

Voice Driven Web Accessibility Tool for Visually Impaired Individuals

Sanika Vinod Modak, Sanskruti Dhairyashil Dhumal, Neha Jalinder Autade,
Disha Ramratan Nimje

Department of Information Technology, Sinhgad Institute of Technology and Science (SPPU), Pune, India

Abstract - The Web has become an indispensable medium for communication, education, and information access; however, many users with visual, cognitive, and learning disabilities continue to encounter challenges while interacting with increasingly complex web content. Existing assistive technologies and screen readers provide basic accessibility support but often suffer from limitations such as linear navigation, limited contextual understanding, and inadequate mechanisms for reducing cognitive load. To address these challenges, this paper presents **VisionaryAI**, an AI-powered voice-driven browser extension designed to provide intelligent and inclusive web interaction. The proposed framework integrates speech recognition, conversational artificial intelligence, and adaptive text-to-speech technologies to enable natural voice-based navigation, webpage reading, AI-generated content summarization, and image description capabilities. In addition, VisionaryAI incorporates cognitive accessibility features including focus mode, decluttering, reading rulers, dyslexic-friendly typography, and automated WCAG 2.2 accessibility auditing. Unlike conventional screen readers that primarily deliver sequential content access, the proposed system offers a multimodal and interactive browsing experience. VisionaryAI demonstrates how the integration of artificial intelligence and human-centered accessibility principles can contribute to a more inclusive, efficient, and user-friendly web environment.

Key Words: Web accessibility, voice-driven interaction, conversational AI, automatic speech recognition, text-to-speech synthesis, web augmentation, intelligent content assistance, cognitive accessibility, inclusive design, assistive technology, browser extension, WCAG compliance, human-computer interaction, image description, AI-powered navigation.

1. INTRODUCTION

The World Wide Web has become an indispensable platform for communication, education, academic research, and digital information exchange. As dynamic web services continue to expand, seamless access to web content has become essential for professional and

academic advancement. However, despite considerable developments in assistive technologies and baseline accessibility standards, many users continue to experience severe barriers when interacting with complex, information-rich websites. Individuals with visual impairments, dyslexia, cognitive disabilities, and motor limitations frequently face compounding difficulties in navigating, understanding, and consuming dense digital content effectively. Ensuring equal, high-efficiency access to digital information therefore remains a significant challenge in the pursuit of an inclusive Web.

Recent advances in conversational artificial intelligence, automatic speech recognition (ASR), and text-to-speech (TTS) technologies have transformed the landscape of human-computer interaction. Voice user interfaces and large language models (LLMs) have increasingly become part of everyday activities, enabling natural, hands-free interaction with computing devices. These technologies hold considerable potential for reducing interaction barriers for users with permanent as well as situational disabilities. Several studies have explored the application of conversational agents for information retrieval and intelligent assistance, highlighting the possibility of making digital systems more adaptive and user centric.

However, while existing screen readers and accessibility tools provide fundamental support for web access, their interaction paradigm remains predominantly linear and sequential. Users are forced to listen to large amounts of non-essential content before reaching relevant information, leading to severe cognitive fatigue and reduced browsing efficiency. Furthermore, traditional assistive technologies lack contextual understanding, intelligent summarization capabilities, and targeted support for highly complex web tasks—such as navigating academic PDFs, parsing data-dense research databases, or filling out sophisticated web forms. As modern web platforms become increasingly dynamic, these technical limitations drastically hinder productivity, particularly for researchers and writers with diverse accessibility needs.

To address these challenges, this paper proposes VisionaryAI, a voice-driven browser extension designed to provide intelligent, inclusive, and high-efficiency web interaction. Optimized for general browsing as well as academic research workflows, the proposed framework combines continuous automatic speech recognition and adaptive text-to-speech synthesis with AI-assisted page summarization, intelligent image description, and voice-based structural navigation. To directly mitigate information overload during complex reading tasks, the system incorporates cognitive accessibility mechanisms, including focus modes, clutter elimination, reading rulers, and specialized dyslexic-friendly typography. Additionally, VisionaryAI features automated accessibility auditing based on WCAG 2.2 guidelines and an inline privacy protection framework via an automated PII (Personally Identifiable Information) scrubber, ensuring secure and private speech interactions.

Unlike conventional screen readers that mainly provide sequential reading capabilities, the proposed system offers a multimodal and interactive browsing paradigm where users can navigate and comprehend web content through natural language commands. By integrating conversational AI with live web augmentation techniques, VisionaryAI aims to bridge the gap between traditional assistive tools and modern AI-driven interaction models. This development is motivated by the broader goal of promoting human-centered, inclusive technologies. Accessibility solutions should not merely focus on providing passive access to raw text, but rather on improving comprehension, reducing cognitive effort, and enabling independent interaction with sophisticated digital environments.

The primary contributions of this work are summarized as follows:

- **Intelligent Multimodal Framework:** The development of a voice-driven browser extension architecture that pairs conversational AI with dynamic web augmentation for platform-independent deployment.
- **Hands-Free Navigation & Parsing:** The integration of continuous speech recognition and adaptive text-to-speech mechanisms tailored for rapid navigation of complex layouts, headers, and content structures.
- **Contextual AI Processing:** The implementation of OpenAI-driven page summarization and vision-based

image descriptions to drastically improve information accessibility and synthesis for research tasks.

- **Cognitive Support Suites:** The incorporation of real-time DOM-manipulation features (focus mode, decluttering, reading rulers, and typography adjustments) engineered to reduce cognitive load and enhance readability.
- **Privacy & Compliance Engineering:** The development of an automated compliance auditor aligned with WCAG 2.2 guidelines, paired with an architectural PII scrubber to guarantee data privacy during auditory synthesis.

2. LITERATURE REVIEW

The increasing complexity of modern websites and the growing demand for inclusive digital experiences have motivated extensive research in web accessibility, assistive technologies, conversational artificial intelligence (AI), and intelligent content understanding. Recent advances in Automatic Speech Recognition (ASR), Text-to-Speech (TTS), large language models (LLMs), and web augmentation techniques have created opportunities for developing intelligent systems that improve accessibility and facilitate more natural human-computer interaction.

Several researchers have investigated the use of conversational agents and voice-based interfaces for improving web accessibility. Akhter et al. [1] proposed a conversational web browsing framework that enables users to navigate websites through natural language interactions, demonstrating the effectiveness of voice-based dialogue for users with visual impairments. Buzzi et al. [2] explored the application of conversational agents for web search and information retrieval, highlighting their potential to provide more intuitive and accessible interaction mechanisms. Caporusso et al. [3] developed methods for automatically generating conversational interfaces from web page content, enabling users to interact with websites using natural language commands. Sano et al. [4] investigated customizable conversational agents that allow end users to extract and access relevant web content through voice-driven interactions.

Researchers have also focused on accessibility-oriented web augmentation techniques. Bigham et al. [5] demonstrated the effectiveness of web augmentation methods for enhancing accessibility and improving user

interaction with existing web pages. Harper and Chen [6] discussed the importance of universal accessibility and emphasized the role of adaptive interfaces in supporting users with diverse abilities. Vigo et al. [7] analyzed the limitations of traditional screen readers and highlighted the need for more intelligent and context-aware browsing mechanisms. Their findings indicate that sequential content reading often increases cognitive load and reduces browsing efficiency.

Recent developments in artificial intelligence have further expanded the capabilities of assistive technologies. OpenAI [8] introduced large language models capable of performing advanced language understanding and content summarization tasks, providing new opportunities for intelligent accessibility systems. Li et al. [9] investigated AI-assisted summarization techniques for improving information comprehension, demonstrating their usefulness in reducing information overload. Almazrouei et al. [10] explored multimodal AI approaches for image understanding and natural language interaction, enabling systems to generate descriptive information for visual content and thereby improving accessibility for visually impaired users.

Several studies have examined cognitive accessibility and adaptive user interfaces. Wobbrock et al. [11] emphasized the importance of human-centered design principles in assistive technologies and advocated the development of systems that accommodate users with varying cognitive and physical capabilities. Cooper et al. [12] investigated adaptive accessibility mechanisms, including focus enhancement and content simplification techniques, to improve readability and reduce cognitive burden. Other researchers have highlighted the effectiveness of specialized typography, reading aids, and distraction reduction techniques in supporting users with dyslexia and cognitive impairments [13].

Although existing approaches have demonstrated promising results in conversational AI, web augmentation, content summarization, and adaptive accessibility, most studies focus on individual components independently. Existing assistive technologies generally provide either voice interaction, content reading, or accessibility enhancement, without integrating these capabilities into a unified framework. Furthermore, many conventional systems continue to rely on static interaction paradigms and lack intelligent content understanding and cognitive support mechanisms.

To address these limitations, VisionaryAI integrates conversational AI, automatic speech recognition, adaptive text-to-speech synthesis, AI-powered summarization, image description, voice-based navigation, cognitive accessibility features, and automated WCAG 2.2 accessibility auditing within a single browser extension. By combining intelligent content understanding with accessibility-oriented web augmentation techniques, the proposed framework aims to provide a comprehensive, scalable, and human-centered solution for inclusive web interaction. Furthermore, the system incorporates personalized interaction settings, multimodal feedback mechanisms, and cognitive support tools to enhance usability and reduce information overload. Through seamless integration of artificial intelligence and accessibility principles, VisionaryAI seeks to improve browsing efficiency and promote a more equitable digital experience for users with diverse accessibility needs

3. PROBLEM DEFINITION

The central problem that VisionaryAI addresses is that existing web accessibility solutions are often too rigid, too linear, and too limited in contextual understanding to provide an efficient and inclusive browsing experience. Conventional screen readers and voice assistants primarily deliver content sequentially, forcing users to traverse large portions of web pages before reaching relevant information. As modern websites become increasingly dynamic and information-dense, this interaction paradigm introduces significant cognitive burden and reduces browsing efficiency, particularly for individuals with visual impairments, dyslexia, cognitive disabilities, and motor limitations.

No single accessibility mechanism is sufficient to address the diverse challenges encountered during web interaction. Voice assistance alone cannot provide contextual understanding, while traditional screen readers lack intelligent summarization and adaptive navigation capabilities. Similarly, cognitive accessibility requires multiple forms of support, including distraction reduction, focus enhancement, readability improvement, and structured content presentation. Existing assistive technologies typically address these requirements independently, resulting in fragmented solutions that fail to provide a unified and natural interaction experience.

Furthermore, modern web content increasingly contains images, multimedia elements, and complex layouts that are difficult to interpret through conventional accessibility

tools. Users often experience information overload caused by advertisements, sidebars, and non-essential content, while the absence of contextual descriptions and content prioritization limits comprehension. In addition, ensuring compliance with accessibility standards remains challenging, and users lack mechanisms for identifying and understanding accessibility issues present within web pages.

To address these limitations, there is a need for an intelligent and multimodal accessibility framework capable of combining conversational AI, voice interaction, content understanding, and cognitive accessibility mechanisms within a single platform. Such a system should enable users to navigate websites naturally, comprehend information efficiently, and interact with digital environments in a manner that promotes accessibility, usability, and inclusivity.

Objectives

To systematically address the problems described above, the following objectives were defined for VisionaryAI:

- Develop a voice-driven browser extension that enables hands-free interaction with websites through natural language commands.
- Implement automatic speech recognition and adaptive text-to-speech mechanisms to provide seamless voice-based navigation and content consumption.
- Incorporate AI-powered page summarization to convert lengthy and complex web content into concise and comprehensible information.
- Implement image description capabilities to improve accessibility for visually impaired users and facilitate understanding of visual content.
- Provide intelligent navigation mechanisms, including heading-based traversal, section navigation, page overview, and location awareness.
- Integrate cognitive accessibility features such as focus mode, declutter mode, reading rulers, and dyslexic-friendly typography to reduce cognitive load and improve readability.
- Implement automated accessibility auditing based on WCAG 2.2 guidelines to identify and highlight accessibility issues within web pages.

- Develop a multimodal interaction framework that combines conversational AI with web augmentation techniques to provide a more natural and inclusive browsing experience.
- Support customizable speech parameters and multiple interaction personas to accommodate diverse user preferences and accessibility requirements.
- Develop a lightweight browser extension architecture capable of operating across different websites while ensuring low latency and platform independence.

4. PROPOSED METHODOLOGY

The proposed system, VisionaryAI, employs a modular AI-driven accessibility framework for intelligent and voice-based web interaction. The methodology is divided into seven major phases: Voice Input Acquisition, Speech Recognition and Intent Understanding, Web Content Extraction, AI-Assisted Content Processing, Cognitive Accessibility Enhancement, Accessibility Auditing, and Response Generation. Each phase contributes to building a comprehensive and inclusive browsing assistant capable of supporting users with visual, cognitive, and motor impairments.

A. Voice Input Acquisition

The interaction begins with the user providing voice commands through the browser extension interface. A wake-word mechanism activates the system and continuously listens for commands related to content reading, navigation, summarization, image description, and accessibility customization. Voice commands constitute the primary input modality, enabling hands-free interaction with web content.

B. Speech Recognition and Intent Understanding

The captured audio is processed using the Web Speech API to convert speech into text. Subsequently, an intent processing engine analyses the textual commands and identifies the user's desired operation. Natural language interpretation enables the system to distinguish between various actions such as "read page", "next heading", "summarize page", "describe image", and "focus mode". This stage ensures accurate mapping between user requests and system functionalities.

C. Web Content Extraction and Navigation

After identifying the user's intent, the system performs Document Object Model (DOM) analysis to retrieve

relevant webpage information. Structural elements such as headings, paragraphs, lists, and images are extracted to facilitate intelligent navigation. Semantic webpage structures are utilized to enable section-based browsing and page overview generation. This approach provides more efficient access to information compared with traditional sequential screen reading.

D. AI-Assisted Content Processing

To improve information comprehension, conversational AI techniques are incorporated into the framework. Page summarization mechanisms transform lengthy content into concise summaries, while image analysis modules generate descriptive information for visual elements. These capabilities provide contextual understanding of webpage content and enhance accessibility for visually impaired users. AI-assisted processing enables users to consume information more efficiently and reduces information overload.

E. Cognitive Accessibility Enhancement

The framework integrates several web augmentation techniques designed to improve readability and reduce cognitive burden. Declutter mode removes distracting webpage components, whereas focus mode highlights the primary content region. Reading rulers assist users in maintaining reading continuity, and dyslexic-friendly typography enhances text readability for individuals with learning disabilities. These mechanisms collectively provide adaptive support for users with diverse accessibility requirements.

F. Accessibility Auditing

An automated accessibility auditing module evaluates webpages according to WCAG 2.2 guidelines. The module analyses webpage elements and identifies accessibility violations that may hinder user interaction. The generated reports provide recommendations for improving website accessibility and promote adherence to inclusive design principles.

G. Response Generation and Speech Synthesis

The processed information is delivered to the user through adaptive Text-to-Speech synthesis. Speech parameters including rate, pitch, and interaction persona can be customized according to individual preferences. Visual overlays and popup notifications complement auditory feedback, creating a multimodal interaction environment. The coordination between speech recognition and speech synthesis ensures seamless communication and prevents conflicts during operation.

H. System Architecture

The proposed VisionaryAI system adopts a layered and modular architecture designed to provide intelligent and accessible web interaction. The architecture consists of six interconnected components: the User Interface Layer, Speech Processing Layer, Content Extraction Module, AI Processing Module, Cognitive Accessibility Module, and Output Generation Module.

The process begins when the user provides a voice command through the browser extension interface. The Speech Processing Layer converts audio into text and forwards it to the Intent Engine, which determines the required action. The Content Extraction Module analyzes the webpage structure using DOM traversal and retrieves relevant information. Simultaneously, the AI Processing Module performs page summarization and image description tasks. The Cognitive Accessibility Module applies focus mode, declutter mode, reading rulers, and dyslexic-friendly typography to improve readability.

The Accessibility Auditing Module evaluates webpage compliance with WCAG 2.2 standards and identifies potential accessibility issues. Finally, the Text-to-Speech engine and visual overlay components communicate the processed information to the user. This layered design ensures low-latency interaction while combining conversational AI and accessibility-oriented web augmentation techniques to provide an inclusive browsing experience.

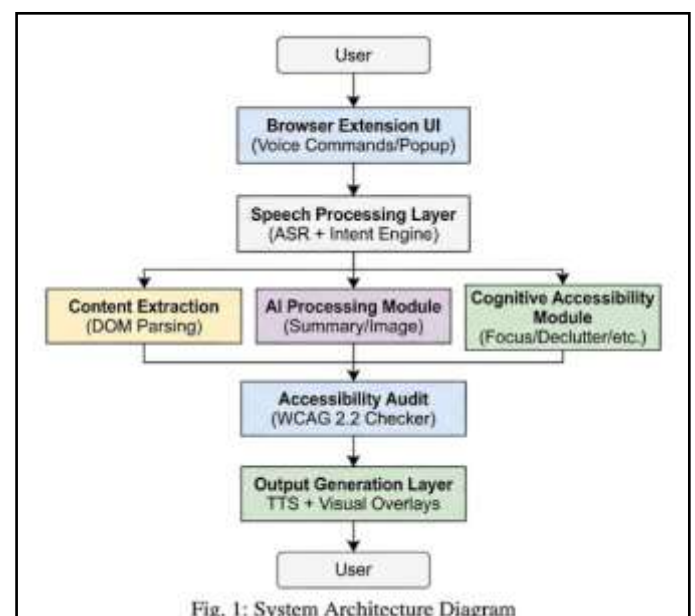


Fig. 1: System Architecture Diagram

I. Use Case Diagram

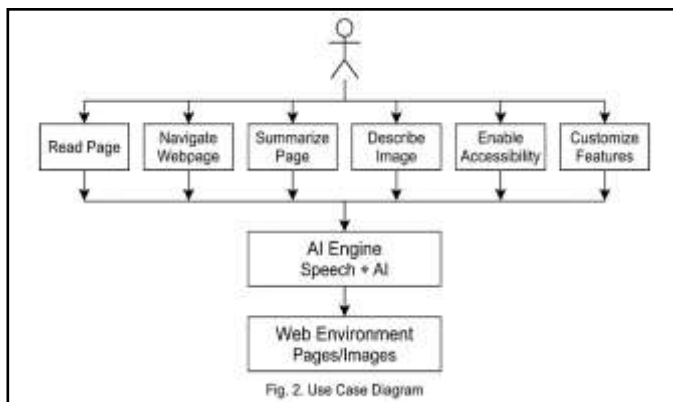
The use case diagram for VisionaryAI illustrates the interactions between three primary actors: the User, the AI Engine, and the Web Environment.

The User represents individuals interacting with the browser extension. Users initiate interaction by activating the voice assistant and issuing commands to read webpages, navigate sections, summarize content, describe images, and enable accessibility features. Users can also customize speech parameters and interaction personas according to their preferences.

The AI Engine represents the intelligent backend responsible for speech recognition, intent processing, summarization, image description, and accessibility auditing. It processes user commands and generates adaptive responses in real time.

The Web Environment represents webpages and multimedia content that provide information to the system. Through DOM traversal and content extraction, the system retrieves webpage elements and presents them to users in an accessible format.

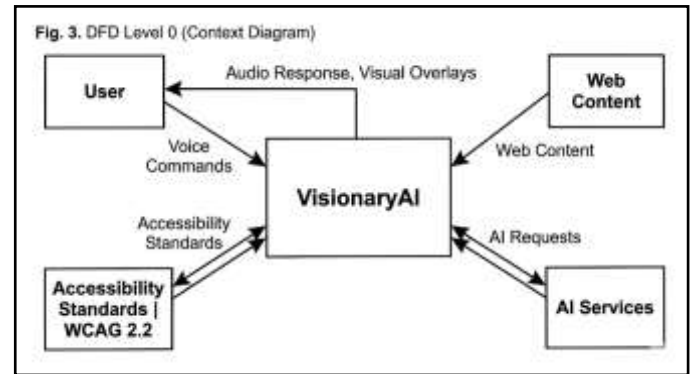
Together, these actors define the complete functional scope of VisionaryAI and demonstrate how conversational AI and accessibility technologies collaborate to deliver intelligent web interaction.



J. Data Flow Diagrams

I. DFD Level 0

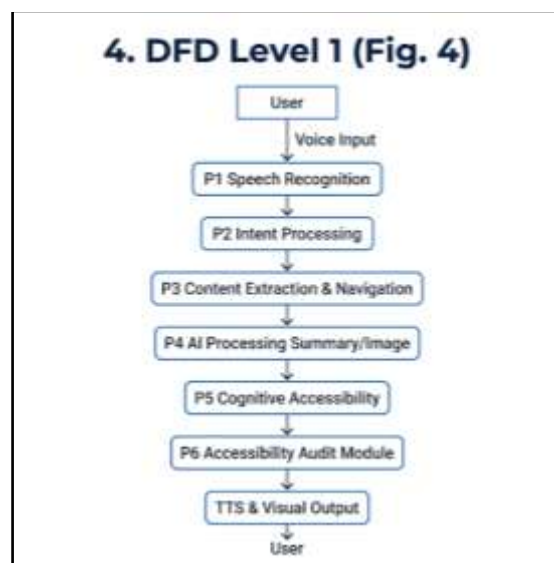
The DFD Level 0 Context Diagram provides a high-level overview of the VisionaryAI system and its interactions with external entities. The User submits voice commands and accessibility preferences to the central VisionaryAI framework. Web Pages supply webpage content and multimedia information required for processing, while AI Services provide support for content summarization and image description. The system processes these inputs and returns audio responses, visual overlays, and accessibility recommendations to the user. This diagram establishes the system boundaries and illustrates the major information flows between external entities and the core accessibility framework.



II. DFD Level 1

The DFD Level 1 diagram decomposes VisionaryAI into six internal processes that together execute the complete accessibility workflow. The process begins when the User provides voice input to P1: Speech Recognition Module, which converts speech into text. The recognized commands are forwarded to P2: Intent Processing Engine, which identifies the requested action. Subsequently, P3: Content Extraction and Navigation Module performs DOM analysis and retrieves webpage information.

When required, P4: AI Processing Module generates page summaries and image descriptions using conversational AI techniques. The processed content is then forwarded to P5: Cognitive Accessibility Module, which applies focus mode, declutter mode, reading rulers, and dyslexic-friendly typography to enhance readability. In parallel, P6: Accessibility Audit Module evaluates webpage compliance with WCAG 2.2 standards and generates accessibility recommendations. Finally, the Text-to-Speech engine produces auditory feedback and delivers the resulting output to the user.



5. IMPLEMENTATION

The proposed system, VisionaryAI, is implemented as a browser-based accessibility extension that provides voice-driven interaction and AI-assisted support for web browsing. The extension integrates speech recognition, text-to-speech, webpage content extraction, and AI-powered features to enhance accessibility and user experience.

Technologies Used:

- Speech Recognition: Chrome Web Speech API
- Text-to-Speech: Chrome Speech Synthesis API, Artyom.js
- AI Processing: OpenAI API
- Web Content Extraction: JavaScript DOM API
- Frontend: HTML, CSS, JavaScript
- Platform: Google Chrome Extension (Manifest V3)

The system operates in real time and allows users to navigate webpages, read content aloud, generate summaries, describe images, and interact with web elements using voice commands.



Fig. 5. VisionaryAI-AI Assistant Interface

As shown in Fig. 5, the VisionaryAI assistant is integrated directly into the webpage through an interactive chat panel. The interface provides AI-generated page summaries, answers user queries, and delivers voice-based assistance while browsing. This enables users to access important information quickly without manually reading the entire webpage.



Fig. 6. VisionaryAI Extension Popup Interface

As shown in Fig. 6, the extension popup serves as the primary control panel of VisionaryAI. It displays webpage headings for quick navigation and provides tabs for actions and settings. Users can access voice controls, accessibility features, and customization options through a simple and user-friendly interface.

6. REFERENCES

1. L. Kumar, A. S. Kushwaha, and A. Jain, "An auditory system interface for augmented accessibility: Empowering the visually impaired," *Engineering Proceedings*, vol. 62, no. 1, p. 8, Mar. 2024, presented at the 2nd Computing Congress (CC 2023), Chennai, India, Dec. 28–29, 2023.
2. J. Nino, S. Ochoa, J. Kiss, G. Edwards, E. Morales, J. Hutson, F. Poncet, and W. Wittich, "Assistive technologies for internet navigation: A review of screen reader solutions for the blind and visually impaired," *International Journal of Recent Engineering Science*, vol. 11, no. 6, pp. 260–274, Dec. 2024.
3. T. Calle-Jimenez and B. Munoz, "Deep analysis of the web accessibility and usability of video conferencing platforms for blind people," in *Usability and User Experience*, T. Ahram and C. Falcao, Eds., AHFE~ Int. Conf. (2023), AHFE Open Access, vol. 110, AHFE International, USA, 2023.
4. V. Bercaru and N. Popescu, "A systematic review of accessibility techniques for online platforms: Current trends and challenges," *Applied Sciences*, vol. 14, no. 22, p. 10337, Nov. 2024.
5. A. Aqle, K. Khowaja, and D. Al-Thani, "Preliminary evaluation of interactive search engine

interface for visually impaired users,” arXiv preprint arXiv:1808.09885, 2018.

6. R. Patel, “Screen Reader AI: A conversational web-accessibility assistant for blind and low-vision users,” *International Journal of Engineering Science and Information Technology*, vol. 5, no. 3, pp. 671–676, Jul. 2025.

7. S., S. MACAKOGLU and S. Peker, ”Web accessibility performance~ analysis using web content accessibility guidelines and automated tools: a systematic literature review,” 2022 International Congress on HumanComputer Interaction, Optimization and Robotic Applications (HORA), Ankara, Turkey, 2022.

8. J. Ara, C. Sik-Lanyi, A. Kelemen, and T. Guzsvinecz, “An inclusive’ framework for automated web content accessibility evaluation,” *Universal Access in the Information Society*, vol. 24, pp. 1581–1607, 2025.

9. R. Devi and S. Kumar, “Accessibility evaluation of selected library websites in Delhi-NCR using an automated tool,” *Annals of Library and Information Studies*, vol. 70, no. 4, pp. 215–225, Dec. 2023.

10. F. Adnan, J. A. Putra, M. D. Agustningsih, E. Oktaviana, and N. A. R. Adawiyah, “Exploring the usability of platforms for individuals with visual impairments: A systematic literature review,” *Frontiers in Computer Science*, vol. 7, 2025.

11. M. M. Moncy, M. Pilli, M. Somasundaram, S. Purkayastha, and C. R. Fulton, “Evaluation of accessibility of open-source EHRs for visually impaired users,” *AMIA Annual Symposium Proceedings*, pp. 1165–1174, 2024.

12. A. Khasawneh, P. Gallagher, J. Jarrett, and E. Jacobson, “Evaluating accessibility in telehealth platforms using heuristic review: A case study of the Pathways application,” *Human Factors in Healthcare*, vol. 9, p. 100119, Jun. 2026.

13. S. H. Kerdar, L. Bachler, and B. M. Kirchhoff, “The accessibility of digital technologies for people with visual impairment and blindness: A scoping review,” *Discover Computing*, vol. 27, no. 1, Art. no. 24, Aug. 2024.