

Title: AI Accessibility: Unified Support for Deaf and Blind Users

Authors: Souvik Dan, Sarthak Gupta, Kunal Raj, Ayush Kumar, Chitra Banerjee*

Affiliation: Vivekananda Global University

Date: October 2, 2025

Abstract

Despite two decades of accessibility standards, the digital world remains largely unnavigable for millions with sensory disabilities, a problem defined not just by a lack of access, but a lack of *meaningful usability*—a "second-level digital divide." Existing assistive technologies are often rigid, placing a significant cognitive burden on users to interpret poorly structured content. This paper introduces the Dynamic Accessibility Transformation Engine (DATE), a novel, AI-driven framework that actively re-engineers web content into accessible, multi-modal formats. Our methodology combines a constructivist grounded theory analysis of user needs with a human-centered AI design process. The developed prototype, an intelligent chatbot integrated into a web application, addresses the critical "semantic gap" in current AI by not just recognizing content but interpreting its context to provide synthesized outputs like contextual image descriptions, thematic summaries, and synchronized audio-visual transcriptions. A mixed-methods usability study was conducted with 10 participants (5 visually impaired, 5 deaf). The results show that the DATE prototype significantly reduced cognitive load, as measured by the NASA-TLX index, and increased task success rates. Thematic analysis of qualitative data revealed themes of user empowerment and reduced frustration. This research contributes a new model for accessibility—one that shifts the burden of adaptation from the user to a proactive, context-aware computational system, offering a vital step toward a more equitable socio-technical future.

1. Introduction

The 21st century is defined by the internet's integration into every facet of modern life. However, this digital revolution has not been equitable. For an estimated 2.2 billion people globally with vision impairment and 466 million with disabling hearing loss (World Health Organization, 2021), the digital world is often a landscape of profound frustration. This has created a second-level digital divide (Hargittai, 2002). Unlike the first-level divide, which concerns physical access to technology, this more insidious gap concerns the skills, context, and interface design required for *meaningful participation*. The technology is present, but it is not universally usable.

This usability failure stems from a fundamental socio-technical gap: a chronic misalignment between the technical implementation of the web and the diverse cognitive and perceptual needs of its human users. Engineers build technically functional systems that often fail in their social context, a problem underscored by landmark legal cases like the Domino's Pizza ADA lawsuit, which affirmed that digital spaces are places of public accommodation. Despite this, compliance with standards like the Web Content Accessibility Guidelines (WCAG) remains alarmingly low. A 2022 analysis of the top one million home pages found that 96.8% had detectable WCAG 2 failures (WebAIM, 2022).

This forces users with disabilities to pay a heavy "cognitive tax" on every interaction. Current assistive technologies, such as screen readers, are

indispensable but act as literal interpreters of code. When the code is poorly written, the interpretation is nonsensical, forcing the user to become a forensic analyst of broken digital architecture rather than a simple consumer of information. They must expend significant mental energy to piece together fragmented content, a burden their non-disabled peers do not face.

This paper challenges the prevailing paradigm of reactive assistive tools. We propose and evaluate the Dynamic Accessibility Transformation Engine (DATE), a framework that proactively re-engineers inaccessible web content. Instead of merely reading what is present, the DATE framework analyzes, interprets, and transforms content into a coherent, multi-modal, and contextually-aware presentation. Our thesis is that by bridging the semantic gap—the inability of current AI to understand context and narrative—the DATE framework can significantly reduce the cognitive load on users with sensory disabilities and provide a more equitable and efficient model for digital inclusion.

2. Literature Review: The Limits of Programmatic and Assistive Solutions

A critical review of the literature reveals three core limitations in the current accessibility landscape that the DATE framework is designed to address.

2.1. The Brittleness of Programmatic Accessibility
The Web Content Accessibility Guidelines (WCAG 2.1) provide a robust technical blueprint for an accessible web (W3C, 2018). However, its efficacy is contingent upon developer adherence, which is notoriously inconsistent. This gap between standards and practice is a well-documented problem, often attributed to a lack of developer training, tight project deadlines that treat accessibility as an afterthought, and a fundamental misunderstanding of the lived experience of disability (Lazar et al., 2017). This leads to significant "accessibility debt," where non-compliance is baked into the foundation of web applications. The ARIA (Accessible Rich Internet Applications) specification, designed to add a layer of semantic context for screen readers, is a powerful tool but is frequently misused. Incorrectly applied

ARIA tags can create a worse experience than no tags at all, leading to what is known as "ARIA-poisoning," rendering parts of a site completely unusable for screen reader users.

2.2. The Cognitive Ergonomics of Assistive Technologies
Assistive technologies, while essential, introduce their own ergonomic challenges. The primary mode of interaction for screen reader users is "auditory scanning"—a serial, linear process of listening to an interface (Asakawa & Takagi, 2000). This stands in stark contrast to the parallel, holistic visual intake of sighted users. This serial process is exacerbated by the "verbosity problem." For example, on a poorly coded menu, a screen reader might announce: "Link, Graphic, Spacer, Text: Home, Clickable, List item 1 of 5." A well-coded menu would simply announce: "Home, link, 1 of 5." This extraneous information, multiplied across hundreds of interactive elements, imposes a significant cognitive load, forcing users to mentally filter a constant stream of noise to find the relevant signal. For deaf and hard of hearing (HoH) users, the cognitive load of processing inaccurate or delayed captions while simultaneously trying to watch visual action creates a split-attention effect that hinders comprehension and memory retention (Gernsbacher, 2015). Gernsbacher's research notably found that captions are broadly beneficial for all users, including second-language learners and those in noisy environments, yet they remain an under-implemented feature.

2.3. The Semantic Gap in AI for Accessibility
Modern AI has made remarkable strides. Microsoft's Seeing AI can recognize text and identify objects in real-time. However, these systems primarily perform recognition, not interpretation, suffering from a significant semantic gap (Semaan & Snively, 2010). For instance, an AI might analyze a news photo and accurately identify its components: "a person in a suit, a podium, a crowd, flags." This is a literal description of objects. It fails, however, to interpret the crucial semantic context: "a political leader is giving a campaign speech at a rally." This deeper understanding of intent, narrative, and social context is the missing link. Without it, AI-generated descriptions are merely lists of objects, not meaningful explanations. This gap is the primary

barrier preventing AI from moving beyond simple description to true comprehension and summarization, a gap the DATE framework is explicitly designed to address by synthesizing visual data with surrounding textual context.

3. Methodology: Human-Centered AI Design

Our research employed a mixed-methods approach rooted in a transdisciplinary, participatory design philosophy. We explicitly chose this path to ensure the technology was driven by the lived experiences of the user community, not by technological solutionism.

3.1. Phase 1: Grounded Theory Needs Assessment

We began with a constructivist grounded theory approach, chosen for its ability to let theory emerge directly from the data. We analyzed 50 user-generated threads on Reddit (r/blind, r/deaf) and technology forums. Through iterative coding and memoing, the core theoretical construct that emerged was the "burden of synthesis." This describes the constant, exhausting mental effort users expend to merge incomplete information from multiple, imperfect channels (e.g., trying to align inaccurate captions with on-screen action, or trying to understand an unlabeled chart based on vague descriptions in the text). This construct became the central problem our prototype needed to solve.

3.2. Phase 2: The DATE Architecture The DATE prototype was architected as a client-server web application. Its three core components are:

1. **DOM Parser & Semantic Analyzer:** This module moves beyond simply reading HTML tags. It analyzes the Document Object Model (DOM) structure, CSS class names (e.g., `class="article-body"`), and element relationships to infer the semantic purpose of content blocks. This allows it to distinguish primary articles from supplementary sidebars, promotional banners, and decorative images with high accuracy.

2. **Multi-Modal AI Pipeline:** When a user issues a command, the relevant content is fed into a parallel processing pipeline. For a video, the audio stream is routed to the Web Speech API for transcription, while the visual frames are sampled and sent to a Hugging Face vision-language model for descriptive captioning of non-verbal action. The results are then time-stamped and synchronized.

3. **User Preference Modulator:** A settings panel allows users to define their output preferences (e.g., "highly concise summaries vs. detailed summaries," "verbose vs. brief image descriptions," "high-contrast captioning"). The AI pipeline's output is then tailored to these stored parameters before being rendered.

3.3. Phase 3: Prototype Development: The AURA Web Application The functional prototype is a web application with an integrated AI chatbot as the primary user interface. This provides a single, consistent point of interaction for all accessibility requests.

- **Technology Stack:** React.js (hosted on Vercel), Node.js/Express (hosted on Render), MongoDB Atlas (Free M0 Cluster), Hugging Face Inference API (for vision-language and summarization models), and the browser-native Web Speech API.
- **User Workflow Example** (`/summarize` command):
 1. The user activates the AURA chatbot overlay on a news article page.
 2. The user types the command `/summarize`.
 3. The DOM Parser & Semantic Analyzer identifies the main `<article>` container on the page, automatically filtering out navigation menus, ads, and footers.
 4. The text content within this container is scraped and sent to the Node.js backend.

- The backend sends the text to a summarization model via the Hugging Face API (e.g., facebook/bart-large-cnn).
- The concise summary is returned to the frontend.
- The chatbot displays the summary in a clean, high-contrast text box and, using the Web Speech API, reads it aloud to the user. This dual-mode output serves both visually impaired and some deaf/HoH users who may prefer textual summaries.

3.4. Phase 4: User Evaluation We conducted a remote usability study with 10 participants (5 visually impaired expert screen reader users, 5 deaf/HoH users who rely on captions). Each participant performed tasks on a sample news website. We measured quantitative metrics and conducted qualitative interviews.

- Quantitative Metrics:** Task completion time, error rates, and cognitive load measured using the NASA-TLX (Task Load Index). The NASA-TLX is a multidimensional assessment that measures workload across six subscales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration.
- Qualitative Metrics:** Post-session semi-structured interviews were recorded, transcribed, and analyzed using Braun & Clarke's (2006) rigorous thematic analysis method to identify recurring patterns in user feedback.

4. Results

The evaluation yielded significant quantitative and qualitative data supporting the AURA framework's efficacy in creating a more accessible and less burdensome user experience.

4.1. Cognitive Load and Performance The most striking result was the dramatic reduction in cognitive load when using the DATE prototype. The overall NASA-TLX workload score, a composite measure of the effort required to use a system, decreased by 45%. As detailed in Table 1, the largest reductions were observed in the "Mental Demand" and "Frustration" subscales, indicating that the system was not only easier to use but also subjectively less stressful.

Table 1: NASA-TLX Cognitive Load Index (Avg. Score 0-100, N=10)

NASA-TLX Subscale	Traditional Tools	DATE Prototype	Reduction
Mental Demand	85.2	41.5	-51.3%
Temporal Demand	60.1	45.0	-25.1%
Effort	78.8	38.6	-51.0%
Performance (Satisfaction)		88.2	+117.8%
Frustration	75.6	15.2	-79.8%
Overall Workload Score	68.0 (High)	37.7 (Low)	-44.6%

This reduction in load correlated with improved performance. Task completion time was reduced by an average of 38%, and task success rates (completing the task without any assistance or errors) improved from 85% to 98%.

4.2. Qualitative Themes Our thematic analysis of the interview transcripts revealed two central, deeply resonant themes that explain the "why" behind the quantitative results.

Table 2: Qualitative Feedback Thematic Analysis

Theme	Description	Description	Illustrative Quote	Illustrative Quote	Illustrative Quote
Fundamental Shift	Participants described a fundamental shift in their role. They felt like simple 'consumers' of information.	Participants described a fundamental shift in their role. They felt like simple 'consumers' of information.	I didn't have to "figure out" the page. I just asked the chatbot what it told me. The amount of information saved is incredible. I was jaded, not fighting with the website.	I didn't have to "figure out" the page. I just asked the chatbot what it told me. The amount of information saved is incredible. I was jaded, not fighting with the website.	I didn't have to "figure out" the page. I just asked the chatbot what it told me. The amount of information saved is incredible. I was jaded, not fighting with the website.
Trustworthy Synthesis	Participants expressed trust in synthesized information by DATE, which they described as a "nightmare" example.	Participants expressed trust in synthesized information by DATE, which they described as a "nightmare" example.	It's the first time an AI just told me what I needed to know. It's probably a "nightmare" example.	It's the first time an AI just told me what I needed to know. It's probably a "nightmare" example.	It's the first time an AI just told me what I needed to know. It's probably a "nightmare" example.

5. Discussion

The study's findings have significant implications for the future of digital accessibility.

5.1. Bridging the Semantic Gap The success of the contextual description feature, praised by users in our qualitative analysis, provides strong evidence that the DATE framework offers a viable path toward bridging the semantic gap (Semaan & Snively, 2010). By synthesizing visual data from an image model with linguistic data from the surrounding text, the system moves beyond literal object recognition toward a rudimentary form of narrative interpretation. This is a critical and novel step, suggesting that future AI for accessibility must be context-aware, not just content-aware.

5.2. A Paradigm Shift in Cognitive Load The dramatic reduction in the NASA-TLX workload score is the study's most significant finding. It provides empirical validation that the DATE framework successfully shifts the "burden of synthesis" from the human user to the computational system. This directly addresses the core ergonomic challenges of assistive technologies, such as the high cognitive load of auditory scanning and the split-attention effect of misaligned captions, as identified in the literature (Asakawa & Takagi, 2000; Gernsbacher, 2015). This represents a paradigm shift: from tools that merely present data to tools that provide synthesized knowledge.

5.3. The "Curb-Cut Effect" and the Specter of Algorithmic Bias Many features of the DATE prototype, such as on-demand summarization and smart content extraction, demonstrate the "curb-cut effect": designing for the margins creates benefits for everyone. A tool that summarizes a long article is valuable not only for a screen reader user but also for any user short on time. However, this potential must be tempered by a critical awareness of algorithmic bias. The large language and vision models that power DATE are trained on vast internet datasets, which contain and can amplify societal biases. An image describer may be less accurate for non-Western cultural contexts, and a summarizer may misrepresent the nuance of an article on a sensitive topic. The ethical imperative for future development is to actively audit and mitigate these biases through techniques like adversarial testing and the use of more diverse and representative training data.

6. Conclusion and Future Work

This research has introduced and validated the Dynamic Accessibility Transformation Engine (DATE), a novel framework that prioritizes context-awareness and multi-modal transformation. Our mixed-methods evaluation demonstrates that this approach is not only feasible using freely available AI tools but is also highly effective in reducing cognitive load, improving performance, and empowering users with sensory disabilities. The primary contribution of this work is the conceptual and empirical shift from a reactive, overlay-based model of accessibility to a proactive, integrated system that performs the labor of interpretation on behalf of the user.

While this study serves as a compelling proof-of-concept, it is a first step. Future work will proceed along three critical tracks:

1. Mitigating Bias and Enhancing Fairness: A significant research effort is required to audit the AI pipeline for demographic and cultural biases and to implement fairness-aware machine learning techniques. This is an essential ethical prerequisite for real-world deployment.
2. Personalization and Learning: Future iterations will develop more sophisticated user preference models that can learn and adapt to an individual's needs over time. A user who frequently skips long descriptions could be served more concise outputs automatically.
3. Ubiquitous Integration via Browser Extension: To maximize impact, the DATE framework must be re-engineered from a destination web application into a lightweight browser extension. This would allow the engine to operate universally across the web, transforming any site into an accessible experience and truly fulfilling the promise of a user-centric accessibility model.

7. References

Asakawa, C., & Takagi, H. (2000). *Annotation-based transcoding for nonvisual web access*. In Proceedings of the fourth international ACM conference on Assistive technologies (pp. 172-179).

Braun, V., & Clarke, V. (2006). *Using thematic analysis in psychology*. *Qualitative Research in Psychology*, 3(2), 77-101.

Gernsbacher, M. A. (2015). *Video captions benefit everyone*. *Policy Insights from the Behavioral and Brain Sciences*, 2(1), 195-202.

Hargittai, E. (2002). *Second-level digital divide: Differences in people's online skills*. *First Monday*, 7(4).

Lazar, J., Goldstein, D. F., & Taylor, A. (2017). *Ensuring digital accessibility through process and policy*. Morgan Kaufmann.

Petrie, H., & Kheir, O. (2014). *The relationship between accessibility and usability of websites*. *ACM SIGCHI Bulletin*, 39(1), 1-10.

Semaan, P., & Snaveley, N. (2010). *The semantic gap in computer vision*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8), 1362-1377.

W3C. (2018). *Web Content Accessibility Guidelines (WCAG) 2.1*. W3C Recommendation. Retrieved from <https://www.w3.org/TR/WCAG21/>

WebAIM. (2022). *The WebAIM Million: The 2022 report on the accessibility of the top 1,000,000 home pages*. Retrieved from <https://webaim.org/projects/million/>

World Health Organization. (2021). *Blindness and vision impairment*.