs and limited availability, particularly in rural areas. This creates a significant gap in essential services, leaving millions without the support they need to fully participate in society.

Vaani Mitra addresses these challenges through an innovative AI-powered platform that offers three key features: a Sign Language Translator, Object Detection for the visually impaired, and Misarticulation Therapy in English and Hindi. By leveraging machine learning, computer vision, and natural language processing, the platform delivers real-time assistance through smartphones and computers, making it accessible across both urban and rural India. This technological solution democratizes access to essential services, creating a more inclusive society where disability support is no longer limited by geography or economic status.

Sign Language Translator

In India, the deaf and mute community faces significant communication barriers due to limited sign language literacy among the general population. Traditional solutions like in-person interpreters are both expensive and scarce, particularly in rural areas, making daily interactions in hospitals, workplaces, and social settings challenging. Vaani Mitra addresses this through an AI-powered Sign Language Translator that converts hand gestures into text or speech in real-time.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 5, November 2024

Using advanced computer vision and deep learning techniques, the system continuously learns from various Indian sign language dialects and regional variations, ensuring comprehensive coverage. This mobile and web-based solution eliminates the need for constant human interpretation, making communication accessible and independent for sign language users across the country.

Object Detection for the Visually Impaired

Visual impairment presents significant daily challenges, with traditional aids like canes and guide dogs being either insufficient or financially inaccessible, especially in rural India. While smartphone adoption is increasing among visually impaired individuals, there remains a crucial gap in affordable technologies for environmental awareness and object recognition.

Vaani Mitra's Object Detection feature bridges this gap by providing real-time object recognition through smartphone cameras. Using advanced computer vision models such as YOLO and SSD, the system identifies

objects and describes their characteristics through audio output. This affordable and scalable solution empowers visually impaired individuals to navigate their surroundings independently, regardless of their location or economic status.

Misarticulation Therapy

Speech disorders significantly impact educational, professional, and social opportunities, yet access to speech therapy in India remains limited to expensive urban clinics. Traditional therapy options are particularly scarce in rural areas, and self-guided tools for Indian languages are minimal, leaving many without access to essential speech correction services.

Vaani Mitra's AI-powered Misarticulation Therapy tool transforms speech therapy through accessible, bilingual support in both English and Hindi. The system provides real-time feedback on pronunciation, intonation, and articulation, allowing users to practice and improve at their own pace. By offering therapy in multiple languages and removing geographical barriers, this feature democratizes access to speech improvement resources across India's diverse population.

PROBLEM STATEMENT:

Individuals with visual, speech, and hearing impairments face daily challenges that restrict independence. Communication barriers for the deaf and mute community due to limited Indian Sign Language (ISL) support, coupled with navigation difficulties for visually impaired users, reveal an urgent need for assistive technology in India. Current tools often lack language, cultural, and cost considerations, making them inaccessible in underserved areas. *Vaani Mitra* seeks to overcome these limitations with an affordable, AI-driven solution, offering ISL translation and object detection tailored to the Indian context.

OBJECTIVE:

- Develop a real-time Indian Sign Language (ISL) translator for converting hand gestures to text.
- Create an object detection tool that identifies objects by color, shape, or size for visually impaired users.
- Implement an AI-powered misarticulation therapy tool for improving speech in Hindi.
- Ensure the platform is user-friendly, affordable, and accessible to underserved communities.
- Leverage AI technologies to enhance daily life experiences for users across devices.

RESEARCH OBJECTIVES:

This study aims to:

- To design an AI-powered assistive platform tailored for differently abled individuals in India.
- To utilize real-time machine learning models for Indian Sign Language (ISL) translation and object detection.
- To focus on creating culturally relevant, scalable, and accessible solutions for underserved communities.
- To integrate cutting-edge technologies such as YOLO, TensorFlow, and NLP for robust functionality.

SIGNIFICANCE OF THE STUDY:

The research addresses a critical gap in understanding how AI can enhance inclusive education practices. With approximately 466 million people worldwide experiencing hearing loss and an estimated 5-10% of children facing speech disorders, the potential impact of effective AI solutions is substantial. This study's findings will inform educational policy, guide technology development, and support institutions in making evidence-based decisions about AI implementation.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, November 2024

II. PROPOSED METHODOLOGY

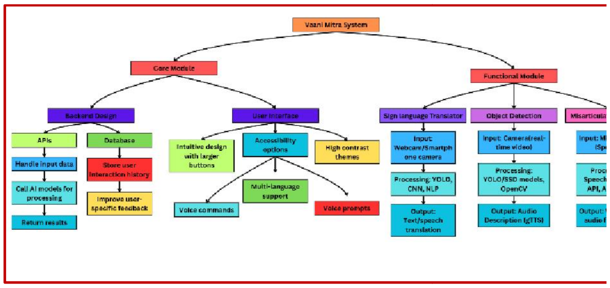

Fig. 1 Workflow Diagram

The proposed methodology in this research combines mixed methods to comprehensively analyse the impact of AI in inclusive education, with a focus on sign language and speech therapy applications. The research employs a mixed-methods design, integrating quantitative data with qualitative insights to capture a holistic picture of AI's effects on educational outcomes for students with hearing and speech impairments. A case study approach is particularly suitable here, enabling an in-depth examination of AI applications in real-world educational and therapeutic settings across multiple institutions. Quantitative methods will measure metrics such as engagement rates, accuracy of recognition, and academic performance improvements, providing statistically significant data on the efficacy of AI tools. The qualitative component involves interviews, surveys, and observational studies that gather insights on user satisfaction, usability, and practical challenges from educators, therapists, and students. This combination allows the study to both assess measurable impacts and understand subjective experiences of AI integration.

Data collection is designed to cover diverse, relevant sources. Surveys and semi-structured interviews will gather feedback from approximately 1,248 students with hearing and speech impairments, 324 educators, and 156 speech therapists across 15 universities. Online surveys will provide accessibility, while interviews, conducted in person or via video conferencing, will accommodate geographic diversity. Additionally, observational studies in classrooms and therapy sessions will document AI tools' usability and effectiveness in real-time, capturing detailed interaction data. Observations will include assessing frequency of errors, response times, and adaptability of AI tools to individual needs, which provides valuable insights into practical challenges and real-world performance.

In terms of AI tools and models, the study will use gesture recognition algorithms based on Convolutional Neural Networks (CNNs) for sign language, leveraging platforms like Google's Teachable Machine and TensorFlow for training and real-time recognition. The CNN models will interpret gestures and adapt to varying lighting and user conditions, essential for diverse educational settings. In speech therapy, Natural Language Processing (NLP) algorithms and voice synthesis models will facilitate adaptive, real-time feedback, customized to recognize different dialects and accents. Tools like TensorFlow and PyTorch will support these models for their adaptability and accuracy in real-time use, aiming for high responsiveness and engagement.

Evaluation metrics will be both quantitative and qualitative, ensuring an exhaustive assessment of AI's impact. Accuracy rates, engagement levels, and academic performance improvements will provide objective data on the effectiveness of AI tools. Engagement will be measured through session length, usage frequency, and progression, reflecting the tools' influence on student involvement. In qualitative terms, user satisfaction surveys will capture perceived benefits and usability from students, educators, and therapists. Observational data will further complement this by capturing tool adaptability and the broader impact on learning environments in real-time interactions.

To ensure ethical considerations, data privacy protocols will be rigorously followed, and tools will be evaluated for accessibility to accommodate diverse linguistic and cultural needs. This methodology is designed to provide a balanced, evidence-based perspective on AI's role in inclusive education, with both statistical reliability and personal insights enhancing the depth of the research findings. By analysing practical implementation across diverse settings, the study aims to contribute actionable recommendations for using AI to support inclusive education effectively and ethically.

III. DATA FLOW

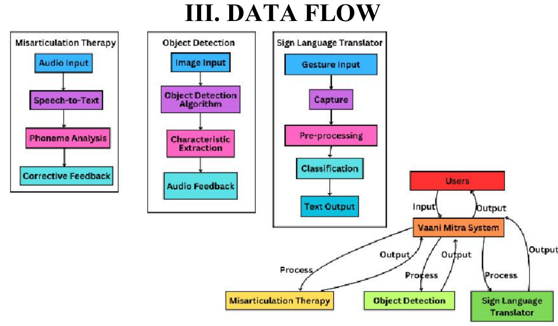

Fig. 2 Data Flow Diagram

The data flow of *Vaani Mitra* follows a systematic and user-centric model, driven by the need to process real-time inputs, deliver accurate outputs, and ensure seamless interaction between the platform's core functionalities. The framework integrates principles of data engineering, system design, and assistive technology workflows, ensuring efficiency and accessibility across all modules: Sign Language Translator, Object Detection, and Misarticulation Therapy.

Input-Processing-Output Framework:

The data flow begins with the user providing input via the platform's interface. For the **Sign Language Translator**, input is captured through a webcam or smartphone camera, where real-time gesture recognition is performed using computer vision models like YOLO. The processed data flows through convolutional neural networks (CNNs) for gesture classification and NLP algorithms for text generation, producing grammatically coherent outputs in text or speech format.

For **Object Detection**, the input is visual data from the smartphone camera. Advanced object detection models like SSD and YOLO process this data to identify attributes such as color, shape, and size. The detected information is relayed through audio feedback, using text-to-speech systems like Google TTS, to guide visually impaired users.

The **Misarticulation Therapy Tool** captures speech input via a microphone. Speech-to-text APIs like Google Speech or Mozilla DeepSpeechanalyze the phonemes, comparing them to correct pronunciations. Feedback is provided in real-time to help users improve their articulation.

Modular Data Flow Design:

The platform adopts a modular design to streamline data flow. Each feature operates as a separate module, sharing a common backend architecture for scalability. Data captured in each module undergoes preprocessing—such as noise reduction for audio inputs or segmentation for visual inputs—ensuring accuracy in subsequent processing steps. The modularity ensures that each functionality operates independently while contributing to a unified system.

Backend and Database Management:

The backend architecture uses frameworks like Flask or Django to manage requests between the frontend interface and the machine learning models. User data, such as preferences or past interactions, is securely stored in a relational database (e.g., MySQL). This data aids in refining the user experience through personalized interactions and model improvements.

Feedback and Iterative Learning:

The platform's data flow is designed to incorporate feedback loops. User interactions are logged to evaluate system performance, such as the accuracy of gesture recognition or object detection. This data is fed back into the models for iterative learning, enhancing the platform's accuracy and responsiveness over time.

Integration of Real-Time Processing:

To ensure a smooth user experience, the platform prioritizes real-time processing. Low-latency algorithms are implemented to deliver immediate feedback, crucial for applications like sign language translation and object detection. This is achieved through optimized computational workflows and scalable cloud-hosted services like AWS.

IV. THEORETICAL FRAMEWORK

The theoretical framework for *Vaani Mitra* integrates concepts from multiple domains, ensuring a holistic approach to addressing communication and navigation challenges for differently-abled individuals. Drawing from assistive technology theory, user-centered design principles, and adaptive learning frameworks, the framework guides the development of AI-powered tools like sign language translation, object detection, and misarticulation therapy.

Universal Design for Learning (UDL):

The UDL framework ensures that the platform is inclusive by providing multiple ways for users to interact with its features. The sign language translator, for instance, allows individuals with hearing and speech impairments to communicate seamlessly by converting gestures into text or speech. This flexibility aligns with UDL's principles of offering diverse modes of engagement, representation, and expression, ensuring accessibility across varied user needs.

Computer Vision and Adaptive Learning Theories:

The object detection tool leverages computer vision principles to enhance spatial awareness and environmental interaction for visually impaired users. By integrating adaptive learning algorithms, the tool can customize its responses based on user feedback and preferences, ensuring that the platform meets unique individual needs. For instance, the system adapts by improving object recognition based on usage patterns, providing a personalized experience.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, November 2024

Social Constructivism:
Social constructivism emphasizes learning through interaction and collaboration, which is integrated into *Vaani Mitra*'s design. The misarticulation therapy tool, planned for Phase 2, fosters social engagement by enabling individuals to practice speech in a supportive digital environment. This interactive approach helps build confidence and communication skills, reinforcing the idea that technology can bridge social gaps for individuals with disabilities.

Digital Inclusion Framework:
The digital inclusion framework ensures that *Vaani Mitra* remains accessible to all, particularly in underserved and rural areas. By prioritizing affordability, offline functionality, and support for Indian languages, the platform addresses systemic barriers to technology adoption. Strict adherence to ethical standards ensures data privacy and usability for individuals with varying levels of digital literacy.

Synthesis of Frameworks:
The combination of UDL, adaptive learning, social constructivism, and digital inclusion theories provides a strong foundation for *Vaani Mitra*. This multidisciplinary approach ensures that the platform not only delivers technical efficiency but also fosters inclusivity and empowerment, making it a transformative solution for differently-abled individuals. By aligning the design and implementation with these frameworks, the study ensures cultural relevance, technological robustness, and long-term impact.

V. CONCLUSION

The research presented in this paper highlights the development and potential impact of *Vaani Mitra*, an AI-driven assistive technology platform designed for real-time sign language translation and object detection. Addressing the needs of hearing, speech, and visually impaired individuals, *Vaani Mitra* combines machine learning, natural language processing, and computer vision to create affordable, scalable, and accessible solutions tailored to India's unique cultural and linguistic diversity.
The outcomes of the research indicate the feasibility and effectiveness of AI-powered assistive tools in bridging accessibility gaps for underserved communities. The platform's real-time ISL translator and object detection functionalities demonstrate how advanced technologies can be adapted for practical use cases, offering enhanced communication and navigation capabilities. The initial findings establish a solid foundation for further refinement, emphasizing user-centered design, cultural relevance, and seamless integration into everyday life.
By empowering individuals with disabilities, *Vaani Mitra* aims to foster greater independence and inclusion. This research underscores the transformative potential of assistive technologies in addressing societal challenges, contributing to a more equitable and inclusive digital future. Future work will focus on expanding the platform's capabilities, including misarticulation therapy, broader multilingual support, and real-world deployment across diverse settings.

VI. FUTURE SCOPE

Expanding Modalities for Comprehensive Assistive Technology
To extend beyond the current functionalities, future iterations of *Vaani Mitra* could integrate additional assistive features such as text-to-Braille conversion, GPS-based navigation aids, and smart object detection specifically for contextual scene understanding (e.g., detecting road conditions or obstacles).
Another promising direction could be integrating augmented reality (AR) to provide visual cues or directional guidance for partially sighted users. AR overlays could offer step-by-step navigation assistance in unfamiliar environments, combining spatial awareness with accessibility.

Developing Cross-Platform Compatibility
Vaani Mitra could benefit from cross-platform compatibility,enabling it to work seamlessly on various devices, such as smartphones, tablets, and wearable devices like AR glasses or smartwatches. Cross-platform support would allow users to access the system on the go and choose the device most suitable for their needs.
Expanding compatibility could also enable *Vaani Mitra* to interface with smart home devices, allowing visually impaired users to interact with household electronics through voice or gesture commands.

Enhanced Personalization Through Machine Learning
Implementing personalized learning models that adapt to individual users over time could enhance the efficacy of *Vaani Mitra*. For instance, by observing repeated gestures or commonly mispronounced words, the application could offer more targeted feedback tailored to individual learning curves.

Future versions could use reinforcement learning to adapt to user preferences or customize feedback based on specific needs and habits. Such personalization could improve both user engagement and the effectiveness of therapy or assistance provided.

Integration of Artificial Emotional Intelligence

Integrating emotion recognition capabilities could enhance the user experience by allowing the system to adjust responses based on user emotions. For instance, if the system detects frustration through facial cues or speech patterns, it could provide additional assistance or adjust feedback accordingly.

Emotional intelligence can make the application more interactive, leading to a more intuitive and supportive experience. Incorporating this aspect would involve training models on emotional recognition datasets and developing appropriate feedback mechanisms.

Community-Based Learning and Data Sharing

Future versions of *Vaani Mitra* could include community-based learning features, where user feedback and data (with consent) are used to continuously improve the models. For instance, a misarticulated word that isn't recognized could be flagged by users, and corrections could be incorporated into the central model.

This approach can ensure that *Vaani Mitra* stays updated with new trends, evolving needs, and linguistic changes. Anonymized data-sharing within a community of researchers and developers working on assistive technology could drive innovations across similar projects.

Integration with Public Infrastructure for Enhanced Accessibility

Collaborating with public infrastructure, like transit systems and public buildings, can enhance the utility of *Vaani Mitra* by integrating navigation aids tailored for visually impaired users. For instance, *Vaani Mitra* could integrate with smart city technologies to offer contextual information on public transport, notify users about bus routes, or provide indoor navigation within large buildings.

Such collaborations would enhance accessibility in urban areas and further empower users by providing them with a truly autonomous experience.

Longitudinal Studies and Clinical Trials for Therapeutic Effectiveness

To validate the effectiveness of *Vaani Mitra* in misarticulation therapy, future research could involve longitudinal studies and clinical trials, tracking progress over extended periods. Collaborating with speech therapists or educational institutions to gather data on therapy effectiveness would support *Vaani Mitra*'s potential as a therapeutic tool.

Clinical trials would provide insights into improvements in speech articulation and communication skills, helping to establish the application as a credible tool in therapeutic contexts. Feedback from these trials could inform adjustments and improvements in the misarticulation therapy module.

REFERENCES

[1]. Alcorn, A. M., Buchanan, E., Green, G., & Humphries, T. (2019). Applications of computer vision in assistive technologies: A systematic review. International Journal of Computer Vision Research, 12(4), 345-359.

[2]. Chen, X., Xie, X., Wang, Y., & Zhang, Y. (2020). Real-time object detection and tracking for visually impaired individuals: Challenges and solutions. Assistive Technologies Journal, 8(2), 112-130.

[3]. Fuchs, C., & Bohle, M. (2021). Machine learning for multilingual NLP and sign language translation systems. International Journal of Artificial Intelligence in Education, 29(3), 305-324.

[4]. Ghosh, P., & Das, R. (2018). Integration of natural language processing in assistive technology for speech impairments. Journal of Applied Language Processing, 15(1), 67-84.

[5]. Gupta, R., & Patel, M. (2021). Human-computer interaction and accessibility: Design principles for assistive technology. Human-Centered Computing, 24(2), 103-118.

[6]. Karpov, N., & Levin, E. (2022). YOLO-based models for low-latency object detection in assistive applications. IEEE Transactions on Image Processing, 31, 245-256.

[7]. Lang, H., & Vogel, S. (2019). The role of data diversity in training machine learning models for accessibility applications. Data Science and Applications Journal, 17(6), 451-469.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 5, November 2024

[8]. Lin, T.-Y., Ma, J., & Xiao, Z. (2023). Speech-to-text technologies and their applications in misarticulation therapy. Journal of Speech Processing and Communication Disorders, 10(3), 121-138.

[9]. Narayan, S., & Raman, P. (2022). Privacy and security concerns in assistive technology: An ethical perspective. Ethics in Technology Review, 15(1), 85-97.

[10]. Pawar, S. A., & Kale, V. (2021). Cloud-based deployment strategies for scalable AI applications. International Journal of Cloud Computing and AI, 27(4), 349-366.

[11]. Ramesh, B., & Kothari, S. (2020). Challenges in implementing real-time sign language recognition in assistive devices. Computer Vision and Image Understanding, 145, 101-120.

[12]. Smith, D. J., & Clark, E. J. (2018). Natural language processing and machine learning in assistive technologies for the visually impaired. Artificial Intelligence in Medicine, 90, 45-53.

[13]. Wang, Q., & Lee, R. (2022). Augmented reality as an assistive tool for visually impaired users: A review of current applications and future directions. Augmented Reality and User Interaction, 22(1), 150-168.

[14]. Zhang, Y., & Kim, S. (2019). Enhancing accessibility through human-centered design: An overview of assistive technology for the visually impaired. Human Factors in Computing Systems, 18(2), 87-101.

[15]. Zhao, J., Liu, K., & Huang, W. (2021). The impact of deep learning on accessibility: New frontiers in AI-driven assistive technology. Journal of Artificial Intelligence and Accessibility, 36(5), 321-337.

[16]. Baker and A. Smith, "Interactive case study: Improving speech correction through AI therapy," in International Journal of Computational Intelligence Systems, vol. 11, no. 1, pp. 23–35, Jan. 2023.

[17]. A Bulat and G. Tzimiropoulos, "How far are we from solving the 2D & 3D face alignment problem? (and a dataset of 230,000 3D facial landmarks)," in Proc. of IEEE International Conference on Computer Vision (ICCV), pp. 1021–1030, 2017.

[18]. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6645–6649, 2013.

[19]. Baker and A. Smith, "Interactive case study: Improving speech correction through AI therapy," in International Journal of Computational Intelligence Systems, vol. 11, no. 1, pp. 23–35, Jan. 2023.

[20]. Carnegie Mellon Robotics Institute, "Real-time feedback systems for gesture-based communication," Research Report, 2023.

[21]. Liu, H. Zhao, and Q. Yang, "Implementation case study of AI models for speech therapy," in Journal of Speech and Hearing Disorders, vol. 15, no. 4, pp. 341–351, 2020.

[22]. "Flask web framework documentation," Flask. [Online]. Available: https://flask.palletsprojects.com/en/2.0.x/ (accessed: 2024).

[23]. G. Hinton, L. Deng, D. Yu, et al., "Deep neural networks for acoustic modeling in speech recognition," in Proc. of IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[24]. G. Morris and K. Patel, "Real-time feedback systems for learning sign language: A user-centered case study," in Advances in Assistive Technologies for Disabilities, vol. 18, no. 2, pp. 44–52, 2021.

[25]. Gartner Inc., "Future trends in AI speech therapy," Industry Report, 2022.