

From Cluttered to Clear: Improving the Web Accessibility Design for Screen Reader Users in E-commerce With Generative AI

Yaman Yu
School of Information Sciences
University of Illinois at Urbana
Champaign
Champaign, Illinois, USA
Mercari R4D
Tokyo, Japan
yamanyu2@illinois.edu

Bektur Ryskeldiev
Mercari R4D
Tokyo, Japan
Research and Development Center for
Digital Nature
University of Tsukuba
Tsukuba, Japan
bektour@mercari.com

Ayaka Tsutsui
Digital Nature Group
University of Tsukuba
Tsukuba, Japan
Mercari R4D
Tokyo, Japan
ayakatsutsui1128@digitalnature.slis.tsukuba.ac.jp

Matthew Gillingham
Mercari R4D
Tokyo, Japan
matthew.gillingham@mercari.com

Yang Wang
University of Illinois at
Urbana-Champaign
Champaign, Illinois, USA
yvw@illinois.edu

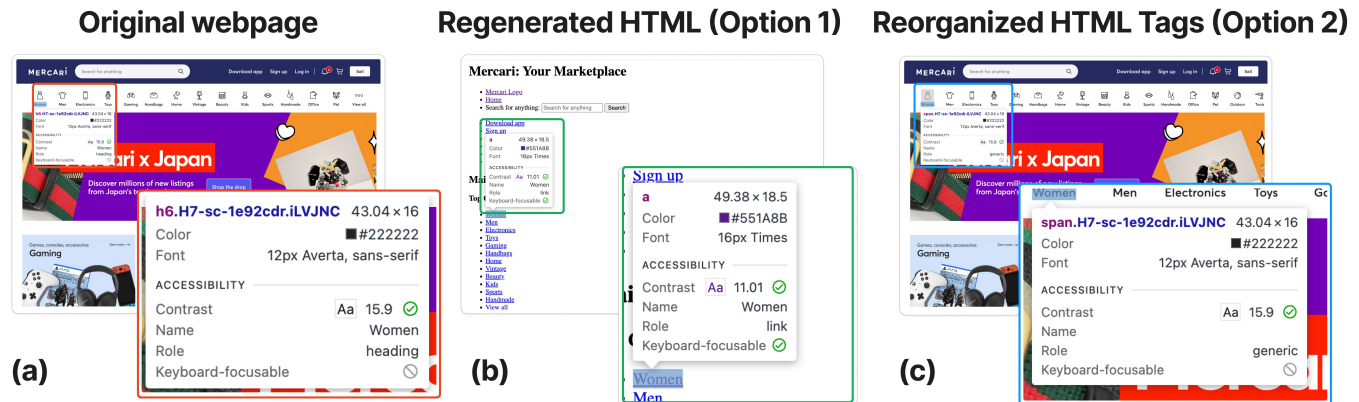


Figure 1: From left to right: (a) the original website, (b) Option 1 (Regenerated HTML), and (c) Option 2 (Reorganized HTML Tags). Option 1 is generated using a GenAI-powered extension that rewrites the HTML to enhance accessibility, while Option 2 reorganizes the original HTML tags to address accessibility issues without altering the visual design. This example highlights the changes made in both versions compared to the original website. Both versions simplify navigation by removing single-category headings, reducing clutter and minimizing fatigue for screen reader users.

Abstract

Blind and low vision users often face significant barriers when navigating online shopping websites using screen readers. Complex layouts, unclear content hierarchies, and visually driven designs create a frustrating and inefficient browsing experience, particularly on unfamiliar platforms. While prior accessibility tools focus on isolated elements such as product descriptions or image alt text, they often fall short of addressing the structural and navigational challenges screen reader users encounter across entire webpages. In this work, we explore how Generative AI (GenAI) can be leveraged

to improve the accessibility of shopping websites by automatically restructuring their HTML content. We conducted a three-phase study: formative interviews with screen reader users, system development of a GenAI-powered browser extension, and user evaluation through both automated audits and real-world testing. Our tool dynamically reorganizes web content to better align with screen reader navigation patterns. Results from user studies with blind and low vision participants show that the GenAI-generated pages significantly improve navigation efficiency, content clarity, and overall usability. Participants highlighted benefits such as more logical section order and reduced browsing fatigue. Our findings demonstrate the potential of GenAI to support comprehensive, user-centered accessibility improvements directly within the structure of existing websites.



This work is licensed under a Creative Commons Attribution 4.0 International License.
ASSETS '25, Denver, CO, USA

© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0676-9/25/10
<https://doi.org/10.1145/3663547.3746353>

CCS Concepts

- **Human-centered computing** → **Accessibility systems and tools**;
- **Computing methodologies** → **Artificial intelligence**;
- **Applied computing** → **Online shopping**.

ACM Reference Format:

Yaman Yu, Bektur Ryskeldiev, Ayaka Tsutsui, Matthew Gillingham, and Yang Wang. 2025. From Cluttered to Clear: Improving the Web Accessibility Design for Screen Reader Users in E-commerce With Generative AI. In *The 27th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '25)*, October 26–29, 2025, Denver, CO, USA. ACM, New York, NY, USA, 19 pages. <https://doi.org/10.1145/3663547.3746353>

1 Introduction

People who are blind or have low vision rely on the Internet to access information and perform everyday tasks, such as shopping for essential items and managing personal finances [46]. However, many blind or low vision users face significant challenges when navigating websites with screen readers, which are tools that convert text to speech for auditory access [41]. Common issues include inaccessible table formats [36], inadequate image descriptions that lack important visual context [19], insufficient navigation aids [29], and confusing or unclear explanations [13, 16]. These accessibility barriers can lead to frustration and a sense of helplessness for blind or low vision users when interacting with websites [4].

For blind or low vision users, navigating unfamiliar websites can be particularly challenging due to the need to understand and interact with unique website elements and layouts, which can be confusing and time-consuming. This problem is especially evident on shopping websites, where visually appealing but complex designs often take precedence over accessibility [22]. Shopping websites frequently use intricate layouts, dynamic content, and numerous filters or categories that may not be accessible to screen readers, further complicating the navigation experience for blind or low vision users. In addition, shopping tasks such as finding product details, accessing reviews, managing shopping carts, and checking out involve multiple steps and specific interactions that are difficult to execute without clear and consistent navigation aids. Although simplified website designs are recommended to improve accessibility [27], the lack of inclusive design on many online shopping platforms makes it difficult for blind or low vision users to perform common tasks, such as comparing products across multiple sites, exploring deals or reading detailed product information.

Researchers have explored various methods to improve web accessibility for blind or low vision users within the field of Human-Computer Interaction (HCI). Studies have focused on creating interactive systems to help users navigate specific elements on product pages or retrieve essential information using AI-based assistants [46, 56]. However, these existing solutions primarily target specific elements, such as product descriptions or image retrieval, rather than ensuring a comprehensive, end-to-end accessible experience across the entire web page, particularly for users who may be unfamiliar with a new platform. They often rely on external tools or assistants that interact with existing site structures, which may not fully align with users' browsing needs and preferences. To fill this gap, we propose using Generative AI (GenAI) to automate the optimization of shopping websites for screen reader users. Our

research goes beyond previous efforts by focusing on improving the overall structure and content of websites, ensuring a more inclusive and user-friendly experience for blind or low vision users using a screen reader. Our study aims to answer the following research questions (RQs):

- RQ1:** What accessibility barriers do screen reader users encounter when navigating shopping websites?
- RQ2:** How can a GenAI-powered tool that restructures HTML address these barriers and improve screen reader usability?
- RQ3:** How effective is the proposed GenAI-based tool in improving navigation, clarity, and task efficiency for screen reader users?

To address these research questions, we conducted a three-phase study that involved formative interviews, system development, and user evaluations. In the formative study, we interviewed six screen reader users to identify the specific challenges they face on shopping websites and to understand their current strategies to overcome these barriers. The insights gained from this phase informed the development of a GenAI-based tool that automatically revises HTML content to improve accessibility for screen reader users. The tool was implemented as a browser extension that dynamically reorganizes web page elements to align with screen reader navigation patterns. Finally, we evaluated the tool through automated accessibility audits and user testing on Mercari, a legitimate and unfamiliar English-language shopping site selected to minimize bias from participants' prior experience.

Our findings indicate that the proposed GenAI solution significantly improves accessibility by restructuring web page content in ways that make it more intuitive and easier to navigate for blind or low vision users. In user evaluations, seven screen reader users tested both the original and revised versions of a shopping website. The results demonstrated that our tool's regenerated HTML version led to more efficient navigation and a clearer understanding of website hierarchy compared to the original version. The participants particularly appreciated the logical reordering of sections, the addition of summary headings, and enhanced navigation flexibility, highlighting the potential of our approach to provide a more comprehensive and inclusive web experience for blind or low vision users.

Our study makes several important contributions to the field of web accessibility. First, we provide an in-depth analysis of the specific accessibility barriers that screen reader users encounter on shopping websites, particularly on non-mainstream platforms with complex layouts and dynamic content. This analysis extends the current understanding of the unique challenges faced by blind or low vision users in navigating shopping environments, such as inconsistent header navigation across different websites, browsing and item comparison fatigue, and how blind or low vision users adopt the latest GenAI technologies to complement the lack of product information in images and description. Second, we introduce the application of GenAI for enhancing web accessibility. Unlike previous approaches that focus on isolated elements or rely on external tools, our GenAI-based solution offers an automated, comprehensive optimization of website structure and content directly on the web page. Third, we developed a browser extension and validated that GenAI can improve the accessibility of shopping websites through

evaluations using both automated tools and screen reader users. Finally, our user evaluations show that the revised HTML content generated by our extension leads to substantial improvements in usability and navigation efficiency for screen reader users.

2 Related Work

2.1 Online (Shopping) Accessibility Challenges

Efforts to improve shopping accessibility can be categorized as offline and online. Offline challenges have been addressed through third-party assistance [63] and integrated technologies [64, 68], yet blind or low vision users with mobility limitations still face difficulties visiting physical stores independently [47, 61]. As a result, many rely on websites for everyday shopping and information gathering [46]. However, online shopping presents its own barriers. Navigating unfamiliar websites is often time-consuming and frustrating for screen reader users due to visual-centric design and mouse-oriented interfaces [9, 41]. Screen readers may misread or skip important content, creating a sense of confusion and helplessness [4]. Prior studies have identified key accessibility issues such as missing navigation aids [29], poorly designed tables [36], lack of alternative text [6, 10, 19, 67], and overly verbose descriptions [13, 16]. These problems limit access to online information resources [30, 32]. Design inconsistencies across websites further increase the burden, particularly for new users or those exploring unfamiliar platforms [27]. Shopping websites, with their complex layouts and varied content types [22], intensify these challenges. Inadequate image descriptions and cluttered structures often make it difficult to retrieve information effectively [44, 60]. To address these issues, HCI researchers have explored various strategies to improve web accessibility for visually impaired users, as detailed in the following sections.

2.2 Improving Web Browsing Accessibility

Prior work has introduced tools to improve web accessibility for visually impaired users, particularly in online shopping contexts. For instance, Wang et al. developed a browser extension for Amazon that extracts product descriptions to help visually impaired users understand item appearance through a question-answering system [56], while Stangl et al. created a system that allows screen reader users to ask questions about products and receive specific, relevant information [46]. These systems highlight the benefits of command-based interaction and targeted information retrieval. However, they often introduce a steep learning curve, requiring users to adapt to an additional interface layer instead of browsing the website naturally. Concerns about trust and the reliability of extracted information also remain. Beyond shopping-specific tools, broader web accessibility efforts have focused on enhancing existing content. WebInSight [7] and SADie [21] insert alternative text or semantic markup, while other tools address specific components, such as command-based browsing [5] and table conversion [58]. Although useful, these approaches address isolated accessibility issues and do not improve the overall content structure. Since effective screen reader navigation depends on coherent page-wide hierarchy and logical content flow, isolated fixes often fall short of supporting seamless interaction.

2.3 AI-Driven Web Accessibility and Design

Recent work has explored AI-driven methods to improve web accessibility, including restructuring data tables [58], converting PDFs to HTML [54, 55], and generating image and audio descriptions [23, 31, 57, 65, 66]. Tools have also supported video comprehension for visually impaired users [50]. Due to the time-intensive nature of manual remediation [2], automated approaches—such as ChatGPT-based accessibility repair [40], evaluation tools [48], and context-aware image description systems [18]—have gained traction. However, Large Language Model (LLM)s face challenges such as hallucinations [1, 8] and the risk of introducing new barriers when not designed with users in mind, as seen in accessibility overlays that disrupt screen readers [15, 34]. Recent studies have begun addressing accessibility in specific contexts such as C2C shopping, improving communication and listing quality [20, 42, 43, 49].

Building on this direction, our research proposes a GenAI-based approach to automate accessibility design for shopping websites. Unlike prior efforts focused on isolated HTML fixes [12, 24], we aim to use LLM to regenerate or reorganize the HTML of existing websites, creating a clearer information hierarchy and facilitating navigation that aligns with screen reader users' natural browsing behaviors, without introducing an additional layer to learn.

3 Methodology Overview

To address our research questions (RQ), we designed a structured methodology comprising four key steps, summarized in Figure 2. In step ①, we conducted a formative study to uncover the challenges screen reader users face, their coping strategies, and their specific needs when navigating shopping websites. These findings informed step ②, where we developed a browser extension system powered by GenAI to automatically revise website HTML for improved accessibility. In steps ③ and ④, we evaluated the improved webpages through technical assessments and user evaluations to validate both their accessibility enhancements and content integrity.

Participants recruitment. A total of 21 participants from the United States participated in our study, with 6 in the formative study and 15 in the user evaluation. Participants were recruited via a screening survey shared on platforms like Prolific¹, as well as mailing lists and word of mouth. The survey assessed vision ability, screen reader usage, and online shopping frequency to identify blind or low vision users of shopping websites. User evaluation participants were distinct from those in the formative study to avoid priming bias. Each participant received a \$25 Amazon gift card per hour. Study sessions were audio-recorded, transcribed, and conducted with participants' consent. The study was approved by the Institutional Review Board (IRB).

4 Formative Study

Prior research has examined the accessibility barriers that blind or low vision users face on commonly used shopping websites [44, 56], primarily focusing on identifying challenges and improving image accessibility. One study [46] specifically noted that blind or low vision users experience difficulties accessing product information due to missing alt text for images and non-compliance with Web

¹<https://www.prolific.com/>

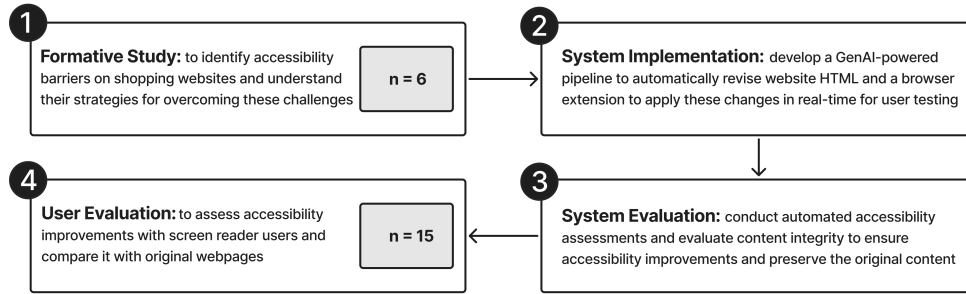


Figure 2: Methodology Overview—Note that participants of the user evaluation (step 4) do not overlap with those in steps 1.

Content Accessibility Guidelines (WCAG) [53]. However, there is limited research on the specific challenges that blind or low vision users face throughout their shopping process due to website design and development, especially on new or unfamiliar websites. Addressing this gap is crucial, as blind or low vision users often want to explore a broader range of shopping options and compare products but are limited to a very narrow selection of accessible websites. There is also a need to understand their strategies for overcoming these challenges and their desired solutions to address these issues. We conducted a formative study to explore these questions.

4.1 Participants Demographics

The formative study included six blind or low vision users who regularly used screen readers for online shopping. These participants, aged 25 to 54 years, were recruited through social networks and mailing lists. Half of them were legally blind ($n = 3$), and the other half had significant visual impairments requiring assistive technologies ($n = 3$). Four participants self-identified as male and two as female. Beyond basic demographics, we collected detailed background information on their online shopping habits to include participants with varying online shopping frequencies (ranging from once a week to 2–3 times a week). However, aside from shopping frequency, our participant sample was fairly homogeneous (e.g., all from similar demographics), which we acknowledge as a limitation. This included people who shop online once a week and those who did so 2–3 times a week, on various e-commerce platforms. More details are provided in the demographic Table 1.

Participants’ screen reader experience varied: some were power users familiar with many hotkeys, while others had more basic proficiency. Furthermore, we did not explicitly record preferred shopping platforms (mobile versus desktop), but several participants mentioned using mobile screen readers (e.g. TalkBack or VoiceOver) in daily life.

4.2 Semi-Structured Interview Procedure

The interviews were conducted online using the Zoom video calling platform and lasted between 45–60 minutes. At the beginning of each interview, we obtained participants’ consent for the study and for recording the session. The interview was divided into two parts. In the first part, we asked participants about their experiences with online shopping, the challenges they encountered, the accessibility issues they faced on previous shopping websites, how they handled

these issues, and their suggestions for improvements. In the second part, participants were asked to “*screen share and think aloud*” while browsing a new shopping website, Mercari. We selected Mercari because it is a legitimate, English-language shopping platform that is not widely known among our participant group. This helped ensure that participants were unfamiliar with the site, allowing us to better observe their natural exploration behaviors and minimize prior experience bias in evaluating accessibility. We instructed them to start by browsing the homepage and exploring the website, then presented them with a scenario of searching for and selecting a blender on the website. Participants were asked to think aloud about the keys they used and their thoughts on the browsing experience with screen readers while performing these tasks.

4.3 Data Analysis

All interview recordings from our formative study were transcribed using Zoom. We employed a deductive thematic analysis approach [11], guided by our initial research questions such as identifying accessibility challenges on unfamiliar shopping websites, but we also remained open to new, inductively identified themes. Three authors independently reviewed two randomly selected transcripts, writing down preliminary codes that captured both explicit statements (for example, “lack of comprehensive alt text”) and implicit ideas (for example, “user frustration with new page layouts”). After this initial pass, the three coders met to reconcile differences, merge overlapping ideas, and refine code definitions into a single preliminary codebook. Each coder then applied this refined codebook to the remaining transcripts, continuing to add or revise codes when new concepts emerged. Throughout the coding process, all coders met regularly to resolve any discrepancies by discussing how best to interpret particular quotes in light of the existing codebook. This process yielded around 30 themes, such as “Accessibility Issues from Improper HTML Tags” and “Preference for Independent Browsing,” which were grouped under broad categories like “Navigation Barriers,” “Workarounds,” and “Desired Enhancements.”

4.4 Interview Findings

All participants encountered accessibility barriers when using online shopping websites. These barriers included inappropriate use of HTML tags and labels, a lack of comprehensive alt text for images, and an offline return process. In the following sections, we will discuss each of these challenges in detail and report the strategies participants used to counter them.

ID	Age	Gender	Visual Ability	Online shopping Frequency	Used Screen Reader
P1	25-34	Male	Completely blind	Once a week	NVDA, Google Talkback
P2	25-34	Male	Blind	Once a week	JAWS, NVDA, VoiceOver
P3	35-44	Male	Totally blind	Once a week	JAWS, VoiceOver
P4	25-34	Female	I have light perception	Once a week	JAWS, NVDA, VoiceOver
P5	25-34	Male	Only some light perception	2-3 times a week	JAWS, NVDA, VoiceOver
P6	45-54	Female	Limited light perception	Once a week	JAWS, NVDA

Table 1: Formative Study Participants Demographic Table

4.4.1 Current Accessibility Challenges on Shopping Websites.

Inappropriate Use of HTML Tags and Labels. All participants highlighted the inappropriate use of HTML tags and labels as the most common accessibility issue on shopping websites. Screen reader users reported relying on shortcut keys to navigate HTML components efficiently. P1 explained, *“Most of the time, if I enter a website, I press H to check if they have headings or not, because this is the fastest and easiest way for us to browse.”* Common shortcuts include pressing “H” to iterate through all headings or numbers 1 to 6 to navigate specific heading levels [28, 59, 62]. Misuse of HTML tags like headings frustrates users with excessive, irrelevant information, making it hard to locate key details or understand crucial information needed for purchase decisions.

Too many headings. Five out of six participants reported that shopping websites often feel overcrowded due to excessive emphasis on non-essential information through headings, causing confusion about the website’s structure and flow, particularly during initial browsing. Overuse of headings makes it hard for screen reader users to navigate quickly, as they must listen to irrelevant content. For instance, some websites place long lists of category headings, like electronics or home, before the main content, forcing users to navigate through numerous headings—often over ten—before reaching key information. P2 noted, *“Categories don’t have to be headings; there could just be a heading titled ‘Categories,’ with items listed as links underneath.”* Similarly, P1 mentioned, *“Sometimes all the filters are marked as headings, making it take a long time to reach the product search results.”*

Lack of headings. On the contrary, some informal yet essential information is not placed in headings or appropriate HTML tags, misaligning with screen reader users’ navigation preferences. Screen reader users typically jump between headings, links, and buttons to quickly access key information [59, 62]. P3 explained, *“When I am searching for something, I really expect it to be in a heading.”* For example, some shopping websites fail to include product titles in headings. Participants often attempted to navigate by headings to locate products but, failing to do so, resorted to using the down arrow key to iterate through elements on the search results page. P4 noted, *“Some smaller store websites are not accessible at all; they are too graphical with not enough headings and links to navigate.”*

Disorganized Heading Hierarchy. Even when appropriate headings are used, inconsistent heading levels and poorly structured layouts can still pose challenges for screen reader users. For

instance, a website might assign a level 1 heading to reviews and a level 6 heading to product descriptions, confusing users who rely on specific heading levels for navigation.

Many participants navigate headings using the “H” key, following their sequence in the HTML code rather than their visual layout. This often leads to a disjointed experience, as visually oriented layouts may prioritize sighted users. For example, if the layout places images and reviews on the left and product names and details on the right, screen reader users will hear about reviews and vague image descriptions first, instead of the crucial product names and details. Screen reader users may encounter irrelevant information, like reviews or vague image descriptions, before crucial product details. P5 noted, *“I just don’t understand why unnecessary information like likes, comments, and share buttons is positioned before the product information.”*

Unclear Labels on Images and Buttons. Participants reported frustration with the lack of clear, descriptive labels for buttons and images, a common issue on less mainstream shopping websites. Unclear labels confuse screen reader users, who rely on meaningful descriptions to navigate interactive elements. For instance, a button might be labeled generically as “button”, providing no indication of its function, such as “add to cart” or “view details.” Similarly, images may lack alt text, leaving their content or relevance unclear. This forces users to guess button functions or spend unnecessary time exploring, reducing usability and accessibility issues. P2 noted, *“Amazon has a few unlabeled buttons that can be a bit confusing sometimes. You can usually figure them out from the context, but they are not labeled correctly.”*

Lack of comprehensive and accurate text and image description for products. Participants noted difficulties understanding product details on shopping websites due to insufficient text and image descriptions. Many sites rely heavily on images to convey key information, such as color, style, condition, and features, without providing adequate text. This creates a significant barrier for screen reader users, who depend on alt text to interpret images. P2 shared, *“The condition of items on eBay is hard to identify because it’s all about images.”* Participants also mentioned inconsistencies between product descriptions and reviews, adding to the confusion. P3 explained, *“When it comes to clothing, I want more detail than just ‘blue shirt.’ I need to know about the fit and style, which is often missing in online descriptions.”*

Hard to compare different products. Three out of six participants reported significant challenges comparing similar products

on shopping websites. While sighted users can quickly assess visual and descriptive information, screen reader users must repeatedly switch between tabs to find key details like prices, descriptions, and specifications. This process is time-consuming and mentally exhausting. P3 explained, *“Comparing products, especially similar ones, is just exhausting. I have to remember all the details in my head and switch back and forth between tabs to check things like the price or specific features.”* Improper use of HTML tags and labels further complicates navigation, making the task even more inefficient. P2 similarly noted, *“It’s really hard to compare products that are very similar. I have to keep switching between tabs to find the price or some specific details. It’s not efficient at all.”*

4.4.2 Strategies for Overcoming Accessibility Challenges.

Relying on External Video Reviews. Participants reported relying on external resources, such as YouTube reviews, to counter accessibility challenges and gain a clearer understanding of products. These video reviews provide in-depth demonstrations and visual details that are often lacking in the brief text descriptions found on shopping sites. For example, P4 mentioned, *“When I’m not sure about how a product looks or functions, I usually check YouTube. The videos give me a much better idea than the short descriptions on the website.”* This approach helps participants make more informed decisions, especially when considering the purchase of products where visual details are important, such as electronics or clothing.

Utilizing User Ratings and Reviews Additionally, participants closely follow user ratings and reviews on shopping platforms to gauge the condition and quality of products. Since product images or brief descriptions can be misleading or incomplete, user reviews become a crucial source of information. These reviews often provide first-hand experiences and specific details that are not immediately visible or described in the official product listings. As P2 explained, *“The product pictures don’t always tell the full story. I go through the reviews to see what people are actually saying about the product condition and quality.”*

Seeking Assistance from Sighted Individuals Participants might ask sighted friends or family members for assistance in describing visual details that are important for making purchase decisions. This practice is necessary when the product’s visual information is critical but inaccessible. P5 shared their experience, stating, *“Sometimes I have to ask someone to look at it to make sure I am seeing the right thing, especially when it is not clear on the website.”* This indicates a reliance on others when the digital content is insufficient, which adds another layer of dependency and effort for screen reader users.

Leveraging GenAI for Accessibility Enhancements Interestingly, most of the participants (4 out of 6) have used GenAI to help mitigate accessibility challenges on shopping websites. They reported using GenAI tools such as ChatGPT [37] and Be My Eyes [38] (specifically the app’s AI-based ‘Be My AI’ feature) to get detailed descriptions of product images and clarify ambiguous information that was not mentioned in the text description or even in the image alt text. P5 shared that *“I used ChatGPT to describe pictures of the product and got a better idea of what the product consists of. It would be great to have an AI that can answer questions about how a product*

looks.” Participants also used GenAI to compare different models of a product by generating a navigable table that highlights key differences, making it easier to understand these differences and make informed purchase decisions. P6 also shared, *“I usually make the AI compare two products by giving it links or screenshots of the pictures to find the differences that are not mentioned in the descriptions.”* However, they also reported limited trust in AI to conduct product searches because it often does not bring up results from preferred websites such as Amazon. P6 explained, *“I do not trust AI to search for products on Amazon. I usually search for the product myself and then use AI to make sure it has the features I am looking for.”*

Preferred Improvements for Screen Reader Users Participants preferred browsing websites themselves over using virtual assistants, which have been proposed as potential solutions [52]. They cited concerns about reliability, language compatibility, and additional accessibility barriers. P1 explained, *“I prefer browsing myself because I can control what I do with my screen reader.”* P2 noted issues with virtual assistants, stating, *“If you are in the same window with a virtual assistant, you are not aware of new messages. That is not friendly for screen reader users at all.”* Instead, participants suggested improvements like a screen reader mode that simplifies web pages by showing only essential elements, such as product descriptions. P3 shared, *“It would be helpful if shopping websites had a screen reader mode that strips away images and just shows the product descriptions.”* This highlights their preference for streamlined, user-controlled experiences that focus on relevant information.

4.5 Design Implications

Our findings strongly support our design intuition that leveraging GenAI to improve HTML tags and labels can significantly enhance accessibility for screen reader users. Participants already use GenAI to tackle certain accessibility challenges, such as extracting detailed descriptions from images or comparing similar products. However, directing GenAI to effectively revise the layout and information hierarchy through HTML tags is challenging and often unfeasible for them. This difficulty highlights the need for more intuitive, user-friendly solutions that seamlessly integrate with existing browsing practices, reducing the burden on users to manually adjust or work around accessibility issues.

Participants expressed a preference for accessibility solutions that operate within their familiar browsing environments rather than relying on external tools that add layers of complexity, such as AI assistants designed to help locate information. This suggests a need for back-end solutions where developers revise a website’s structure to better align with how screen reader users navigate, creating a more seamless and integrated user experience. However, this is a challenging task for developers, as it requires specialized knowledge of both accessibility standards and screen reader behavior, along with substantial time and effort to implement these changes manually.

We developed a GenAI-driven tool to enhance online shopping website accessibility by optimizing HTML tags, labels, and information hierarchy. The system addresses common barriers for screen reader users, such as inconsistent headings, missing alt text, and unclear button labels. By automating these adjustments, the tool

reduces developer effort and bridges knowledge gaps, offering a more efficient way to create accessible web experiences.

5 System Implementation

5.1 System Overview and Purpose

To address the accessibility challenges identified in our formative study, we proposed the concept of using Generative AI (GenAI) to dynamically restructure webpage content for screen reader users. While participants identified a lack of descriptive image labels as an issue, we excluded image description generation, as participants and prior research have already addressed this using GenAI [23, 46]. Our approach aims to improve navigation efficiency, logical content hierarchy, and labeling of interactive elements. To test this concept, we developed a browser extension as a proof-of-concept system to evaluate its feasibility and gather user feedback on preferences and experiences².

The system is designed to operate in two modes:

- (1) **Rewrite the Document: Regenerated HTML Version (Option 1).** In this option, we produce an HTML document which is re-created from scratch to provide an accessible screen reader experience. The content is preserved, but all other aspects of the site can be modified. The elements can be re-ordered, and visual design aspects may be removed, since the usability of the screen reader is the only criteria under consideration.
- (2) **Replace the Tags: Reorganized HTML Tags Version (Option 2).** The second option adjusts existing tags without altering the visual layout, focusing on minimal changes to preserve content integrity while improving navigability. This goal is achieved by re-writing part of the website's Document Object Model (DOM) to reorganize HTML tags, fix misused elements, and provide better labels.

5.2 Technical Architecture

We developed a Google Chrome Extension and tested it in Google Chrome Official Build 127.0.6533.122 (arm64). Google Chrome was the most widely-used browser and had largest number of screen readers deployed in a desktop context [59]. The pipeline of our system is applicable to other browsers and computing environments, including mobile web browsers that support plugins. We implemented the prototype leveraging Chrome's Web APIs for real-time DOM access and manipulation. The extension extracts webpage content, processes it with GPT-4o via the OpenAI API, and reinserts the modified HTML or tag updates back into the live webpage. We selected GPT-4o for its support of large input/output sizes (128K/16K tokens), which is critical for processing full HTML pages. The system processes content in segments, preserving coherence through sequential prompting and chunk-level context injection.

- **Option 1 (see Figure 3)** shows how the system decomposes a webpage into chunks that fit the model's context window, sends these chunks to GPT-4o along with prior context, and reconstructs the full page by stitching outputs together.

- **Option 2 (see Figure 4)** follows a similar pipeline but focuses on tag-level editing instead of complete rewriting. The model receives structured segments and returns modified HTML tags, which are directly updated in place. This method preserves the original visual layout and reduces the risk of content discrepancies.

We configured the LLM with deterministic parameters (*temperature* = 0.2, *max_tokens* = 16,384, *top_p* = 1, *frequency_penalty* = 0, and *presence_penalty* = 0) to ensure reproducibility. Furthermore, we also tried our prototype with similar models such as Google Gemini 1.5 Pro [17] and Anthropic Claude 3.5 Sonnet [3], and found that in all cases the models could generate valid HTML code consistent with the intended prompt. Therefore, we believe that our approach can be sufficiently robust in multiple advanced LLMs. We ultimately chose GPT-4o for our implementation due to its strong performance, wide availability, and efficient integration with our existing development environment. Additional implementation details and prompt templates are available in Appendix A.5.2.

To illustrate the types of enhancements made under both versions, Figure 1 presents a before-and-after example showing how HTML is transformed for improved accessibility for both versions. While Option 1 applies broader restructuring to the entire page, Option 2 preserves layout while correcting semantic structure. A summary of representative changes for each version is provided in Appendix A.5.

5.2.1 Mitigating Content Discrepancies in LLM-Generated Accessibility Enhancements. Large language models (LLMs) can effectively revise and improve accessibility design; however, they are prone to content discrepancies [33], where certain details may be unintentionally omitted or altered when regenerating a webpage. In our context, this can manifest as missing content (e.g., product links) or unintended text in the HTML output. To address these challenges, we adopted two key strategies. First, we refined prompts and implemented a similarity check function to identify discrepancies between the regenerated page and the original. When discrepancies were detected, such as missing product links on extensive pages like Amazon's search results, we programmatically reinserted these details to ensure consistency with the source. The LLM-generated result was only rendered on the front end if the similarity measure exceeded a threshold of 90%. This threshold was chosen based on preliminary tests, which showed that values below 90% often reflected differences due to reformatting rather than critical content changes. Second, we limited some revisions to HTML tags only in our Option 2 approach, which minimized the risk of losing key information but also reduced our ability to reorder visual layouts for better accessibility (see Section 7).

6 System Evaluation

6.1 Automatic Accessibility Evaluation

To assess whether our GenAI-enhanced webpages comply with WCAG guidelines, we conducted an automatic audit using three widely adopted accessibility testing tools, Google Lighthouse³, Sort-Site⁴, and AChecker⁵. We ran each tool on the original webpages

²GitHub repository: <https://github.com/yutouzai/GenAIforA11y>

³<https://developer.chrome.com/docs/lighthouse/overview>

⁴<https://www.powermapper.com/products/sortsite/>

⁵<https://websitesaccessibilitychecker.com/checker/index.php>

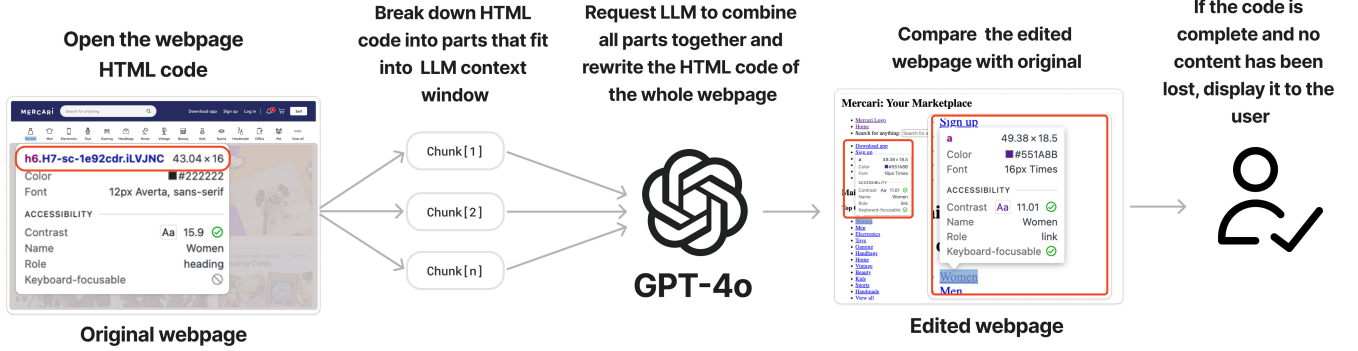


Figure 3: Option 1 pipeline overview, prompting GPT-4o model to completely rewrite HTML code

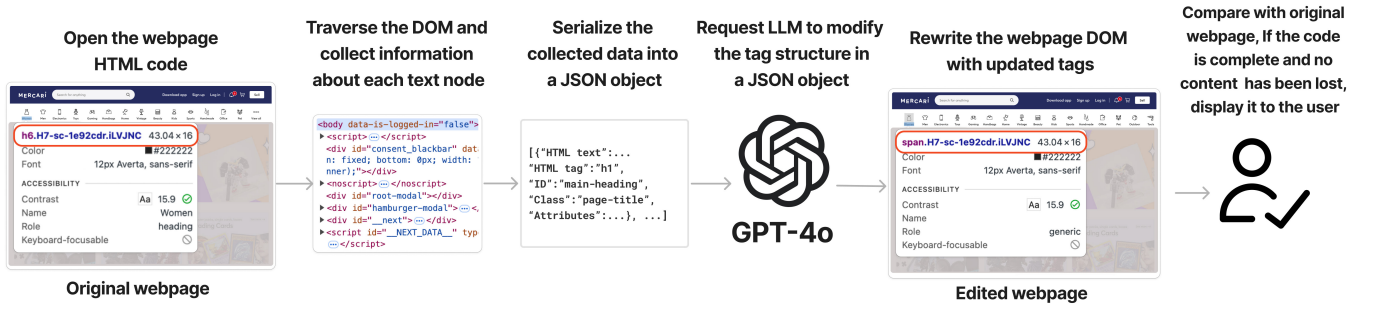


Figure 4: Option 2 pipeline overview, prompting GPT-4o to rewrite only some parts of the HTML code

and their LLM-regenerated and LLM-reorganized version. For the evaluation, we chose three e-commerce websites—Amazon, Nordstrom, and Mercari—based and tested three common pages: Home page, Search Results Page, and Product Page. This selection was informed by user interviews in the formative study, which provided insights into platform familiarity and usage patterns. Specifically, we included: a widely used platform (e.g., Amazon), a platform that users recognized but rarely used (e.g., Nordstrom), and a platform that was largely unfamiliar (e.g., Mercari). This range allowed us to capture a variety of accessibility challenges experienced by screen reader users across websites with differing levels of user familiarity and interface complexity. As shown in Table 2, both the LLM-regenerated and LLM-reorganized versions consistently reduced the number of accessibility violations across all three tools (including mismatching ARIA⁶ attributes and their roles, as well as heading elements not following a sequentially descending order). For example, on the Amazon Product Page, Google Lighthouse reported a drop from 16 violations in the original to one in the regenerated version and three in the reorganized version. Similarly, SortSite reported five violations on the Nordstrom Home Page in the original, which dropped to one in both the regenerated and reorganized versions. Across all platforms and page types, the regenerated version showed the most substantial improvements, particularly in reducing issues related to heading hierarchy, ARIA roles, and unlabeled interactive elements.

⁶<https://www.w3.org/WAI/standards-guidelines/aria/>

6.2 Content Integrity Evaluation

While improving accessibility, it is equally important that the regenerated webpages preserve the original content. To evaluate this, we measured Aggregated Semantic Similarity between the original and LLM-generated versions. We selected this metric because it aligns with the real-world experience of screen reader users, and is recommended in Microsoft’s AI evaluation guidelines [35] and supported by prior research [14]. This metric captures whether the meaning and intent of text within HTML elements remain consistent, even when phrasing or structure changes.

To ensure meaningful alignment, we extracted all screen-reader-accessible HTML elements, such as visible text, ARIA labels, and alternative text, from both the original and regenerated versions, and compared their aggregated semantic embeddings. Unlike string-level comparisons, this metric accounts for paraphrasing, content reordering, and summarization—which are common outcomes of LLM-based regeneration. For instance, moving “Product Details” above “Customer Reviews” or collapsing redundant filter headings may improve usability while maintaining semantic equivalence.

We applied this evaluation to the same three websites and page types used in the accessibility audit. Since Option 2 only modifies HTML tags without changing content, we focused the analysis on Option 1 (Regenerated HTML). As shown in Table 3, the regenerated pages achieved high semantic similarity scores, frequently exceeding 95%, indicating strong content preservation. In a few cases where the scores were slightly lower, such as the Nordstrom

Page	Category	Lighthouse	SortSite	AChecker	Page	Category	Lighthouse	SortSite	AChecker
Mercari Home Pages	Original	6	4	6	Nordstrom Home Pages	Original	3	5	3
	Option 1	1	0	0		Option 1	1	1	0
	Option 2	2	2	3		Option 2	1	1	1
Mercari Search Result Pages	Original	5	5	6	Nordstrom Search Result Pages	Original	3	3	3
	Option 1	1	0	1		Option 1	1	1	0
	Option 2	1	2	2		Option 2	2	1	1
Mercari Product Pages	Original	6	4	5	Nordstrom Product Pages	Original	2	6	3
	Option 1	1	1	1		Option 1	1	1	0
	Option 2	2	2	3		Option 2	1	3	1
Amazon Home Pages	Original	6	6	7	Amazon Search Result Pages	Original	12	17	13
	Option 1	1	1	2		Option 1	1	0	1
	Option 2	2	2	3		Option 2	3	4	3
Amazon Product Pages	Original	16	18	15					
	Option 1	1	1	1					
	Option 2	3	4	4					

Table 2: Comparison of Level A accessibility issues detected on various webpages using three automated evaluation tools: Lighthouse, SortSite, and AChecker (lower is better). A lower number indicates improved accessibility. The table shows the number of accessibility problems identified for the original version of each page, Regenerated HTML Version (Option1), and Reorganized HTML Tags Version (Option2). While Option 1 achieved near-perfect results in most cases, minor issues like one/two broken links persisted. Option 2 addressed structural problems but retained challenges such as inaccessible heading sequences and visual design conflicts due to its inability to alter the original layout or content order.

homepage, we observed that the LLM had condensed or reorganized content—for example, grouping scattered promotional items or combining fragmented text elements into single, coherent sections. These changes slightly reduced the similarity score, but reflected practical adjustments that enhance clarity and navigation for screen reader users.

7 User Evaluation

While automated audits provide useful indicators of accessibility, prior research cautions against relying solely on these tools due to inconsistencies in their coverage and reliability [51]. To complement our technical evaluation and better understand the real-world impact of our system, we conducted a user evaluation with screen reader users. We focused the user evaluation on Mercari.com because it is a legitimate but unfamiliar English-language shopping site, which helps minimize prior experience bias. In addition, since each session lasted up to 90 minutes, we limited testing to one website to avoid overburdening participants and to allow for in-depth comparison of the three versions.

7.1 Participants Demographics

15 screen reader users participated in the study, consisting of nine females and six males. The participants’ ages ranged from 18 to 44 years. All demographic data were self-reported by the participants (see Table 4). Among the participants, 11 were completely blind, one had limited vision, and three could perceive only light. Participant recruitment details are in section 3.

7.2 Study Procedure

We conducted a within-subjects user evaluation over Zoom, lasting 60-90 minutes. This duration included time for setting up the experiment (e.g., installing an extension), completing all three tasks across the three website versions, filling out an evaluation survey after

each version (three surveys total), and engaging in think-aloud protocols and discussions with researchers to capture detailed feedback and insights. At the beginning of the study, we obtained participants’ consent for both participation and recording. We started with general questions about their previous experiences with online shopping and screen reader usage. This was followed by user testing of three different versions of the Mercari.com website: (1) the original website, (2) the LLM-regenerated HTML version implemented through Extension Option 1, and (3) the LLM-reorganized HTML tags version implemented through Extension Option 2, as described in Section 5.

To prevent learning effects, we counterbalanced the sequence of versions each participant tested. The testing sequence for each participant is detailed in Table 5. Drawing from prior research [26], we designed three tasks that simulate key navigation and decision-making processes, reflecting real-world shopping scenarios.

For each version, participants started by browsing the homepage to familiarize themselves with the layout of a new shopping website (Task 1). Once they completed this, they were asked to imagine they wanted to buy a blender and search for it on the site. We chose a blender as it is a gender-neutral item, helping us avoid content bias that might influence the browsing experience. Participants then reviewed the search results, exploring the products as they typically would when searching on an e-commerce website, and selected a product they were interested in exploring further (Task 2). Finally, they clicked on their chosen product and browsed the product detail page to gather the information they needed (Task 3).

During these tasks, participants were encouraged to browse as they normally would while using their screen readers and to think aloud about their experience. This included sharing how they navigated the site, any challenges they encountered, and their overall thoughts on the browsing experience.

Platform	Page Type	Aggregated Semantic Similarity (%)
Mercari	Home Page	99.36
Mercari	Search Results Page	97.09
Mercari	Product Page	98.35
Nordstrom	Home Page	91.60
Nordstrom	Search Results Page	93.19
Nordstrom	Product Page	92.06
Amazon	Home Page	97.47
Amazon	Search Results Page	98.56
Amazon	Product Page	98.89

Table 3: Aggregated Semantic Similarity scores between the original and LLM-regenerated webpages across three e-commerce platforms (Amazon, Nordstrom, and Mercari) and three page types (Home, Search Results, and Product). These results demonstrate strong preservation of content meaning, even when structure and phrasing were altered to improve accessibility. Lower scores, such as on the Nordstrom homepage, reflect cases where the LLM reorganized or merged content for better screen reader usability.

ID	Age	Gender	Visual Ability	Online shopping Frequency	Used Screen Reader
P1	25 - 34	Female	I am totally blind	4-6 times a week	JAWS, NVDA, VoiceOver
P2	25 - 34	Female	Completely blind	4-6 times a week	JAWS
P3	25 - 34	Female	Blind with light perception	Once a week	JAWS, NVDA, Google Talkback
P4	18 - 24	Female	No vision	Once a week	JAWS, NVDA, VoiceOver
P5	35 - 44	Female	Totally blind	Once a week	JAWS, NVDA
P6	25 - 34	Male	Blind, light perception only	2-3 times a week	JAWS, NVDA
P7	25 - 34	Male	I have limited vision	2-3 times a week	JAWS
P8	25 - 34	Male	Fully blind	4-6 times a week	JAWS, NVDA, Google Talkback, VoiceOver
P9	35 - 44	Female	Legally blind	4-6 times a week	JAWS, Google Talkback, VoiceOver
P10	35 - 44	Male	Blind	Once a week	JAWS, VoiceOver
P11	25 - 34	Female	Completely blind	2-3 times a week	JAWS, NVDA, Google Talkback, VoiceOver
P12	25 - 34	Female	Zero vision	Once a week	JAWS, NVDA, Google Talkback, VoiceOver
P13	25 - 34	Female	Blind but can perceive light	Once a week	JAWS, NVDA, Google Talkback, VoiceOver
P14	35 - 44	Male	Legally blind	2-3 times a week	JAWS, VoiceOver
P15	25 - 34	Male	Blind	Once a week	JAWS, Google Talkback

Table 4: User Evaluation Participant Demographic Table

Testing Sequence	Participants
HTML → Tag → Original	P4, P12
HTML → Original → Tag	P2, P11, P15
Tag → HTML → Original	P5, P9
Tag → Original → HTML	P6, P10, P14
Original → HTML → Tag	P1, P7, P13
Original → Tag → HTML	P3, P8

Table 5: Sequence of Version Testing by Participants

(1) **Task 1: Homepage** Participants navigated the homepage of the shopping website using a screen reader as they normally would.

(2) **Task 2: Search Page** Participants used the screen reader to search for a blender on the website and browsed through the search results.

- (3) **Task 3: Product Page** Participants used the screen reader to view the details of a specific product they were interested in purchasing.

To better understand how well each version addressed participants' accessibility needs and navigation preferences on e-commerce webpages, we designed a brief Likert-scale survey with five questions. These questions were adapted from a previous study [26] and informed by findings from our formative study, tailored specifically to accessibility concerns and shopping scenarios.

- (1) **Q1:** How would you rate your overall experience using the screen reader on this website?
- (2) **Q2:** How clear and meaningful did you find the headings on this website?
- (3) **Q3:** How helpful were the headings in understanding the hierarchy of content on this website?
- (4) **Q4:** How efficiently could you locate key sections or information on this website using headings?
- (5) **Q5:** How easy was it to access key features (e.g., search bar, product categories, account options) using your screen reader on this website?

Participants rated each question on a 5-point likert scale, where 1 indicated "Very Difficult / Poor" and 5 indicated "Very Easy / Good." To ensure consistent evaluation of accessibility and usability, participants completed the survey after testing each version, resulting in three sets of responses within a single interview. After participants completed browsing all three versions, we asked follow-up questions such as how similar they found the content across the three versions, how they would compare these versions, and to rank them based on their experience. Finally, we debriefed participants about the differences between the versions and asked for additional suggestions for design improvements.

7.3 Data Analysis

We analyzed the user evaluation data (think-aloud and post-task interviews transcripts) using an inductive thematic analysis [11]. Two researchers independently read through each transcript, developing preliminary code set grounded in participants' navigational preference and comparative opinions of each website version. These codes included recurring observation related to preferred heading structure, button labeling and reorganized content sequence. They met each week during the coding process to discuss about new codes and reconcile the discrepancies on code definition and scope. Ultimately, they identified approximately 50 themes, including "Trust in Consistent Labeling" and "Heading Overload".

7.4 Quantitative Results

7.4.1 Statistical Analysis of Likert Scale Survey Results. To analyze the results from the 5-question survey for each version, We conducted Wilcoxon signed-rank tests for all pairwise comparisons across five questions (Q1–Q5), comparing three different conditions: Original, HTML, and Tag. To account for the increased risk of Type I errors due to multiple comparisons, a Bonferroni correction was applied for each question, adjusting the significance threshold to $\alpha_{\text{adjusted}} = 0.0167$. Results indicate that the HTML version consistently outperformed the Original website across all five questions, while the Tag version showed significant improvements only in the

overall user experience (Q1) compared to the Original. The comparisons between HTML and Tag versions revealed a significant difference only in Q1, suggesting that HTML enhancements had a more pronounced impact on improving user experience than Tag reorganizations.

Overall User Experience (Q1): Participants rated the overall user experience higher for both HTML and Tag versions compared to the Original website. The Wilcoxon signed-rank tests showed significant differences between the Original website and HTML ($W = 0$, $p = 0.00585$) and between the Original website and Tag ($W = 0$, $p = 0.00585$). Additionally, the comparison between HTML and Tag ($W = 0$, $p = 0.01170$) was also significant. All these p-values are below the adjusted significance threshold ($\alpha_{\text{adjusted}} = 0.0167$), indicating that both HTML and Tag versions significantly improved the overall navigation experience compared to the Original website.

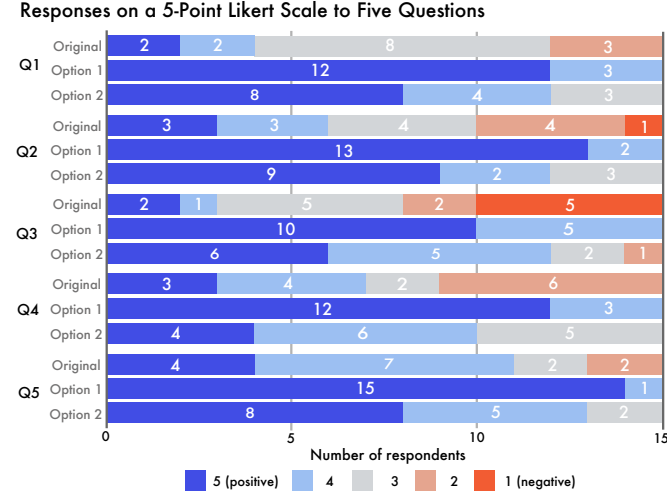
Clarity of Headings (Q2): Clarity of headings was rated higher in both HTML and Tag versions compared to the Original website. The comparison between the Original website and HTML ($W = 0$, $p = 0.00585$) was significant, indicating improved clarity. However, the comparison between the Original website and Tag ($W = 0$, $p = 0.04500$) and between HTML and Tag ($W = 0$, $p = 0.18750$) were not significant after Bonferroni correction.

Understanding of Content Hierarchy (Q3): Participants reported improved understanding of content hierarchy in both HTML and Tag versions compared to the Original website. The comparison between the Original website and HTML ($W = 0$, $p = 0.00585$) was significant, indicating a better understanding. However, the comparisons between the Original website and Tag ($W = 0$, $p = 0.04500$) and between HTML and Tag ($W = 5$, $p = 0.18750$) were not significant after correction.

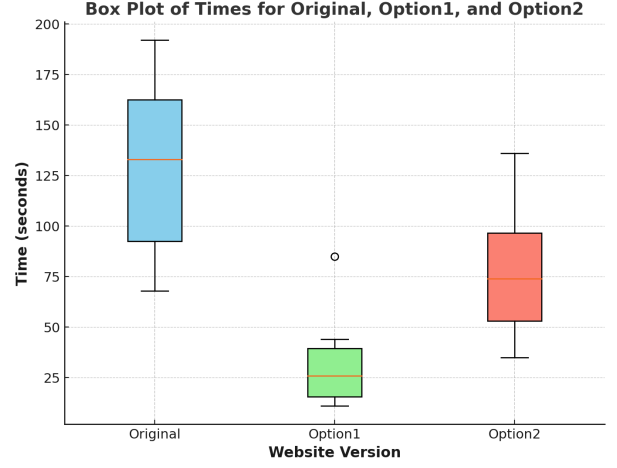
Efficiency in Locating Key Sections (Q4): Option HTML outperformed the Original website in efficiency ($W = 0$, $p = 0.00585$), which was significant. Comparisons between the Original website and Tag ($W = 2.5$, $p = 0.16470$) and between HTML and Tag ($W = 0$, $p = 0.17550$) were not significant after Bonferroni correction.

Ease of Access to Key Features (Q5): Ease of access to key features improved in the HTML version compared to the Original website ($W = 0$, $p = 0.00585$), which was significant. However, comparisons between the Original website and Tag ($W = 5.5$, $p = 0.10260$) and between HTML and Tag ($W = 0$, $p = 0.35100$) did not reach statistical significance after correction.

7.4.2 Task Completion Time in Different Version. We instructed participants to complete tasks 1–3 that simulated their real-life shopping experiences. They were encouraged to freely browse products without a strict goal of locating specific information or identifying a particular product on the search results page. Participants were also prompted to think aloud as much as possible during the tasks, sharing their thoughts on accessibility and navigation. Given this open-ended setup, comparing the time it took participants to complete each task (e.g., browsing the homepage or search results page) would be unreasonable, which would not accurately reflect the differences in accessibility design between versions. Some participants naturally verbalized more thoughts or explored more products on the search results page than others. However, in Task 3, while browsing the product details page, all participants consistently began by locating the product description.



(a) The 5-point Likert scale results from Q1 to Q5, as answered by 15 participants when testing the three versions. From top to bottom, “Original” refers to the original website, “Option 1” refers to the LLM-regenerated HTML version implemented through Extension Option 1, and “Option 2” refers to the LLM-reorganized HTML tags version implemented through Extension Option 2.



(b) Box plot comparing task completion times across the three versions (Original, Option 1, and Option 2), showing variability in performance and highlighting differences in accessibility design.

Figure 5: Comparison of participant responses and task performance across the three webpage versions.

This behavior allowed us to calculate the time each participant spent finding the product description as a reliable quantitative metric for comparing the accessibility of different versions [25]. All participants naturally prioritized this step, making it a reliable and meaningful metric for comparing accessibility design across the three versions. The task completion times, as visualized in the box plot (Figure 5b), reveal significant differences across the three webpage versions: Original, Option 1 (Regenerated HTML), and Option 2 (Reorganized HTML Tags).

To analyze these differences, we conducted a Friedman test followed by Wilcoxon signed-rank tests for pairwise comparisons. Given that three pairwise tests were performed, we applied the Bonferroni correction to control for the increased risk of Type I errors. Specifically, the original significance level ($\alpha = 0.05$) was divided by the number of comparisons ($m = 3$), resulting in an adjusted significance level ($\alpha_{\text{adjusted}} = 0.0167$). The Friedman test indicated a significant overall difference among the versions ($\chi^2(2) = 30.00$, $p < 0.001$). Subsequent pairwise comparisons using the Wilcoxon signed-rank test, adjusted with the Bonferroni correction, demonstrated that Option 1 significantly outperformed the Original website ($W = 0.0$, $p < 0.001$) and Option 2 ($W = 0.0$, $p < 0.001$). Additionally, Option 2 showed a significant improvement over the Original version ($W = 0.0$, $p < 0.001$) but exhibited greater variability in task completion times. Option 1 consistently achieved the shortest task completion times, reflecting its effective structural and navigational enhancements. Option 2 provided moderate improvements but was limited by the unchanged layout, while the Original website showed the longest times and the most inconsistencies due to unstructured content and missing navigation labels. These results highlight the importance of combining structural and

usability improvements to create a more accessible and efficient user experience.

7.5 Qualitative Results

7.5.1 Browsing Patterns of Screen Reader Users on Unfamiliar Website. The most commonly used navigation combination by all participants is the *heading key* paired with the *down arrow key*. The *heading key* allows users to jump between headings on a webpage by pressing “H”, similar to how sighted users scan visually for bold or large text to find different sections. The *down arrow key* is used to navigate through HTML elements one by one, enabling users to read and explore content sequentially. When visiting a new website, participants typically began by pressing the down arrow to read line by line and discover whether the page made effective use of headings. Once they recognized headings were present, they switched to pressing *heading key* (H) to move among heading levels more swiftly. If they found a heading relevant (for instance, product details labeled with an `<h3>` or `<h4>`), they returned to the down arrow key to read underlying elements in sequence. Sometimes, they may also navigate by specific heading levels (number keys 1–6). For example, if they browse a product title labeled as a heading level 2, and then want to read the product details, they might press the number 3 key to jump to the next heading level, assuming the details are structured logically under that level. In addition to the heading and down arrow keys, some participants also mentioned using the *button key* and *edit key* to interact with interactive elements on a webpage. The button key allows users to quickly navigate to buttons, such as “Add to Cart” or “Submit”, while the edit key helps them jump directly to input fields, making it easier to fill out forms or enter search queries.

7.5.2 User Preferences for Regenerated HTML Versions. All participants found both the regenerated HTML version and the reorganized HTML tags version to be more intuitive for navigating with a screen reader, significantly enhancing accessibility. The regenerated HTML version was especially preferred for its logical reordering of sections and enhanced navigational HTML labels. The participants' average evaluation of the screen reader browsing experience on the regenerated HTML version was 5 on a scale from 1 (Very Poor) to 5.00 (Excellent). The average rating for the reorganized HTML tags version was 4.57, while the original website received an average rating of 3.14. 13 participants reported that the regenerated HTML version provided the best experience, even though the reorganized HTML tags version also showed improvements over the original website. P6 described the navigation experience on regenerated HTML version as "highly structured. Even for people who just started using screen readers, they can navigate quickly without confusion." P3 also mentioned that "(regenerated HTML) version has a title in the heading 'Purchase Options' that summarizes buttons and information related to purchasing, which is something I've never seen on other websites." One participant felt that both versions offered an equally good screen reader navigation experience, while another preferred the reorganized HTML tags version over the regenerated HTML version.

7.5.3 Accessibility Challenges of Original Website. While browsing the original website, participants often experienced inconvenience and confusion due to the inappropriate use of HTML tags and a layout that did not align with screen reader browsing patterns. P2 noted, "The website is cluttered, maybe like too many uses of the headings, but there's nothing I could do." Participants also acknowledged that this was a common problem on other mainstream shopping websites. Although it is not convenient for screen reader users, they have become accustomed to it and consider it acceptable. As P2 explained, "Amazon is an insanely cluttered website. There are so many headings. The only way to navigate Amazon is to just arrow down each item instead of using headings." Participants also reported frustration when important information, such as product titles, was not marked as headings. For example, P3 noted, "So the problem is the product itself is not a heading, but a lot of other information that is not necessarily needed is in headings." Additionally, the structure and sequence of the website posed challenges for participants in understanding and navigation. For instance, some websites place product images and reviews on the left, while product names and details are on the right. This layout may make sense from a sighted user's perspective due to visual design principles, but it creates confusion for blind or low vision users. Because review titles are also marked as headings, when users press "H" to navigate through headings, reviews appear before the product title. P4 expressed confusion with this structure, "It is good that they put reviews in headings, but not as the first heading we encounter, because it serves as a signal that it is already the end part of the product information." P9 also expressed concern about the complexity of the layout, stating, "Whenever I hit H, it would bring me to a completely different part of the page."

7.5.4 Correct Labeling for Essential Information. Participants perceived that both the regenerated HTML version (Option 2) and the reorganized HTML tags version (version 3) improved the heading

structure compared to the original website. They rated the helpfulness of headings for understanding the website's hierarchy with an average score (Q3) of 4.86 for the regenerated HTML version and 4.43 for the reorganized HTML tags version, in contrast to a score of 2.71 for the original version. Participants reported significant improvements in both versions 2 and 3 in two key aspects: (1) unnecessary headings, such as individual category names (e.g., women, men, sports), were removed. For example, P2 explained, "Those categories are just regular links [on the regenerated HTML version and reorganized HTML tags version], and it is faster to navigate through the page because you can easily skip that whole section of the website if it's not what I'm interested in." (2) Important information, like product names, was correctly labeled as headings. P4 mentioned, "I see that links, such as the title of the product, are also displayed as headings [on the regenerated HTML version and reorganized HTML tags version], which is quite convenient for me because now I can simply press the number key for the product." P13 noted that important information appears as headings, stating, "Headings are very much in the key part of the page, and they really help me quickly navigate, revealing that they also assist in navigation."

7.5.5 Why Regenerated HTML Works Better. The regenerated HTML version (Option 2, $Q4 \bar{x} = 4.86$) was perceived as even better than the reorganized HTML tags version (version 3, $Q4 \bar{x} = 4.28$) in terms of locating essential information and understanding the hierarchy of website content. In Option 2, the LLM regenerated the entire HTML file, leading to several unique improvements that version 3 could not achieve:

- (1) **Logical Reordering of Sections:** The regenerated HTML reorganized the page's sections into a more logical, linear sequence that better suits screen reader navigation. For example, participants observed that in the regenerated version, the reviews and seller profile appear after the product name and details, but before the purchase options. This follows a logical order that aligns with the linear browsing patterns of screen readers. In contrast, on the original website, elements were arranged for visual aesthetics—product images and reviews were positioned on the left side, which in the HTML code would be read first by screen readers, even before the product name and details. Additionally, the purchase options were presented before the product details in the HTML structure. P6 highlighted this improvement, "[The regenerated HTML version] is highly structured. Even for people who just started using screen readers, they can definitely navigate quickly without confusion."
- (2) **Addition of Summary Headings:** The regenerated HTML version added summary text to sections and aligned them as headings, which participants found extremely helpful. These changes made Option 2 stand out not only from the original website but also from other shopping websites. Participants considered this a uniquely nuanced accessibility improvement. As P3 noted, "[Option 2] has a title in the heading 'purchase options' that summarizes buttons and information related to purchasing, something I have never seen on any other websites. Others normally just put all buttons in headings. Now I can quickly jump through the purchase information like 'add to cart' and 'buy now' while browsing; that's impressive." P5

also mentioned *“It is very easy to read because it has both headings and lists.”*

- (3) **Enhanced Navigation Flexibility:** The regenerated version used different labels on the same element, enriching screen readers’ options for navigation. P7 mentioned that each product title is not only a heading but also a link and part of a list structure: *“It’s very flexible to navigate; I can use list navigation, heading navigation, or link navigation.”* P14 noted *“The number of times I had to search through headings to reach the results has significantly decreased.”*

7.5.6 User Perception on Content Similarity. When it comes to self-reported perception of content similarity, all participants noted the high similarity in content between versions, only noting the changes in design or layout of the website. For instance, P9 has mentioned that *“I don’t think there’s any difference. It was just where they were located and how long it took to get to them.”* While P8 mentioned that *“overall, it looked pretty similar.”* and similarly P10 has mentioned *“It definitely seems like a lot of the same content, you know, just in different orders depending on how you click on stuff.”*

8 Discussion

Our formative study revealed key mismatches between shopping website designs and the navigation behaviors of screen reader users. Even when pages followed accessibility guidelines, users still faced confusion due to inconsistent HTML structures, overloaded headings, and poor labeling. These issues disrupted linear navigation, especially on unfamiliar sites, where users struggled to locate essential information within cluttered or visually prioritized layouts.

To address these barriers, we developed a GenAI-based pipeline that reorganizes and regenerates HTML to align with screen reader navigation patterns. This proof-of-concept demonstrated clear improvements in labeling, content hierarchy, and usability. In our user evaluation, participants strongly preferred the revised versions and highlighted preferences for clear section order, consistent headings, and reduced visual clutter.

These findings reinforce the need for accessibility tools that go beyond technical compliance and support the actual browsing logic of screen reader users. By grounding our system design in formative insights, we show how GenAI can deliver practical, user-centered improvements for accessible web design. We reflect on our findings to discuss limitations and opportunities for future work:

8.1 Generative AI as a Tool for Accessibility: Supporting Developers, Not Shifting Responsibility

The proposed Generative AI (GenAI) approach to revising HTML for better accessibility is not intended to absolve companies of their responsibility to design inclusive interfaces. While our implemented browser extension dynamically reorganizes or regenerates HTML to enhance screen reader navigation, it is not intended as a permanent solution that shifts the burden of accessibility onto users. Instead, the primary purpose of our proposed method and implementation serves as a proof-of-concept, which is to demonstrate the potential of such technology in improving web accessibility and to evaluate user preferences for future website content structures.

Companies should take full responsibility for designing accessible websites from the outset and integrating GenAI into their development workflows. By doing so, they can ensure that accessibility improvements are implemented and controlled during the design and testing phases rather than relying on end-users to apply external tools to address these issues. As highlighted in prior research, tools like **accessibility overlays** have been criticized for creating new barriers rather than resolving existing ones, further emphasizing the need for developer-led interventions [6, 34].

By leveraging GenAI responsibly within development pipelines, companies can enhance web accessibility while maintaining control over the design process. This ensures that accessibility improvements are both effective and sustainable, reducing the burden on end-users who rely on assistive technologies. For instance, future work could explore real-time feedback mechanisms that provide accessibility suggestions and quick fixes, allowing developers to continuously refine their systems and enhance their adaptability across different browsing environments. Specifically, the pipeline could be used to generate or reorganize a developer’s HTML product and visualize the changes and reasoning in a sidebar, enabling developers to make informed decisions. This system could then be further evaluated and improved based on developer feedback. By extending the pipeline’s capabilities in these ways, the tool could become both a practical solution for developers looking to create more inclusive web experiences.

8.2 Broadening GenAI Accessibility Enhancements to Diverse Websites

Generative AI (GenAI) offers significant potential for improving web accessibility across diverse contexts, extending beyond the e-commerce platforms explored in this study. While our research primarily focuses on shopping websites, the challenges identified such as poor information hierarchy, inadequate labeling, and navigation inefficiencies are prevalent across many types of websites. These barriers can impact screen reader users navigating other contexts, including news platforms, educational resources, and service-based websites. For example, news platforms often contain complex layouts with articles, headlines, sidebars, and advertisements competing for attention. GenAI could reorganize these sections to create a logical flow that enhances readability for screen reader users. Prior research emphasize the importance of clear and concise content presentation for visually impaired users navigating dense information spaces, which aligns with the potential application of GenAI in this domain [45, 58].

While our system showed promising results on several mainstream e-commerce websites, the generalizability of our approach to more dynamic, JavaScript-heavy platforms remains an open challenge. Sites that rely heavily on asynchronous content loading, infinite scroll, or script-based DOM manipulation may hinder consistent HTML extraction and reinsertion during runtime. Websites with dynamic or user-generated content may require real-time adjustments to maintain accessibility standards. Additionally, different website types have distinct user requirements; for instance, news platforms prioritize readability while service-based sites focus on task efficiency. Ensuring compliance with WCAG guidelines while adapting to diverse contexts is also critical for widespread adoption.

of GenAI solutions. Future research should explore the generalizability of GenAI-powered tools by conducting studies on various website categories to identify specific barriers and refine GenAI models accordingly. Hybrid solutions combining automated GenAI outputs with manual adjustments could address context-specific needs while maintaining flexibility in accessibility improvements. Furthermore, developing tools that allow developers to customize GenAI-generated outputs based on user feedback and accessibility requirements would enhance the practical application of these technologies.

8.3 Risks of Content Loss and Deployment Considerations

While our system showed improvements, LLM-driven HTML restructuring still carries risks, particularly when essential content is lost, altered, or inaccurately generated. Although we implemented semantic similarity checks and content reinsertion to reduce discrepancies, subtle omissions such as missing product links or seller information may still occur. These failures, though sometimes undetected by automated tools, can affect users' ability to complete tasks or make informed decisions.

This concern is especially relevant for screen reader users, who often depend on structured and authoritative content to navigate unfamiliar websites. Prior work suggests that users with disabilities may place greater trust in information presented through formal interfaces like screen readers. As a result, even minor hallucinations, such as incorrect headings or summaries, can lead to confusion or misjudgments. These findings underscore the need to treat GenAI-based accessibility tools as intermediaries of critical content. Developers should consider safeguards including content validation, fallback options, and transparency indicators that alert users to AI-modified sections. Moreover, our evaluation focused on desktop use, and did not account for mobile or tablet navigation scenarios, which are highly relevant for screen reader users. Future work should assess cross-device performance and explore user controls for comparing original and AI-enhanced versions.

9 Conclusion

Accessibility in online shopping remains an open problem for blind or low vision screen reader users. In this study, we investigated the issues experienced by screen reader users when shopping online and developed a web browser plugin that applies GenAI to improve the accessibility of the website for screen readers. Our evaluation shows improvement (as reported by accessibility checkers and self-reported by users) in the navigation and labeling of items and can potentially be used by both users and developers to improve the accessibility of their online shopping experience.

We believe that our work proves that application of GenAI and LLMs in particular can automatically improve the web browsing accessibility and reduce the burden on both users and developers. We believe that our work contributes to the topic of web development and accessibility in the coming age of ubiquitous human-AI interactions, and we hope that our study would contribute to the work of researchers and practitioners in the fields of AI and HCI.

References

- [1] Rudaiba Adnin and Maitraye Das. 2024. "I look at it as the king of knowledge": How Blind People Use and Understand Generative AI Tools. In *Proc. of the 26th Int. ACM SIGACCESS Conf. on Computers and Accessibility* (St. John's, NL, Canada) (ASSETS '24). Association for Computing Machinery, New York, NY, USA, Article 64, 14 pages. <https://doi.org/10.1145/3663548.3675631>
- [2] Aljazi Nasser Al Jabor, Fadi Adnan, Mike Park, and Achraf Othman. 2021. Mada Web Accessibility Monitor Tool. In *2021 8th Int. Conf. on ICT & Accessibility (ICTA)*. 1–5. <https://doi.org/10.1109/ICTA54582.2021.9809423>
- [3] Anthropic. 2024. Claude 3.5 Sonnet. Website. Retrieved April 15 2025 from <https://www.anthropic.com/news/claude-3-5-sonnet>.
- [4] Stacey Menzel Baker, James W Gentry, and Terri L Rittenburg. 2005. Building understanding of the domain of consumer vulnerability. *Journal of macromarketing* 25, 2 (2005), 128–139.
- [5] Natã M. Barbosa, Jordan Hayes, Smirity Kaushik, and Yang Wang. 2022. "Every Website Is a Puzzle!": Facilitating Access to Common Website Features for People with Visual Impairments. *ACM Trans. Access. Comput.* 15, 3, Article 19 (jul 2022), 35 pages. <https://doi.org/10.1145/3519032>
- [6] Jeffrey P Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samuel White, et al. 2010. Vizwiz: nearly real-time answers to visual questions. In *Proc. of the 23rd annual ACM Symp. on User interface software and technology*. 333–342.
- [7] Jeffrey P Bigham, Ryan S Kaminsky, Richard E Ladner, Oscar M Danielsson, and Gordon L Hempton. 2006. WebNLSight: making web images accessible. In *Proc. of the 8th Int. ACM SIGACCESS Conf. on Computers and Accessibility*. 181–188.
- [8] Jeffrey P Bigham, Irene Lin, and Saiph Savage. 2017. The Effects of "Not Knowing What You Don't Know" on Web Accessibility for Blind Web Users. In *Proc. of the 19th Int. ACM SIGACCESS Int. on computers and accessibility*. 101–109.
- [9] Yevgen Borodin, Jeffrey P. Bigham, Glenn Dausch, and I. V. Ramakrishnan. 2010. More than meets the eye: a survey of screen-reader browsing strategies. In *Proc. of the 2010 Int. Cross Disciplinary Conf. on Web Accessibility (W4A)* (Raleigh, North Carolina) (W4A '10). Association for Computing Machinery, New York, NY, USA, Article 13, 10 pages. <https://doi.org/10.1145/1805986.1806005>
- [10] Erin Brady, Meredith Ringel Morris, Yu Zhong, Samuel White, and Jeffrey P. Bigham. 2013. Visual challenges in the everyday lives of blind people. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems* (Paris, France) (CHI '13). Association for Computing Machinery, New York, NY, USA, 2117–2126. <https://doi.org/10.1145/2470654.2481291>
- [11] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
- [12] Jeremy T. Brudvik, Jeffrey P. Bigham, Anna C. Cavender, and Richard E. Ladner. 2008. Hunting for headings: sighted labeling vs. automatic classification of headings. In *Proc. of the 10th Int. ACM SIGACCESS Conf. on Computers and Accessibility* (Halifax, Nova Scotia, Canada) (Assets '08). Association for Computing Machinery, New York, NY, USA, 201–208. <https://doi.org/10.1145/1414471.1414508>
- [13] Michael Crystian Nepomuceno Carvalho, Felipe Silva Dias, Aline Grazielle Silva Reis, and André Pimenta Freire. 2018. Accessibility and usability problems encountered on websites and applications in mobile devices by blind and normal-vision users. In *Proc. of the 33rd Annual ACM Symp. on Applied Computing* (Pau, France) (SAC '18). Association for Computing Machinery, New York, NY, USA, 2022–2029. <https://doi.org/10.1145/3167132.3167349>
- [14] Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde De Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, et al. 2021. Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374* (2021).
- [15] Niklas Egger, Gottfried Zimmermann, and Christophe Strobbe. 2022. Overlay Tools as a Support for Accessible Websites – Possibilities and Limitations. In *Computers Helping People with Special Needs: 18th Int. Conf., ICCHP-AAATE 2022, Lecco, Italy, July 11–15, 2022, Proc., Part II* (Milan, Italy). Springer-Verlag, Berlin, Heidelberg, 6–17. https://doi.org/10.1007/978-3-031-08645-8_2
- [16] Ombretta Gaggi, Giacomo Quadrio, and Armir Bujari. 2019. Accessibility for the Visually Impaired: State of the Art and Open Issues. In *2019 16th IEEE Annual Consumer Communications & Networking Conf. (CCNC)*. 1–6. <https://doi.org/10.1109/CCNC.2019.8651676>
- [17] Google. 2024. Generate content with the Gemini API. Website. Retrieved September 10 2024 from <https://cloud.google.com/vertex-ai/generative-ai/docs/model-reference/inference>.
- [18] Ananya Gubbi Mohanbabu and Amy Pavel. 2024. Context-Aware Image Descriptions for Web Accessibility. In *Proc. of the 26th Int. ACM SIGACCESS Conf. on Computers and Accessibility* (St. John's, NL, Canada) (ASSETS '24). Association for Computing Machinery, New York, NY, USA, Article 62, 17 pages. <https://doi.org/10.1145/3663548.3675658>
- [19] Darren Guinness, Edward Cutrell, and Meredith Ringel Morris. 2018. Caption Crawler: Enabling Reusable Alternative Text Descriptions using Reverse Image Search. In *Proc. of the 2018 CHI Conf. on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI '18). Association for Computing Machinery, New York, NY, USA, 1–11. <https://doi.org/10.1145/3173574.3174092>

- [20] Kunal Gupta, Ken Takaki, Ayaka Tsutsui, Kentaro Teramoto, and Bektur Ryskeldiev. 2025. AccessibleShip: Exploring Packaging and Shipping Barriers for Visually Impaired Users in C2C E-Commerce. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 656, 8 pages. <https://doi.org/10.1145/3706599.3706665>
- [21] Simon Harper and Sean Bechhofer. 2007. Sadie: Structural semantics for accessibility and device independence. *ACM Transactions on Computer-Human Interaction (TOCHI)* 14, 2 (2007), 10–es.
- [22] Gerald Häubl and Valerie Trifts. 2000. Consumer decision making in online shopping environments: The effects of interactive decision aids. *Marketing science* 19, 1 (2000), 4–21.
- [23] Mina Huh, Yi-Hao Peng, and Amy Pavel. 2023. GenAssist: Making image generation accessible. In *Proc. of the 36th Annual ACM Symp. on User Interface Software and Technology*. 1–17.
- [24] Muhammad Asif Islam, Yevgen Borodin, and I. V. Ramakrishnan. 2010. Mixture model based label association techniques for web accessibility. In *Proc. of the 23rd Annual ACM Symp. on User Interface Software and Technology (New York, New York, USA) (UIST '10)*. Association for Computing Machinery, New York, NY, USA, 67–76. <https://doi.org/10.1145/1866029.1866041>
- [25] Md Touhidul Islam and Syed Masum Billah. 2023. SpaceX mag: an automatic, scalable, and rapid space compactor for optimizing smartphone app interfaces for low-vision users. *Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 2 (2023), 1–36.
- [26] Shuli Jones, Isabella Pedraza Pineros, Daniel Hajas, Jonathan Zong, and Arvind Satyanarayan. 2024. "Customization is Key": Reconfigurable Textual Tokens for Accessible Data Visualizations. , Article 47 (2024), 14 pages. <https://doi.org/10.1145/3613904.3641970>
- [27] Shaun K. Kane, Jacob O. Wobbrock, and Richard E. Ladner. 2011. Usable gestures for blind people: understanding preference and performance. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems (Vancouver, BC, Canada) (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 413–422. <https://doi.org/10.1145/1978942.1979001>
- [28] Smirity Kaushik, Natā M Barbosa, Yaman Yu, Tanusree Sharma, Zachary Kilhoffer, JooYoung Seo, Sauvik Das, and Yang Wang. 2023. {GuardLens}: Supporting Safer Online Browsing for People with Visual Impairments. In *Nineteenth Symp. on Usable Privacy and Security (SOUPS 2023)*. 361–380.
- [29] Rushil Khurana, Duncan McIsaac, Elliot Lockerman, and Jennifer Mankoff. 2018. Nonvisual Interaction Techniques at the Keyboard Surface. In *Proc. of the 2018 CHI Conf. on Human Factors in Computing Systems (Montreal QC, Canada) (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3173574.3173585>
- [30] Andrew Kirkpatrick, Richard Rutter, Christian Heilmann, Jim Thatcher, and Cynthia Waddell. 2006. *Web Accessibility: Web Standards and Regulatory Compliance*. friends of ED.
- [31] Jie Lei, Liwei Wang, Yelong Shen, Dong Yu, Tamara L Berg, and Mohit Bansal. 2020. Mart: Memory-augmented recurrent transformer for coherent video paragraph captioning. *arXiv preprint arXiv:2005.05402* (2020).
- [32] Stefan Leuthold, Javier A Vargas-Avila, and Klaus Opwis. 2008. Beyond web content accessibility guidelines: Design of enhanced text user interfaces for blind internet users. *Int. Journal of Human-Computer Studies* 66, 4 (2008), 257–270.
- [33] Jie Lu, Anjin Liu, Fan Dong, Feng Gu, João Gama, and Guangquan Zhang. 2019. Learning under Concept Drift: A Review. *IEEE Transactions on Knowledge and Data Engineering* 31, 12 (2019), 2346–2363. <https://doi.org/10.1109/TKDE.2018.2876857>
- [34] Tlamele Makati, Garreth W Tigwell, and Kristen Shinohara. 2024. The Promise and Pitfalls of Web Accessibility Overlays for Blind and Low Vision Users. In *Proc. of the 26th Int. ACM SIGACCESS Conf. on Computers and Accessibility*. 1–12.
- [35] Microsoft. 2023. List of Evaluation Metrics for Generative AI. https://learn.microsoft.com/en-us/ai/playbook/technology-guidance/generative-ai/working-with-llms/evaluation/list-of-eval-metrics?utm_source=chatgpt.com. Accessed: 2025-04-15.
- [36] Toshiya Oogane and Chieko Asakawa. 1998. An interactive method for accessing tables in HTML. In *Proc. of the Third Int. ACM Conf. on Assistive Technologies (Marina del Rey, California, USA) (Assets '98)*. Association for Computing Machinery, New York, NY, USA, 126–128. <https://doi.org/10.1145/274497.274521>
- [37] OpenAI. 2023. ChatGPT. <https://chat.openai.com>. Accessed: 2025-07-31.
- [38] OpenAI. 2024. Be My Eyes uses GPT-4 to transform visual accessibility. <https://openai.com/index/be-my-eyes/>. Accessed: 2025-04-15.
- [39] OpenAI. 2024. GPT-4o. Website. Retrieved April 15 2025 from <https://platform.openai.com/docs/models/gpt-4o>.
- [40] Achraf Othman, Amira Dhoubi, and Aljazi Nasser Al Jabor. 2023. Fostering websites accessibility: A case study on the use of the Large Language Models ChatGPT for automatic remediation. In *Proc. of the 16th Int. Conf. on Pervasive Technologies Related to Assistive Environments (Corfu, Greece) (PE-TRA '23)*. Association for Computing Machinery, New York, NY, USA, 707–713. <https://doi.org/10.1145/3594806.3596542>
- [41] Christopher Power, André Freire, Helen Petrie, and David Swallow. 2012. Guidelines are only half of the story: accessibility problems encountered by blind users on the web. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems (Austin, Texas, USA) (CHI '12)*. Association for Computing Machinery, New York, NY, USA, 433–442. <https://doi.org/10.1145/2207676.2207736>
- [42] Bektur Ryskeldiev, Kotaro Hara, Mariko Kobayashi, and Koki Kusano. 2022. Investigating Accessibility Challenges and Opportunities for Users with Low Vision Disabilities in Customer-to-Customer (C2C) Marketplaces. In *Proc. of the 24th Int. ACM SIGACCESS Conf. on Computers and Accessibility (Athens, Greece) (ASSETS '22)*. Association for Computing Machinery, New York, NY, USA, Article 79, 4 pages. <https://doi.org/10.1145/3517428.3550390>
- [43] Bektur Ryskeldiev, Mariko Kobayashi, Kentaro Teramoto, and Koki Kusano. 2023. Towards Immersive Inclusivity for C2C: How Immersive Multimodal Interactions Can Make Online Customer-to-Customer Shopping More Inclusive. In *Proc. of the Asian HCI Symp. 2023 (Online, Indonesia) (Asian CHI '23)*. Association for Computing Machinery, New York, NY, USA, 84–88. <https://doi.org/10.1145/3604571.3604586>
- [44] Abigale Stangl, Meredith Ringel Morris, and Danna Gurari. 2020. "Person, Shoes, Tree. Is the Person Naked?" What People with Vision Impairments Want in Image Descriptions. In *Proc. of the 2020 CHI Conf. on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376404>
- [45] Abigale Stangl, Nitin Verma, Kenneth R Fleischmann, Meredith Ringel Morris, and Danna Gurari. 2021. Going beyond one-size-fits-all image descriptions to satisfy the information wants of people who are blind or have low vision. In *Proc. of the 23rd Int. ACM SIGACCESS Int. on computers and accessibility*. 1–15.
- [46] Abigale J. Stangl, Esha Kothari, Suyog D. Jain, Tom Yeh, Kristen Grauman, and Danna Gurari. 2018. BrowseWithMe: An Online Clothes Shopping Assistant for People with Visual Impairments. In *Proc. of the 20th Int. ACM SIGACCESS Conf. on Computers and Accessibility (Galway, Ireland) (ASSETS '18)*. Association for Computing Machinery, New York, NY, USA, 107–118. <https://doi.org/10.1145/3234695.3236337>
- [47] Sarit Szpiro, Yuhang Zhao, and Shiri Azenkot. 2016. Finding a store, searching for a product: a study of daily challenges of low vision people. In *Proc. of the 2016 ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing (Heidelberg, Germany) (UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 61–72. <https://doi.org/10.1145/2971648.2971723>
- [48] Mahan Tafreshipour, Anmol Deshpande, Forough Mehralian, Iftekhar Ahmed, and Sam Malek. 2024. Ma11y: A Mutation Framework for Web Accessibility Testing. In *Proc. of the 33rd ACM SIGSOFT Int. Symp. on Software Testing and Analysis (Vienna, Austria) (ISSTA 2024)*. Association for Computing Machinery, New York, NY, USA, 100–111. <https://doi.org/10.1145/3650212.3652113>
- [49] Ayaka Tsutsui, Kengo Tanaka, Yaman Yu, and Bektur Ryskeldiev. 2025. "I Am a Blind Seller!": Picture Taking Assistance for Visually Impaired Individuals for Participation as Sellers in Customer to Customer (C2C) Marketplaces. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 10, 10 pages. <https://doi.org/10.1145/3706599.3720150>
- [50] Tess Van Daele, Akhil Iyer, Yuning Zhang, Jalyn C Derry, Mina Huh, and Amy Pavel. 2024. Making Short-Form Videos Accessible with Hierarchical Video Summaries. In *Proc. of the CHI Conf. on Human Factors in Computing Systems*. 1–17.
- [51] Markel Vigo, Justin Brown, and Vivienne Conway. 2013. Benchmarking web accessibility evaluation tools: measuring the harm of sole reliance on automated tests. In *Proc. of the 10th Int. Cross-Disciplinary Conf. on Web Accessibility (Rio de Janeiro, Brazil) (W4A '13)*. Association for Computing Machinery, New York, NY, USA, Article 1, 10 pages. <https://doi.org/10.1145/2461121.2461124>
- [52] Alexandra Vtyurina, Adam Fourney, Meredith Ringel Morris, Leah Findlater, and Ryen W White. 2019. Verse: Bridging screen readers and voice assistants for enhanced eyes-free web search. In *Proc. of the 21st Int. ACM SIGACCESS Conf. on Computers and Accessibility*. 414–426.
- [53] World Wide Web Consortium (W3C). 2018. Web Content Accessibility Guidelines (WCAG) 2.1. <https://www.w3.org/TR/WCAG21/>. Accessed: 2025-04-15.
- [54] Lucy Lu Wang, Isabel Cachola, Jonathan Bragg, Evie Yu-Yen Cheng, Chelsea Haupt, Matt Latzke, Bailey Kuehl, Madeleine van Zuylen, Linda Wagner, and Daniel S Weld. 2021. Improving the accessibility of scientific documents: Current state, user needs, and a system solution to enhance scientific PDF accessibility for blind and low vision users. *arXiv preprint arXiv:2105.00076* (2021).
- [55] Lucy Lu Wang, Isabel Cachola, Jonathan Bragg, Evie Yu-Yen Cheng, Chelsea Haupt, Matt Latzke, Bailey Kuehl, Madeleine N van Zuylen, Linda Wagner, and Daniel Weld. 2021. SciA11y: Converting Scientific Papers to Accessible HTML. In *Proc. of the 23rd Int. ACM SIGACCESS Conf. on Computers and Accessibility (Virtual Event, USA) (ASSETS '21)*. Association for Computing Machinery, New York, NY, USA, Article 85, 4 pages. <https://doi.org/10.1145/3441852.3476545>
- [56] Ruolin Wang, Zixuan Chen, Mingrui Ray Zhang, Zhaocheng Li, Zhixiu Liu, Zihan Dang, Chun Yu, and Xiang 'Anthony' Chen. 2021. Revamp: Enhancing Accessible Information Seeking Experience of Online Shopping for Blind or Low Vision Users. In *Proc. of the 2021 CHI Conf. on Human Factors in Computing Systems*

- (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 494, 14 pages. <https://doi.org/10.1145/3411764.3445547>
- [57] Yujia Wang, Wei Liang, Haikun Huang, Yongqi Zhang, Dingzeyu Li, and Lap-Fai Yu. 2021. Toward Automatic Audio Description Generation for Accessible Videos. In *Proc. of the 2021 CHI Conf. on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 277, 12 pages. <https://doi.org/10.1145/3411764.3445347>
- [58] Yanan Wang, Ruobin Wang, Crescentia Jung, and Yea-Seul Kim. 2022. What makes web data tables accessible? Insights and a tool for rendering accessible tables for people with visual impairments. In *Proc. of the 2022 CHI Conf. on Human Factors in Computing Systems*. 1–20.
- [59] WebAIM. 2021. WebAIM Screen Reader User Survey #10. <https://webaim.org/projects/screenreadersurvey10/>. Accessed: 2025-04-15.
- [60] WebAIM. 2023. Screen Reader User Survey. <https://webaim.org/projects/screenreadersurvey/>. Accessed: 2025-04-15.
- [61] Michele A. Williams, Amy Hurst, and Shaun K. Kane. 2013. "Pray before you step out": describing personal and situational blind navigation behaviors. In *Proc. of the 15th Int. ACM SIGACCESS Conf. on Computers and Accessibility* (Bellevue, Washington) (ASSETS '13). Association for Computing Machinery, New York, NY, USA, Article 28, 8 pages. <https://doi.org/10.1145/2513383.2513449>
- [62] Yaman Yu, Saidivya Ashok, Smirity Kaushik, Yang Wang, and Gang Wang. 2023. Design and evaluation of inclusive email security indicators for people with visual impairments. In *2023 IEEE Symp. on Security and Privacy (SP)*. IEEE, 2885–2902.
- [63] Chien Wen Yuan, Benjamin V. Hanrahan, Sooyeon Lee, Mary Beth Rosson, and John M. Carroll. 2017. I Didn't Know that You Knew I Knew: Collaborative Shopping Practices between People with Visual Impairment and People with Vision. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 118 (Dec. 2017), 18 pages. <https://doi.org/10.1145/3134753>
- [64] Chien Wen Yuan, Benjamin V. Hanrahan, Sooyeon Lee, Mary Beth Rosson, and John M. Carroll. 2019. Constructing a holistic view of shopping with people with visual impairment: a participatory design approach. *Universal Access in the Information Society* 18 (2019), 127–140.
- [65] Beste F. Yuksel, Pooyan Fazli, Umang Mathur, Vaishali Bisht, Soo Jung Kim, Joshua Junhee Lee, Seung Jung Jin, Yue-Ting Siu, Joshua A. Miele, and Ilmi Yoon. 2020. Human-in-the-Loop Machine Learning to Increase Video Accessibility for Visually Impaired and Blind Users. In *Proc. of the 2020 ACM Designing Interactive Systems Conf.* (Eindhoven, Netherlands) (DIS '20). Association for Computing Machinery, New York, NY, USA, 47–60. <https://doi.org/10.1145/3357236.3395433>
- [66] Beste F. Yuksel, Soo Jung Kim, Seung Jung Jin, Joshua Junhee Lee, Pooyan Fazli, Umang Mathur, Vaishali Bisht, Ilmi Yoon, Yue-Ting Siu, and Joshua A. Miele. 2020. Increasing Video Accessibility for Visually Impaired Users with Human-in-the-Loop Machine Learning. In *Extended Abstracts of the 2020 CHI Conf. on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI EA '20). Association for Computing Machinery, New York, NY, USA, 1–9. <https://doi.org/10.1145/3334480.3382821>
- [67] Yu Zhong, Walter S. Lasecki, Erin Brady, and Jeffrey P. Bigham. 2015. RegionSpeak: Quick Comprehensive Spatial Descriptions of Complex Images for Blind Users. In *Proc. of the 33rd Annual ACM Conf. on Human Factors in Computing Systems* (Seoul, Republic of Korea) (CHI '15). Association for Computing Machinery, New York, NY, USA, 2353–2362. <https://doi.org/10.1145/2702123.2702437>
- [68] Peter A. Zientara, Sooyeon Lee, Gus H. Smith, Rorry Brenner, Laurent Itti, Mary B. Rosson, John M. Carroll, Kevin M. Irick, and Vijaykrishnan Narayanan. 2017. Third Eye: A Shopping Assistant for the Visually Impaired. *Computer* 50, 2 (2017), 16–24. <https://doi.org/10.1109/MC.2017.36>

A Appendix

A Implementation Details

A.1 System Configuration and LLM Setup

The Chrome extension's background script uses the OpenAI API to access GPT-4o [39], implemented through the browser's fetch Web API. We selected GPT-4o for both regeneration (Option 1) and tag reorganization (Option 2) due to its large input (128,000 tokens) and output (16,384 tokens) capacities [39], which allow extensive content processing while maintaining structure and context.

To optimize for deterministic and consistent outputs, the following parameters were used: `temperature = 0.2`, `max_tokens = 16,384`, `top_p = 1`, `frequency_penalty = 0`, and `presence_penalty = 0`. A low temperature helps maintain formatting consistency in HTML, while `top-p = 1` ensures complete content generation. The penalties are set to 0 to avoid suppressing necessary recurring elements, such as repeated headings or button labels.

Each webpage required 1–5 minutes to process, with around 220,000 tokens consumed per session (combined input and output). According to OpenAI's pricing⁷, the cost per page ranged from \$0.50 to \$2.20. This could be reduced in production by caching outputs or using lighter models.

A.2 Option 1: Regenerated HTML Version

A.2.1 Step 1: HTML Collection. We used a Google Chrome extension to extract complete webpage HTML. A **content script** interacts directly with the Document Object Model (DOM) and retrieves the full HTML using `document.documentElement.outerHTML`. The captured HTML, including dynamic JavaScript-rendered elements, is passed to the **background script**.

A.2.2 Step 2: Preprocessing. The background script segments the raw HTML based on DOM structure to ensure each chunk fits within the token limit of GPT-4o. Segmentation is done at logical HTML boundaries (e.g., top-level `<div>` or `<section>` blocks) to preserve context. Tokenization ensures each chunk respects model constraints.

A.2.3 Step 3: Sequential LLM Calls. Each HTML chunk is processed using OpenAI's chat/completions API. A system prompt, user prompt, and assistant context (see Section A.4) guide the model to generate an accessible HTML version. The assistant message includes prior output to maintain continuity across chunks.

A.2.4 Step 4: Similarity Check. Before injecting the generated output into the live webpage, our system conducts a similarity check to ensure that the transformed HTML retains the core content and meaning of the original page. To maintain consistency between implementation and evaluation, we used the same metric—**Aggregated Semantic Similarity**—as described in Section 6.2.

We compute semantic similarity by extracting all screen-reader-accessible HTML elements (e.g., visible text, ARIA labels, alt attributes) from both the original and generated versions, concatenating them, and comparing their embeddings using cosine similarity. The transformation is accepted only if the aggregated semantic

⁷<https://openai.com/api/pricing>

similarity score meets or exceeds a **90% threshold**. This threshold was chosen based on preliminary testing to ensure the balance between allowing accessibility-enhancing changes (e.g., restructuring or paraphrasing) and preserving the original meaning. If the score falls below the threshold, the system automatically reprocesses the content to improve alignment.

A.2.5 Step 5: Page Replacement. Once the final HTML passes the checks, it replaces the original webpage DOM using JavaScript injection. All functional elements (e.g., forms, buttons, links) remain intact. If any issues are detected post-rendering, additional regeneration is triggered until the results are usable and accessible.

A.3 Option 2: Reorganized HTML Tags

A.3.1 Step 1: DOM Extraction and JSON Serialization. Option 2 focuses on minimally modifying the webpage structure while improving accessibility. HTML is extracted similarly via the extension, but converted into a structured JSON format for fine-grained tag-level manipulation. This includes tag names, attributes, and content.

A.3.2 Step 2: Preprocessing. We filter out unnecessary elements (e.g., styles, scripts) and chunk the JSON for compatibility with the GPT-4o model’s 32,768-token input window. Segments are created at logical DOM subtrees to preserve structural coherence.

A.3.3 Step 3: LLM-Powered Tag Optimization. Each chunk of JSON is sent via the same Web API using system and user prompts (Section A.4) tailored to instruct the model to adjust headings, labels, and ARIA attributes for improved screen reader navigation—without introducing or removing any content.

A.3.4 Step 4: Similarity Check. Although Option 2 only modifies the HTML tags and not the visible content or layout, we still perform a similarity check to ensure that the tag reorganization does not unintentionally alter or omit content relevant to screen reader users. We apply the same **Aggregated Semantic Similarity** metric used in Option 1 and in our evaluation (Section 6.2) to compare the original and tag-updated versions.

We extract all accessible HTML elements (including headings, labels, and text content) from both versions, concatenate them, and compute cosine similarity between their semantic embeddings. The updated version is accepted only if the aggregated semantic similarity score is at least **90%**. This threshold ensures that any tag changes preserve the original content’s meaning and utility for screen reader users. If the similarity score is below the threshold, the chunk is flagged for reprocessing.

A.3.5 Step 5: Live DOM Update. Once validated, the extension updates the active page using a `replaceTags()` function, replacing only the tags while keeping the text and visual layout unchanged. All interactive elements remain fully functional, now better marked up for screen reader interpretation.

A.4 Prompt Design

In both options, LLM prompts are composed of a **System Prompt** to establish behavior, a **User Prompt** to pass specific HTML or JSON input, and an **Assistant Message** to maintain continuity across chunks.

A.4.1 Prompts for Option 1 (Regeneration). **System Prompt:**

Your task is to enhance the accessibility of the shopping website for blind screen reader users. Screen reader users navigate websites sequentially and use shortcuts to jump across headings and links. Remove non-essential elements (e.g., pictures, `<style>`, `<script>`) and ensure all content from the original HTML is included and properly structured.

User Prompt:

Generate executable, text-only HTML optimized for screen reader users. Reorganize headings to reflect information hierarchy and enhance clarity. Prioritize content that is most relevant for screen reader navigation.

Assistant Prompt:

This is the previous response from the system: `<HTML content>`

Follow-up User Prompt:

This is part X of the HTML document. Please continue processing it while maintaining the previous structure. `<next HTML segment>`

A.4.2 Prompts for Option 2 (Reorganization). **System Prompt:**

You are an accessibility expert. Screen reader users navigate content sequentially, relying on well-structured headings, landmarks, and clear labels. Your task is to revise the tags of an existing webpage to improve accessibility without modifying its visual layout or content.

User Prompt:

Please adjust tags to enhance accessibility for screen reader users. Do not introduce or remove elements. Only reorganize or relabel existing tags based on structural clarity and semantic accuracy.

Assistant Prompt:

This is the previous response from the system: `<JSON content>`

Final Message (if needed):

This is the remaining JSON content that needs to be processed. Ensure completeness and do not duplicate any previously generated elements.

A.5 Example Accessibility Enhancements

To demonstrate the types of accessibility enhancements applied by our system, we summarize typical modifications made under each version.

A.5.1 Option 1: Regenerated HTML.

- **Consistent and Semantic Heading Structure:** The regenerated HTML (Table 6, center) ensures clear and consistent use of heading tags (`<h1>`, `<h2>`, `<h3>`, etc.) throughout the pages, following a logical order. This makes it easier for screen reader users to understand the page layout and navigate to relevant sections. For example, on the homepage, the `<h1>` tag is appropriately used for the main title, “Shop Popular

Table 6: Extended example of HTML changes for an e-commerce search-results page.

Original HTML	Option 1 (Regenerated HTML)	Option 2 (Reorganized HTML Tags)
<pre>1-16 of over 10,000 results for "blender" <!-- ... --> <h2>nutribullet SmartSense Blender ... Black</h2> <div>Rating: 4.668</div></pre>	<pre><header> <h1>Search Results for "Blender"</h1> <p>1-16 of over 10,000 results</p> </header> <main> <section> <h2>Product Listings</h2> <article> <h3>nutribullet SmartSense Blender ... Black</h3> <p>Rating: 4.668</p> </article> <!-- ... --> </section> </main></pre>	<pre><h1>1-16 of over 10,000 results for "blender"</h1> <!-- ... --> <h2>nutribullet SmartSense Blender ... Black</h2> <p>Rating: 4.668</p> <!-- ... --></pre>

Items,” while sub-sections such as “Featured Categories” and “Today’s Picks” use <h2>. This prevents the confusion caused by random or missing headings in the original version.

- Replace headings to distinguish and highlight information: The content is logically organized with unnecessary elements removed from headings, and summarizing titles added to enhance clarity and reduce clutter. This makes it easier for users to focus on the relevant information without distractions. For example, on the search result page, instead of showing all filter options and categories as separate headings, the LLM-regenerated version groups them under a single “Filter Results” section, allowing users to navigate efficiently.
- Reordered Sections for Logical Screen Reader Browsing Sequence: The order of sections on each page is rearranged to provide a more logical flow for screen reader users. This adjustment ensures that the most important information is encountered first, reducing the need for excessive navigation. For example, on the product details page, the regenerated HTML places “Product Name” and “Product Details” before sections such as “Reviews” and “Similar Products.” This contrasts with the original layout where reviews could appear before key product information. This change aligns better with the linear browsing patterns of screen reader users, who prefer accessing critical information first.
- Enhanced Use of ARIA Roles and Attributes: ARIA roles (e.g., role=“navigation”, role = “main”, role = “banner”) and attributes (e.g., aria-label, aria-labelledby) are added to provide additional context for assistive technologies. This allows screen readers to better interpret the purpose of various sections and elements. For exmaple, on the product details page, buttons such as “Add to Cart” are given specific aria-label attributes (aria-label = “Add item to cart”), enhancing clarity for screen reader users who cannot see the button’s text.

A.5.2 Option 2: Reorganized HTML Tags Version .

- Replace headings to distinguish and highlight information: this version focuses (Table 6, right) on ensuring that headings are used appropriately to reflect the content hierarchy, removing unnecessary elements from being tagged as headings

and adding key information into headings. Uses appropriate headings (<h1>, <h2>, <h3>, etc.) and paragraphs (<p>) to present content hierarchically and semantically. For example, the product name “oster blender” is an <h1>, ensuring it stands out as the primary focus.

- Enhanced Heading Hierarchy for Logical Screen Reader Navigation: This version focuses on adjusting the levels of headings to emphasize the content’s structure and hierarchy without altering the page layout. By refining the heading levels, it provides a clearer and more logical flow for screen readers. For example, on the search result page, the filter sections such as “Filter by,” “Status,” “Item origin,” and “Size” are clearly marked with <h2>, <h3>, and <h4> tags, providing a hierarchical structure that is more accessible for screen readers and improves the overall user experience.