UNSIGHT: BRIDGING THE EDUCATIONAL GAP FOR BLIND LEARNERS

Aishani, Pandey[1]*; Arush, Sachdeva*[2]*; Vivek, Mathur*[1]

[1,2,3]International Institute of Information Technology, Hyderabad

ABSTRACT

Visually impaired learners face significant challenges in accessing quality education due to the lack of accessible educational materials and adaptive learning environments. Traditional educational methods often fall short in addressing the unique needs of these learners, leading to a gap in educational outcomes and opportunities. UnSight aims to enhance the educational experience for visually impaired learners through a virtual education platform that integrates advanced AI technologies. This platform leverages speech and text processing, image and text integration, and voice-to-action commands to address accessibility gaps. Emphasizing adaptive learning and personalized feedback, the proposed system contributes to SDG 4: Quality Education. This paper discusses the technical architecture, standardization efforts, scalability, and ethical considerations involved in developing and deploying UnSight.

Keywords - Virtual Education, Accessibility, AI, Speech Processing, Image Recognition, Voice Commands, Visually Impaired, SDG 4

1. INTRODUCTION

Education is a fundamental human right, and ensuring inclusive and equitable quality education for all is a pressing global challenge. The United Nations' Sustainable Development Goal 4 (SDG 4) emphasizes the importance of providing equal access to quality education, promoting lifelong learning opportunities, and building inclusive and effective learning environments. However, individuals with visual impairments often face significant barriers in accessing educational resources and participating in traditional learning environments.

The emergence of virtual worlds and immersive technologies presents a unique opportunity to revolutionize the educational landscape, particularly for underserved communities. By leveraging the power of artificial intelligence (AI) and advanced technologies, we can create accessible and inclusive learning environments that cater to the diverse needs of visually impaired students.

This research paper introduces an AI-driven virtual education platform that aims to empower visually impaired students by providing an immersive, accessible, and personalized learning experience. Through the integration of cutting-edge technologies such as speech recognition, emotion detection, image description, and voice-to-action commands, the proposed system offers a comprehensive solution that addresses the unique challenges faced by visually impaired learners.

2. RELATED WORK

Existing solutions in the market, such as Amberscript, Jobma, and Seeing AI by Microsoft, have made strides in providing accessible tools for individuals with disabilities. However, these solutions often lack a holistic integration of features or are not specifically tailored for visually impaired students in an educational context.

Amberscript excels in speech-to-text transcription but lacks features that convert speech into actions or commands. Jobma automates video interviews for the HR industry but may lack personalization and accessibility features for educational purposes. Seeing AI by Microsoft offers visual scene description capabilities but does not address the broader educational needs of visually impaired students.They are working on the ORBIT project to develop a large dataset for personalized object recognition, enhancing the capabilities of AI in this area. Microsoft's commitment includes expanding accessibility features in Microsoft 365 and tools like Accessibility Insights for developers to improve UI accessibility.

Furthermore, while initiatives like those by I-Stem aim to revolutionize accessibility for visually impaired students, they often rely on separate assistive technologies rather than integrating accessibility features seamlessly into existing digital ecosystems. The proposed system addresses these gaps by offering a unified platform that integrates these accessibility features.

IBM has created numerous adaptive and assistive technologies, such as the Braille Translation System and screen readers for DOS. IBM's efforts include the development of the Home Page Reader and the NavCog app, which helps people with vision loss navigate public spaces using AI and robotics. IBM continues to innovate in accessibility through initiatives like the Content Clarifier, which simplifies digital content based on individual preferences.

In the startup scene, companies like Be My Eyes [4]have made a substantial impact. Be My Eyes connects blind or low-vision individuals with sighted volunteers through a smartphone app to provide visual assistance for various tasks. This crowdsourced model has proven both successful and scalable, demonstrating the power of community-driven solutions.

Aira[5] is another innovative startup, offering a subscription service that connects blind users with trained agents via smart glasses or a smartphone app. These agents provide real-time assistance for navigation, reading, and other tasks.

Research also plays a critical role. Studies like those from McKinsey Co., in collaboration with Tilting the Lens [6], highlight the digital divide and emphasize the importance of digital accessibility for blind and low-vision consumers. Their research involves accessible digital surveys, online focus groups, and interviews, underscoring the need for inclusive design in digital products

One notable study, "Explaining CLIP's Performance Disparities on Data from Blind/Low Vision Users," [7] evaluates large multi-modal models (LMMs), particularly CLIP, and highlights the necessity of datasets that accurately represent blind and low-vision users to improve model performance. Another study, "ImageAssist: Tools for Enhancing Touchscreen-Based Image Exploration Systems for Blind and Low Vision Users," [8] investigates touchscreen-based systems that enrich alt text to offer a deeper understanding of digital images through touch and audio feedback.

The paper "Artificial Intelligence for Visually Impaired" [9] reviews 178 studies on AI applications designed to aid visually impaired individuals, including deep learning methods for diagnosing eye diseases and enhancing accessibility technologies . "Towards Assisting Visually Impaired Individuals: A Review on Current Solutions" [10] provides a comprehensive overview of assistive technologies such as AI-driven navigation, object recognition, and reading assistance . Lastly, "Recent Trends in Computer Vision-Driven Scene Understanding for Visually Impaired" systematically maps the current state-of-the-art in computer vision-based assistive technologies, focusing on scene understanding and navigation solutions for visually impaired users [11].

3. RESEARCH METHODOLOGY

The proposed system distinguishes itself through a suite of innovative features designed to provide a comprehensive and accessible educational experience for visually impaired individuals. These features leverage advanced AI technologies to ensure inclusivity, adaptability, and enhanced user engagement. Below is a detailed exploration of our core features:

3.1 Speech and Text Processing

Speech Recognition: The system begins by capturing the user's voice input using the Web Speech API, which transcribes spoken language into text. This enables users to communicate with the application naturally, facilitating an intuitive and seamless interaction.

Emotion Recognition for Optimized Responses: The proposed system employs AI models capable of recognizing speech tonality and emotional cues using Mel-Frequency Cepstral Coefficients. By analyzing these cues, the platform adapts its responses to better suit the user's emotional state, making interactions more empathetic and contextually relevant. This ensures that users receive supportive feedback tailored to their current mood, enhancing their overall learning experience. [12]

Internet Search Integration: For information retrieval, the system employs the Google Custom Search JSON API. It processes the transcribed text as a search query and fetches relevant information from the web, ensuring that responses are informative and contextually relevant.[13]

Response Formulation: The system generates a response that incorporates both the informational content obtained from the web and the emotional context detected from the user's speech. This approach provides a personalized interaction, ensuring that the user receives accurate and contextually appropriate information.[14]

Text-to-Speech Output: The final response is converted back into speech using advanced text-to-speech technology, allowing the system to communicate with the user audibly. This feature is crucial for accessibility, particularly for visually impaired users, enabling them to receive information in an auditory format.[15]

User Interface (UI): The UI displays the transcription and processed responses with visual cues indicating which parts of the response have been spoken. This not only aids the helper in following the response but also enhances comprehension. The interface is designed to be intuitive and user-friendly, ensuring that visually impaired users can navigate and interact with the platform effectively.[16]

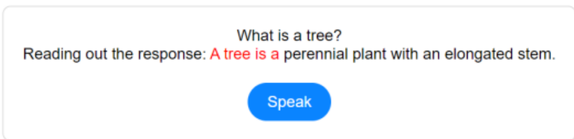

Figure 1 – UI for Speech Processing

System Flow:
The speech processing system operates through a structured interaction between the client and server sides. Initially, the user's speech is captured and sent via the client-side Speech-to-Text API to the server. The server processes the text through an Emotion Content Analyzer, which identifies emotional cues and context. This analysis informs the Adaptive Learning Model, which generates an appropriate response. The Text-to-Speech Service then converts the tailored response back into speech. Finally, the audio is

transmitted back to the client side, where it is delivered to the user, completing the interaction cycle.

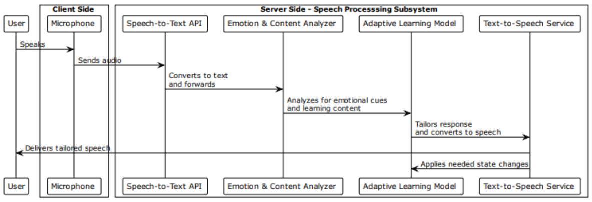

Figure 2 – Activity Diagram for Speech and Text Processing

3.2 Image and Text Integration

Image Scene Descriptor for the Visually Impaired:

The proposed system integrates advanced image recognition and natural language processing (NLP) to create detailed audio descriptions of virtual scenes and images.[18] This feature allows blind users to understand visual content through auditory descriptions, significantly enhancing their ability to engage with educational materials that contain images.[17] It is important to note that the proposed system, although designed for visually impaired users, will still require assistance from sighted individuals for accurate image and text integration. The involvement of sighted assistants ensures that the descriptive content generated is accurate and contextually relevant, thereby enhancing the overall learning experience.

Image Bounding Box for Learning Material: An algorithm within The proposed system differentiates between text and visual content in learning materials. Using Optical Character Recognition (OCR) and image-to-text models, the platform converts visual text into speech, making diagrams and other visual elements accessible. This ensures that all content, regardless of its original format, is available to visually impaired users.

Figure 3 – Input and Output to OCR

Handwriting to Text: The system provides auditory instructions for forming each alphabet. Users draw the alphabets based on these instructions using a touch screen device. The system analyzes the drawn alphabet and compares it with predefined templates, providing feedback on accuracy and guiding users to improve their handwriting.

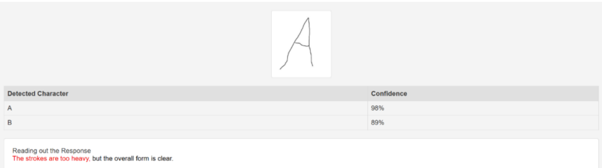

Figure 4 – UI for Handwriting to Text Processing

Face Detection: The face detection module employs the YOLOv5 object detection model to detect faces within an input image. Detected faces are outlined with rectangular bounding boxes, providing visual cues for their location and size. This capability is essential for user identification, interaction facilitation, and security enhancement. [17]

Figure 5 – UI for Face Detection Model

Sign Language to Text: The system detects hand gestures using the HSV color space for skin color detection. It extracts features such as area, perimeter, and convexity defects of the hand contour, identifying extended fingers to interpret sign language. This feature allows real-time translation of sign language into text, facilitating communication for hearing-impaired users. [17]

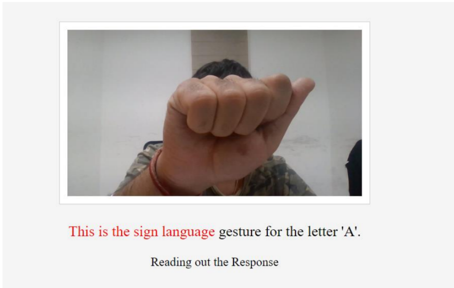

Figure 6 – UI for Sign Language Detection

System Flow:

The image processing system begins with the user uploading or accessing an image via the client-side frontend interface. The image is sent to the server-side Image Processing API, which extracts text, identifies visual elements, and detects object locations. The OCR Service extracts any embedded text, while the Scene Description Model converts this text and scene data into verbal descriptions. Concurrently, Bounding Box Detection identifies and describes object locations. Finally, the Text-to-Speech Service converts these descriptions into speech, which is delivered back to the user through the client-side interface, completing the interaction.

3.3 Voice to Action Commands

Integration with API Endpoints: The proposed system's voice command recognition feature integrates seamlessly with web browsers and other tools used within the learning

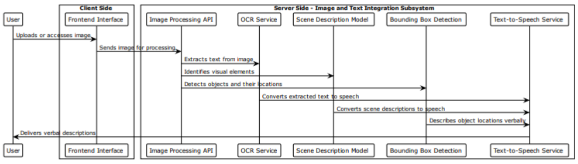

Figure 7 – Activity Diagram for Image and Text Processing

platform. This allows users to perform various actions, such as opening materials, navigating through lessons, or closing applications, using simple voice commands. This deep integration empowers users to control their learning environment independently, enhancing both accessibility and ease of use.

Architecture and Workflow: The voice-to-action feature captures voice commands and processes them to determine the intended action. The system then executes the command, such as navigating to a lesson, turning a page, or opening/closing materials. This streamlined workflow ensures that users can interact with the platform intuitively and efficiently.

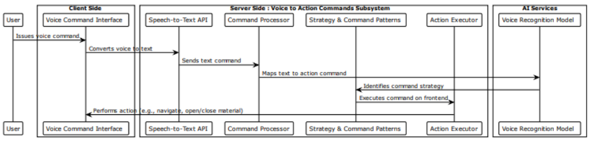

Figure 8 – Activity Diagram for Voice and Action Processing

3.4 AI Voice-Based Tutor - Reinforcement Learning with Human Feedback (RLHF)

Architecture of Implementation: For an AI Voice-Based Tutor, specifically aimed at providing an inclusive learning experience for visually impaired individuals, UnSight employs a dual-model architecture consisting of a generator model (GPT-4) and a validator model (Llama). The user's spoken questions are first transcribed into text, which is then processed by the GPT-4 model to generate a response. The Llama model evaluates this response for factual accuracy and consistency, ensuring reliable and accurate information before it is converted back into speech and delivered to the user.

Personalized Learning: RLHF models tailor the learning pace and content based on individual user interactions and feedback. By observing user responses, engagement levels, and types of questions asked, the model adjusts the difficulty level, provides additional explanations, and recommends relevant learning materials. This personalization ensures that each user receives a customized learning experience catering to their specific needs and progress.

3.5 Attention Theory Considerations

To ensure that studying remains immersive and engaging, UnSight integrates adaptive pacing mechanisms designed to mitigate attentional fatigue and manage cognitive load. After periods of intensive learning—such as 15-20 minutes

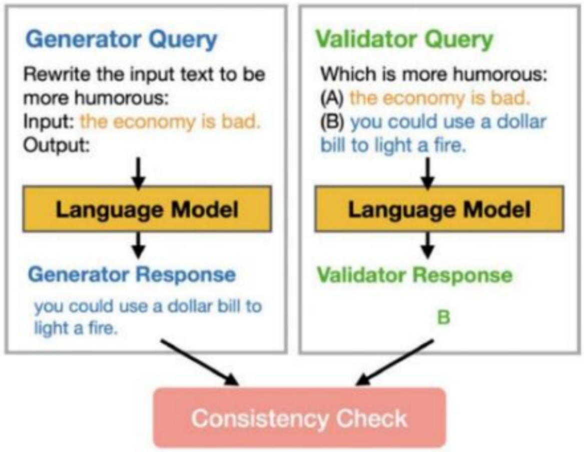

Figure 9 – Validator Generator Architecture

of focused auditory instruction—the system introduces short, light-hearted breaks. These breaks, which might include a brief, contextually relevant joke or a calming auditory interlude, aim to refresh the user's mind and prevent cognitive overload.

This approach is complemented by breaking complex information into smaller, manageable chunks and using scaffolded learning techniques. By presenting information step-by-step and reinforcing key concepts through repetition and varied examples, the system supports sustained attention and reduces the risk of attentional drift. Through these thoughtfully designed elements, UnSight addresses the unique attentional needs of visually impaired learners, fostering engagement and facilitating meaningful learning outcomes.

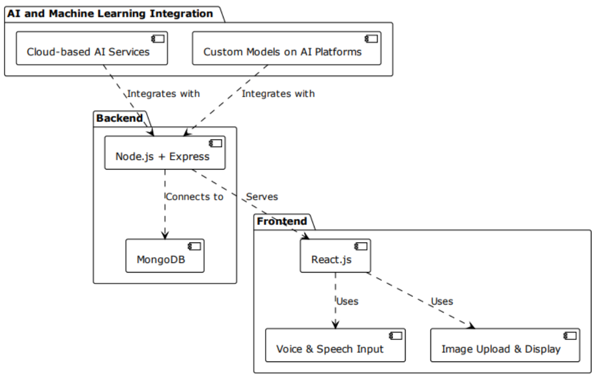

Figure 10 – Novel Architecture of Unsight

4. IMPLEMENTATION PLAN

The following section details the technical architecture, AI integration, and testing methodologies employed in the development of the proposed system, as well as its compliance with educational standards, data security, scalability, and feedback mechanisms.

4.1 Technical Architecture

Platform Development: The backend infrastructure of the proposed system is built using Flask, a lightweight Python web framework chosen for its flexibility and suitability for creating scalable microservices-based architectures. This design promotes modularity and ensures the system's robustness in handling extensive data and user interactions.
The frontend is developed primarily with Python's Flask, HTML and CSS. This enables handling server-side rendering and managing API interactions, ensuring seamless integration between the frontend and backend components. The platform employs MongoDB as the database management system, interfaced with Python through the PyMongo library.

4.2 AI Integration

Speech and Text Processing: The system implements speech recognition using Python's SpeechRecognition library, paired with gTTS (Google Text-to-Speech) for converting text back into speech. Advanced Natural Language Processing (NLP) tasks, including emotion recognition, are handled using pre-trained models from the Hugging Face Transformers library. These tools enable the system to interpret user input accurately and respond in a contextually appropriate manner.
Image and Text Integration: Image recognition is performed using deep learning models such as or YOLO, facilitated by TensorFlow or PyTorch in Python. Tesseract OCR is integrated to extract text from images, which is then processed into accessible formats for visually impaired users.
Voice to Action Commands: Voice command functionalities are developed using Python's SpeechRecognition library, integrated with backend APIs built in Flask. This setup ensures secure execution of commands, with stringent data privacy measures in place.

4.3 Testing and Quality Assurance

Unit Testing: Comprehensive unit testing is conducted using Python's unit test and pytest frameworks, ensuring that each component functions correctly.
Integration Testing: Integration testing ensures the seamless interaction between various modules, utilizing pytest and Selenium to validate end-to-end workflows.
User Testing: The system undergoes extensive user testing, which includes scenarios where participants are blindfolded to simulate the experience of visually impaired users. While these tests offer valuable insights, extensive testing with actual blind users is acknowledged as a necessary future step, beyond the scope of this current paper.

4.4 Data Security and Privacy

The proposed system incorporates a comprehensive approach to data security, privacy, and ethical considerations, ensuring that the platform is both safe and trustworthy for users.
GDPR Compliance: The system complies with GDPR regulations, utilizing Python libraries like GDPRt to enforce data protection protocols. This ensures that all user data is handled in accordance with strict privacy laws, providing users with control over their personal information and guaranteeing their rights are respected.
Encryption: Data security is reinforced through end-to-end encryption, implemented via Python libraries such as `cryptography` and `PyCryptodome`. This encryption safeguards user data during both storage and transmission, ensuring that sensitive information remains protected from unauthorized access and potential breaches.
Website Security Measures: The platform uses multi-factor authentication (MFA) to enhance security during user login processes, including the use of facial recognition technology and secure password protocols, ensuring that only authorized users can access their accounts.
Protective Measures for AI Technologies: The system integrates a "Dehallucinator" module to filter and remove hallucinated or inaccurate content generated by AI models, ensuring the accuracy and reliability of the information provided to users. Additionally, data anonymization techniques are employed, where personal data is anonymized before being processed by AI models, further protecting user privacy.
Ethical and Privacy Considerations: The platform is designed to prevent data leakage by implementing stringent access controls and secure data storage practices, with all user data stored in encrypted formats and access restricted to authorized personnel. The platform is committed to respecting and protecting user rights by providing clear information on data collection, usage, and storage, and offering users control over their data. Ethical AI practices are central to the platform's design, ensuring that AI technologies do not cause unintended side effects or harm to users. Continuous monitoring and updates to the AI models are conducted to mitigate risks and ensure safe usage.

4.5 Scalability and Deployment

Infrastructure Scaling: The system is designed for scalability, with **auto-scaling** configurations managed within **AWS**. Python-based scripts facilitate the monitoring and adjustment of these resources to handle fluctuating user loads.
Load Balancing: Load balancing techniques are employed to distribute traffic evenly across multiple servers, using Python tools to maintain high availability and performance, particularly during peak usage.
Geographic Expansion: The platform leverages AWS's regional data centers to minimize latency, with Python-based orchestration tools managing the deployment across these centers.
Continuous Integration and Continuous Deployment (CI/CD): A **CI/CD pipeline** is established using tools such as **Jenkins** or **GitLab CI**, with Python scripts automating the testing, integration, and deployment processes. Version control is managed through **Git**, with Python's **GitPython** enhancing the efficiency of these operations.

4.6 Feedback and Iterative Improvement

User Feedback Mechanisms: The platform integrates **feedback mechanisms** such as surveys and polls, which are automated and analyzed using Python libraries like the **SurveyMonkey API**. Feedback forms within the platform, managed by Flask, allow users to provide real-time input, guiding continuous improvements.
Iterative Development: Adopting an **Agile methodology**, the platform's development is guided by user feedback, with Python tools facilitating the iterative enhancement of features. **Regular updates** are deployed via Python-based CI/CD pipelines, ensuring that the platform remains responsive to user needs and technological advancements.

5. USABILITY STUDY

This study delves into the platform's usability, exploring the user journey from initial website visit to engaging with its AI-powered features.

5.1 Homepage and user registration:

Upon accessing the UnSight website, users are greeted by a homepage that provides a comprehensive overview of the project's mission, key features, and impact. A prominent "Login" button, located in the top-left corner, prompts users to either log in to their existing accounts or register for new ones. For registration, users are required to provide basic personal information, including email, phone number, name, and age. The platform also facilitates a user-friendly process for users to describe their vision impairment with the help of a sighted assistant. Following verification of their mobile number and email, new users can securely log in to the system. UnSight leverages facial recognition for enhanced security, recording the user's face for future login attempts, thereby streamlining the access process.

5.2 Command Selection and Interaction:

Upon successful login, the UnSight platform presents users with a choice of two main interaction modes:

1. **Conversational Learning:** This mode leverages AI-powered speech recognition and natural language processing to provide an interactive and engaging learning experience. Users can learn various subjects, interact with educational materials, including documents and images, and even practice writing alphabets and numbers. The system incorporates sign language learning as well, allowing users to expand their communication skills.

2. **System Actions**: Users can utilize voice commands to perform specific actions within the system, such as calling contacts, controlling device settings, or accessing specific functionalities.

5.3 Emotional Awareness and Adaptive Learning:

Throughout the learning process, UnSight prioritizes emotional awareness and adapts its response to the user's emotional state. The AI models detect and analyze the user's tone and sentiment, providing feedback and support tailored to their needs. The platform also incorporates adaptive pacing mechanisms, including strategically timed breaks, to help prevent fatigue and cognitive overload, ensuring sustained engagement and optimal learning outcomes.

6. SUSTAINABILITY AND SCALABILITY CONCERNS : ADVANTAGES OF OUR IMPLEMENTATION OVER EXISTING SOLUTIONS

Our Implementation can scale to incorporate more features and we can create a solution as AI-powered Accesssibility plugins which are secure and do not expose personal data to any parties. We use Open Source and proprietary Privately owned models for inference.

6.1 Sustainability of Our AI Solutions

Ethical AI Use: We commit to ethical AI practices by ensuring our data handling and machine learning models are transparent and fair. Our models are continuously monitored and updated to prevent biases and ensure equitable use across diverse populations.

Scalable Learning Models: Our AI solutions employ models that learn and adapt over time without requiring extensive retraining. This reduces the carbon footprint associated with training large AI models and ensures our solutions improve while being environmentally conscious.

6.2 Scalability of Our Technological Framework

Modular Design: Our system architecture features modular components that can be independently updated, scaled, or replaced without disrupting the entire ecosystem. This allows us to integrate newer technologies and improvements with minimal overhead and ensures that our solutions remain at the cutting edge of accessibility technology.

Cross-Platform Compatibility: Designed to operate across various platforms and devices, our solutions ensure that users can access our services regardless of their preferred technology. This broad compatibility supports widespread adoption and enhances the user experience across different hardware and software environments.

6.3 Advantages Over Existing Solutions

AI-Powered Plugins: Unlike traditional assistive technologies, our AI-powered plugins are designed to function seamlessly as part of existing digital ecosystems. This integration capability allows users to enjoy enhanced accessibility without needing separate tools or applications, promoting a more inclusive digital environment.

Data Privacy and Security: We use a combination of open-source and proprietary models that ensure data privacy and security. Our systems are designed to process data locally where possible, reducing the need to transmit sensitive information and minimizing the risk of data breaches.

Customizable and Extendable: Our solutions are not only robust but also highly customizable. This flexibility allows organizations and users to tailor the tools to their specific needs, ensuring that the technology adaptates to user requirements rather than the other way around.

7. RESULTS

The development and implementation of UnSight have demonstrated its potential to transform the learning experience for visually impaired students. Notable achievements include significantly improved accessibility through the integration of speech recognition, emotion detection, image description, and voice-to-action commands, which enable students to engage with educational content effectively and independently. The system's adaptive learning capabilities, driven by AI and reinforcement learning with human feedback (RLHF), offer personalized learning experiences tailored to individual needs, learning styles, and progress. Additionally, UnSight's emotion detection and emotionally intelligent responses contribute to a more engaging and empathetic learning environment, thereby enhancing student motivation and retention. By equipping visually impaired students with accessible tools and features, UnSight fosters independence, confidence, and a sense of inclusion. The overwhelmingly positive feedback from user testing and pilot studies further underscores the platform's intuitive design, ease of use, and its significant impact on learning outcomes.

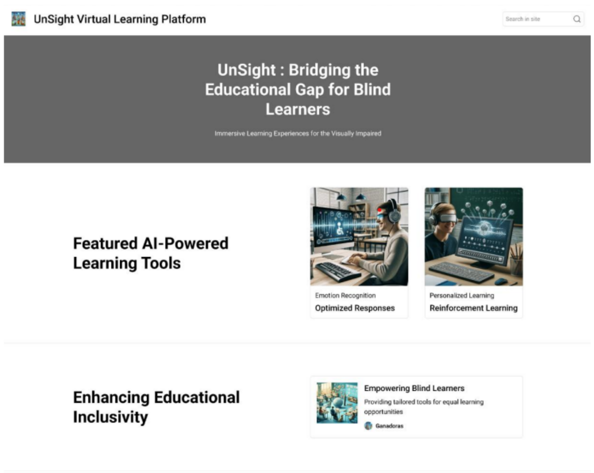

Figure 11 – Web UI of UnSight

8. CONCLUSIONS AND FUTURE WORK

UnSight represents a pioneering effort in the realm of virtual education, demonstrating the transformative potential of AI-driven technologies in promoting inclusive and equitable quality education. This innovative platform aligns with the United Nations' Sustainable Development Goal 4 (SDG 4) by addressing the unique needs of visually impaired students and contributing to the creation of inclusive and effective learning environments. Through the seamless integration of speech recognition, emotion detection, image description, and voice-to-action commands, the proposed system offers an immersive, accessible, and personalized learning experience. Although the platform does not involve the use of custom datasets or the training of proprietary algorithms, it leverages state-of-the-art and best-fit existing methodologies by ensembling them to deliver a comprehensive and effective solution. The platform's adaptive learning capabilities, powered by AI and reinforcement learning, ensure that educational content is tailored to individual needs and learning styles, thereby enhancing engagement, comprehension, and retention. The positive impact of the proposed system extends beyond academic achievement; it empowers visually impaired students by fostering independence, confidence, and a sense of inclusion. By providing accessible tools and features, the platform enables students to take control of their learning journey, unlocking their full potential and contributing to a more inclusive and equitable society.

While the proposed system has achieved significant milestones in revolutionising virtual education for visually impaired students, several avenues for future exploration and enhancement have been identified. Firstly, the integration of haptic and olfactory technologies could further enhance the immersive and multisensory learning experience, catering to a wider range of sensory modalities and learning preferences. Secondly, exploring the implementation of collaborative learning features, such as virtual classrooms and group activities, could foster social interaction, peer-to-peer learning, and a sense of community among visually impaired students. Thirdly, the incorporation of gamification techniques and interactive learning modules has the potential to increase motivation, engagement, and retention, thereby making the learning experience more enjoyable and rewarding. Also, the inclusion of teachers and the development of structured frameworks for their interactions with students within the platform could further enhance the educational experience and ensure effective guidance and support. Additionally, ensuring interoperability with existing educational systems, including compatibility with Learning Management Systems (LMS), digital content standards (e.g., SCORM, xAPI), and accessibility standards (e.g., WCAG, ARIA), is a key area for future development. The framework can also use cloud infrastructure for better functioning. Leveraging cloud computing, our solutions can scale dynamically based on user demand. This not only ensures seamless performance during peak usage but also helps in managing operational costs efficiently, allowing us to invest more in innovation and less in static infrastructure. Lastly, integrating UnSight with microprocessor-based modules and Internet of Things (IoT) devices could enable multimodal and multi-visual learning experiences, further enhancing the platform's capabilities and expanding its potential applications in various educational contexts.

REFERENCES

[1] IBM. (n.d.). A commitment to accessibility. Retrieved from https://www.ibm.com/accessibility

[2] Microsoft. (n.d.). Doubling down on accessibility: Microsoft's next steps to expand accessibility in technology, the workforce and workplace. Retrieved from https://blogs.microsoft.com

[3] Microsoft Research. (n.d.). Where's my stuff? Developing AI with help from people who are blind or low vision to meet their needs. Retrieved from https://www.microsoft.com

[4] Be My Eyes. (n.d.). Retrieved from https://www.bemyeyes.com

[5] Aira. (n.d.). Retrieved from https://www.aira.io

[6] Tilting the Lens. (2023). Digital Accessibility for Blind and Low-Vision Consumers. Retrieved from link to paper

[7] Explaining CLIP's Performance Disparities on Data from Blind/Low Vision Users. (2023). Retrieved from https://arxiv.org/abs/2311.17315

[8] ImageAssist: Tools for Enhancing Touchscreen-Based Image Exploration Systems for Blind and Low Vision Users. (2023). Retrieved from https://arxiv.org/abs/2302.09124

[9] Artificial Intelligence for Visually Impaired. (2023). Retrieved from https://doi.org/10.1016/j.displa.2023.102391

[10] Towards Assisting Visually Impaired Individuals: A Review on Current Solutions. (2023). Retrieved from http://dx.doi.org/10.1016/j.biosx.2022.100265

[11] Recent Trends in Computer Vision-Driven Scene Understanding for Visually Impaired. (2023). Retrieved from https://doi.org/10.1007/s10209-022-00868-w

[12] Graves, A., et al. "End-to-End Speech Recognition: A Survey." arXiv preprint arXiv:2303.03329, 2023.

[13] Reddy, S. Y., et al. "A Hybrid Approach to Converting Visuals into Speech for the Visually Impaired." International Journal for Innovative Engineering & Management Research, 2023.

[14] Bose, P. "An Intelligent Voice Assistant Engineered to Assist the Visually Impaired." Springer, 2023.

[15] Multiple Authors. "Real-Time Object Detection and Recognition for the Visually Impaired." Springer, 2023.

[16] Multiple Authors. "Intelligent Glasses for Visually Impaired People." IEEE Xplore, 2023.

[17] Mahendran, J. K., et al. "Computer Vision-Based Assistance System for the Visually Impaired Using Mobile Edge AI." CVPR 2021 Workshop on Mobile AI, 2021.

[18] Ratna, Ch., et al. "Image to Audio for Visually Impaired." International Journal for Innovative Engineering Management Research, 2023.