Adapting Online Customer Reviews for Blind Users: A Case Study of Restaurant Reviews

Mohan Sunkara
Old Dominion University
Department of Computer Science
Norfolk, Virginia, USA
msunk001@odu.edu

Akshay Kolgar Nayak
Old Dominion University
Department of Computer Science
Norfolk, Virginia, USA
anaya001@odu.edu

Sandeep Kalari
Old Dominion University
Department of Computer Science
Norfolk, Virginia, USA
skala003@odu.edu

Yash Prakash
Old Dominion University
Department of Computer Science
Norfolk, Virginia, USA
yprak001@odu.edu

Sampath Jayarathna
Old Dominion University
Department of Computer Science
Norfolk, Virginia, USA
sampath@cs.odu.edu

Hae-Na Lee
Michigan State University
Department of Computer Science and Engineering
East Lansing, Michigan, USA
leehaena@msu.edu

Vikas Ashok
Old Dominion University
Department of Computer Science
Norfolk, Virginia, USA
vganjigu@odu.edu

ABSTRACT

Online reviews have become an integral aspect of consumer decision-making on e-commerce websites, especially in the restaurant industry. Unlike sighted users who can visually skim through the reviews, perusing reviews remains challenging for blind users, who rely on screen reader assistive technology that supports predominantly one-dimensional narration of content via keyboard shortcuts. In an interview study, we uncovered numerous pain points of blind screen reader users with online restaurant reviews, notably, the listening fatigue and frustration after going through only the first few reviews. To address these issues, we developed QuickCue assistive tool that performs aspect-focused sentiment-driven summarization to reorganize the information in the reviews into an alternative, thematically-organized presentation that is conveniently perusable with a screen reader. At its core, QuickCue utilizes a large language model to perform aspect-based joint classification for grouping reviews, followed by focused summarizations within the groups to generate concise representations of reviewers' opinions, which are then presented to the screen reader users via an accessible interface. Evaluation of QuickCue in a user study with 10 participants showed significant improvements in overall usability and task workload compared to the status quo screen reader.

W4A '25, April 28-29, 2025, Sydney, Australia

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/XXXXXXX.XXXXXXX

CCS CONCEPTS

• **Human-centered computing → Accessibility technologies**; *Empirical studies in accessibility*.

KEYWORDS

blind, screen reader, visual impairment, assistive technology, online discussion forum, large language model

ACM Reference Format:
Mohan Sunkara, Akshay Kolgar Nayak, Sandeep Kalari, Yash Prakash, Sampath Jayarathna, Hae-Na Lee, and Vikas Ashok. 2025. Adapting Online Customer Reviews for Blind Users: A Case Study of Restaurant Reviews. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (W4A '25)*. ACM, New York, NY, USA, 12 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Online reviews have become a cornerstone of modern consumer decision-making, offering valuable insights into products, services, and experiences [44, 69, 91]. This has especially been the case in the restaurant industry, with reviews and ratings providing diners with information about food quality, ambiance, and service, thereby helping them make informed choices [33, 59]. Therefore, the presentation of information in user reviews must be as holistic and fair as possible, to avoid inducing consumer biases and harming a restaurant's reputation. While present applications (e.g., Google Maps) do include an assortment of features in their interfaces to help prospective diners make fully-informed decisions, these features are presently insufficient for blind users who rely on screen reader assistive technology to interact with these applications.

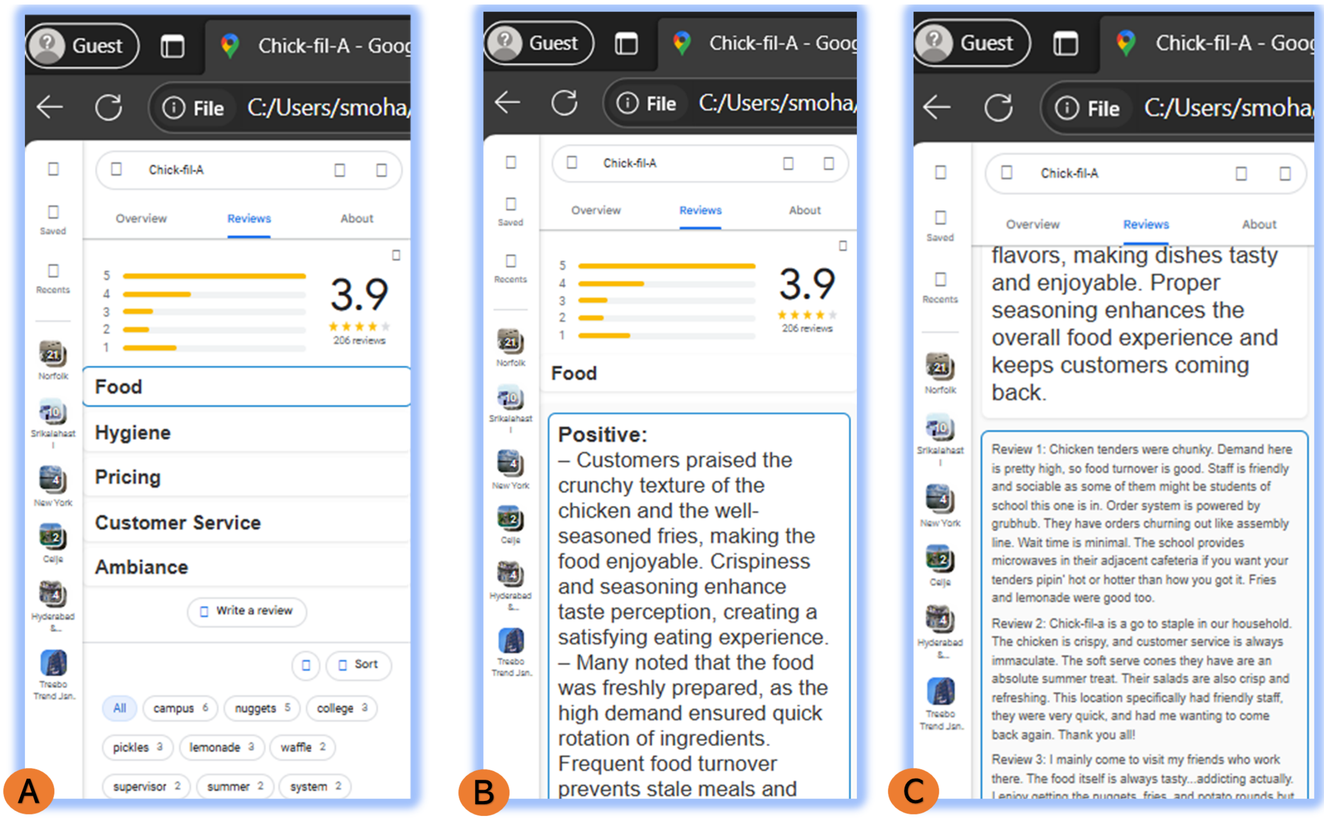

Figure 1: QuickCue's interface hierarchy: (A) displays the default list of five aspects, (B) shows the positives and negatives of each aspect, and (C) shows the original reviews associated with each aspect-sentiment pair.

A screen reader (e.g., JAWS [38], NVDA [63], VoiceOver [7]) narrates web-application content based on the order in the webpage's document object model (DOM), essentially enforcing a one-dimensional interaction with the content. Although a screen reader offers numerous keyboard shortcuts to aid navigation, the "press-and-listen" paradigm inherently limits the efficiency and usability of accessing and consuming information, including user reviews, in websites [50, 67]. In contrast, sighted users can effortlessly scan and skim online reviews, by leveraging visual cues to quickly pinpoint relevant content. In an interview study with 30 blind screen reader users, we found that participants often experienced listening fatigue after perusing only a few reviews, they faced difficulties in finding specific information, and in general they struggled with navigating unstructured and repetitive reviews. Almost all participants expressed a need for an alternative thematic presentation of reviews (i.e., grouping according to food quality, ambiance, service, etc.), with further bifurcation of information along the lines of 'positives' and 'negatives' (e.g., good/bad experiences about service, food, or ambiance).

To address the aforementioned interaction challenges as well as adapt the presentation of information to match users' needs and preferences, we developed QuickCue, a novel assistive tool embodied as a browser extension, that augments the existing interface (see Figure 1) with additional content comprising aspect-based organization of information mined from reviews. To generate such as an alternative presentation, QuickCue performs the following two core tasks: (i) Joint classification of reviews to determine the aspect-sentiment pairs covered in the reviews; followed by (ii) Focused summarization of select reviews associated with each aspect-sentiment pair. QuickCue performs both these tasks by leveraging the GPT-4 large large model (LLM), specifically using the *clue and reasoning prompt engineering* strategy [76] for joint classification, and *directed stimulus prompt engineering* strategy [55] for focused summarization of select reviews. The user interface, as shown in Figure 1, hierarchically presents the generated information, with only aspects listed in the first level (Figure 1(A)), to the positive and negative summaries in the second level (Figure 1(B)), to the subset of reviews associated with the aspect-sentiment pair in the last level (Figure 1(C)).

A user study with 10 blind participants showed that QuickCue significantly improved usability and reduced interaction workload for blind screen reader users while perusing user reviews, compared to the status quo. Furthermore, a majority of the participants stated that QuickCue would enable them to make more informed decisions regarding choice of restaurants. In sum, this paper makes the following contributions:

- Interview study with 30 blind users uncovering their interaction challenges and needs regarding online reviews.

- The design, development, and evaluation of QuickCue, a novel assistive tool for blind users to efficiently consume customer reviews.

2 RELATED WORK

2.1 Online Review Systems

Extensive research has explored the impact of online reviews on consumer decision making [26, 56, 72, 73]. Liu et al. [56] found that online reviews allow consumers to make more informed choices, acting as digital word-of-mouth. The impact of reviews extends to sales performance, as even slight changes in ratings can lead to substantial variations in revenue [58]. Credibility is a key factor in the effectiveness of online reviews, with detailed, balanced, and authentic feedback being perceived as more trustworthy [36].

In the restaurant domain, consumer reviews typically include detailed feedback on food quality, service, ambiance, and price, all of which are key factors influencing a prospective diner's restaurant selection. Campos et al. [72] applied natural language processing techniques to analyze sentiments in online reviews, showing how positive or negative sentiments directly impact restaurant reputation and customer expectations. Dash et al. [26] further extended this work by demonstrating the effectiveness of deep learning models in extracting relevant attributes from online reviews to recommend dishes, highlight popular items, and even tailor experiences based on individual customer preferences. The above research works underline the critical role of online reviews in shaping consumer perceptions and behaviors. However, these works do not explore users' interaction and engagement with the content in reviews. While online review systems play a crucial role in shaping consumer preferences, blind screen reader users are presently unable to fully exploit these systems, as the present user interfaces are mostly designed for visual interaction. We address this issue in this paper by building QuickCue that dynamically augments the current review systems with a screen-reader friendly interface to conveniently peruse information in reviews.

2.2 Web Interaction with Screen Readers

As mentioned earlier, blind people interact with digital applications, including web applications, using screen-reader assistive technology such as JAWS, NVDA, or VoiceOver. A screen reader transforms the two-dimensional graphical interface of a web page into a linear, one-dimensional list of on-screen elements (such as headers, text, buttons, and menus) for auditory navigation. This sequential press-and-listen method of navigation has been found to create significant accessibility and usability challenges for blind users [3, 10, 11, 15, 40, 61, 64, 92], despite the availability of several accessibility guidelines [21, 24, 25] and accessibility checking-aids for web developers [1, 2, 18, 49].

While the accessibility challenges have been extensively investigated in prior works [48, 57, 88], relatively fewer studies have focused on the usability of web interaction for screen reader users [12, 34, 52]. Usability, i.e., the ease, efficiency and satisfaction with which users can accomplish tasks, is equally important in web interaction for blind users, with many studies showing that screen reader users are typically an order of magnitude slower than sighted peers in doing the same web tasks [30, 60]. While usability-enhancing solutions for blind users have been proposed in the literature [9, 19, 32, 34, 35, 47, 51–53, 67, 68, 77], these have predominantly focused on the general efficiency of webpage navigation, and as such they are inadequate in their ability to address domain-specific challenges involved in online review systems. Review systems are not only text-heavy with significant information redundancy, but often require nuanced understanding of tone, sentiment, and other specific preferences or features that traditional screen readers and other extant usability solutions are currently unable to support for assisting blind users. A tailored solution is therefore needed to address this issue and make online review systems more usable for blind screen reader users, which is the focus of this paper.

2.3 User Interfaces of Review Systems

The effective organization of information plays a crucial role in enhancing user experience, especially in text-heavy online review systems. Therefore, prior works have looked into methods such as clustering, sentiment-based categorization, and personalized filtering to structure data in reviews into more digestible formats [45, 46, 89]. For instance, faceted navigation has been widely applied in e-commerce platforms to allow users to filter reviews based on specific attributes, such as taste, portion size, and service quality [83]. Similarly, researchers have explored sentiment-based organization of restaurant reviews, finding that customers tend to prioritize attributes such as food quality and pricing when assessing menu items [8, 93]. However, these works have all focused mostly on sighted-user interaction, and as such do not fully address the unique needs of blind screen reader users.

Another factor to consider while presenting reviews to blind users is the redundancy of information. As screen reader interaction consumes significant time and effort [50, 71], repetitive reviews with little new information can be burdensome for blind users, since they are unable to quickly skim through the reviews like their sighted counterparts. Text summarization [31, 78], therefore, can be a valuable tool to address this issue. Especially, the recent large language models such as GPT-4 [4], Gemini [79], and Llama [81] which have demonstrated remarkable summarization capabilities across diverse domains with either prompting [23, 43, 74] or minimal fine-tuning [90], can be very useful to compact information in reviews across different aspects and granularity before providing them to the blind users. However, ensuring factual accuracy and maintaining relevance in generated summaries requires domain-specific adaptation (e.g., through tailored few-shot prompts) and is currently an active area of research [39, 84]. While some applications, e.g., Google Maps, are leveraging these models to summarize information in reviews, these are mostly 'high-level' short summaries capturing multiple aspects; these applications do not provide specific summaries pertaining to individual aspects, which can be more informative to users.

3 INTERVIEW STUDY

We conducted an IRB-approved interview study with 30 screen reader users to investigate their needs and challenges regarding online customer reviews, particularly in the restaurant domain. The participants were recruited via email lists and word-of-mouth. The average age of the participants was 43.2 years (median: 44, min: 22, max: 63), and the gender distribution was 13 male and 17 female. The inclusion criteria were: (i) Proficiency in web screen reading; (ii) Familiarity with online review systems including restaurant reviews; and (iii) Proficiency in English. The interviews were semi-structured to allow deeper discussions into the participants' needs and issues, with each discussion initiated by a set of careful crafted seed questions. Examples of seed questions included: *What are the primary challenges you encounter while navigating restaurant reviews?*, *What key information do you prioritize when browsing reviews?*, *How do you typically search for the key details in the reviews?*, and *What improvements would make it easier for you to find relevant and helpful information in restaurant reviews?*. The interview feedback was qualitatively analyzed using the standard open coding and axial coding methods [70], where we iteratively examined the transcribed interview data to identify key insights and pain points.

3.1 Findings

Information overload and listening fatigue. A majority (26) of the participants reported struggling with navigating large volumes of reviews, describing it as a significant challenge due to repetitive and redundant content that caused frustration and fatigue. In this regard, one participant P8 stated, "It's frustrating to go through 10 reviews that say the same thing – great food, nice ambiance. I need more details, like whether the restaurant can accommodate dietary restrictions or if the seating is comfortable and accessible." Another participant P4 shared a similar sentiment, "Sometimes I just give up because the information feels repetitive and boring." More than half (17) of the participants expressed that often stopped perusing reviews after listening to only the first few reviews due to listening fatigue. Eleven participants further felt that the present interface was not 'fair' to them in this regard, as they could not obtain sufficient information to make informed decisions. Towards this, one participant P12 mentioned, "I cannot go beyond 4 to 5 comments without getting tired ... sometimes even less if the first couple of reviews are too long ... so I don't get the full picture of what is good and what is bad ... just have to decide based on opinions of a couple of folks, which is obviously unfair."

Outdated information and reviews. All participants mentioned that they often came across reviews that contained outdated information about the different aspects of the restaurants. For examples, P6 said, "I first check the menu and then look at the reviews to see which items are good. But some of the reviews don't make any sense, since they mention dishes that don't exist in the menu ... perhaps the menu has been changed after the review was written, who knows". Another participant P23 echoed, "A lot of things changed after COVID ... many reviews before COVID are no longer useful." When probed regarding their preferred time threshold for reviews, 17 participants mentioned that they wished to only view reviews in the past 1 year, 7 participants indicated two years as their preferred limit, whereas the remaining 6 participants mentioned that they were only interested in reviews from the past six months. In this regard, one participant P17 stated, "Restaurant staff, food quality, and service keep changing all the time. What was good a few years ago may not be good anymore ... some of the bad stuff might have also improved over time."

Thematic organization of information in reviews. Most (28) of the participants expressed a need for an alternative aspect-based (e.g., food quality, hygiene, and ambiance) organization of information in reviews, as they felt that this type of organization helpful because it allowed them to quickly identify the aspects most relevant to their needs. Sixteen participants further suggested summarizing reviews within each aspect to avoid redundancy in the content and also drown out vague uninformative reviews (e.g., *good food!*). As for the preferred aspects, food quality and pricing were unsurprisingly specified as high priority (by 26 and 22 participants respectively). The customer service, hygiene, and ambiance aspects were also mentioned as important by a sizable chunk of the participant pool (19, 16, and 12 participants respectively). For instance, P5 stated, "I attend for the food; however, if the environment is noisy or cramped, it detracts from the overall experience." Another participant P27 reiterated this perspective: "Good service is essential. A rude waiter can make even a tasty meal forgettable."

Sentiment-based insights. Nearly two-thirds (19) of the participants expressed a desire for sentiment-based segregation of information in the reviews. For instance, P16 said, " I simply like to know what is good and what is bad. What is nice about the food... which dishes to avoid.. is the place too crowded on the weekends... are the prices reasonable. If I can easily get this information without having to search for it myself, it will save me a lot of time." The preferences for the 'positives' vs. 'negatives' however varied across the different aspects. While some of the participants were more interested in the positive feedback regarding food quality (e.g., 'What items are most recommended here?' – P14), others were more interested in the negative feedback regarding hygiene (e.g., 'Do they have a hand sanitizer at the entrance?' – P23). Similarly, regarding the pricing, customer service, and ambiance, the participants leaned more towards negative, negative, and positive feedback respectively.

Summary. The participants pointed out several interaction issues with the current presentation of customer reviews, including listening fatigue, frustration after listening to only a few reviews, content redundancy, and difficulty in searching for specific information. To address these issues, most participants suggested an alternative presentation of information in the customer reviews, specifically along the concepts of themes (or aspects) and customer sentiment. With Google Maps as the vehicle for our investigation, we developed QuickCue prototype that generates such an alternative screen reader-friendly thematic presentation of information in restaurant reviews, as described next.

4 QUICKCUE PROTOTYPE DESIGN

4.1 Overview

Figure 2 presents an operational schematic of the QuickCue browser extension we built to thematically organize information in customer reviews and then present this processed information via a screen reader-friendly user interface (Figure 1). For the case study, we chose the Google Maps platform, given that it is most popular platform for sharing reviews online, especially regarding restaurants [27]. As seen in the figure, QuickCue augments the existing Google Maps user interface with additional accessible content in which the information in reviews are organized based on both their underlying aspect (e.g., food quality, pricing) and their sentiment

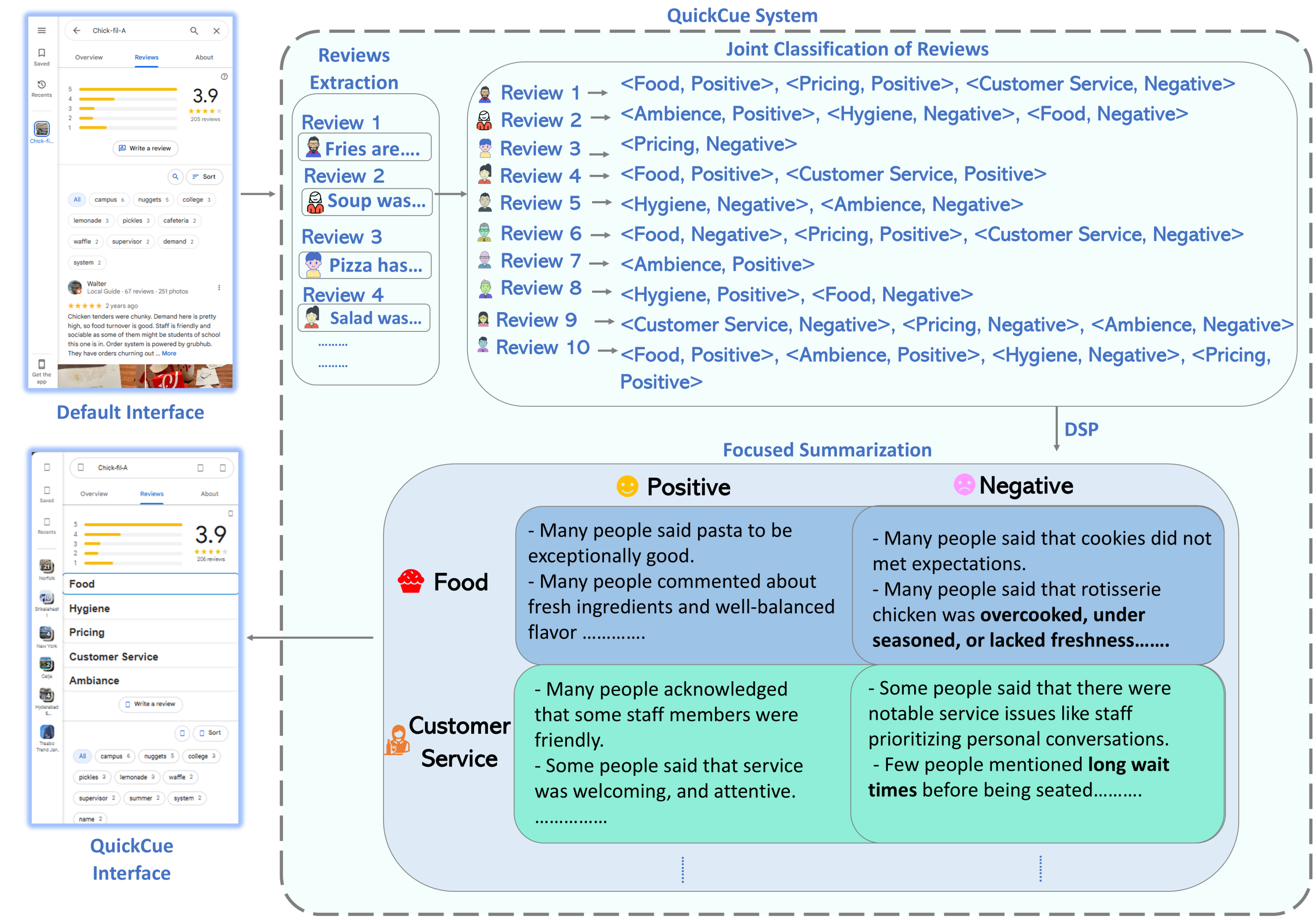

Figure 2: A workflow schematic depicting QuickCue's architecture.

(positive, negative). The first level comprises a list of five drop-down buttons, each corresponding to an aspect. The next level comprises positive and negative summaries pertaining to each of the five aspects. The last level simply lists the raw reviews classified as belonging to each of the aspect/sentiment pairs. This arrangement, designed based on the findings of the interview study, not only reduces information redundancy by enabling users to get a quick overview of the 'good' and the 'bad' of aspects the users care about, but also helps them focus on a subset of reviews pertaining to a specific aspect of interest, e.g., reviews that shed light on the negative experiences with customer service.

To generate such an alternative presentation of information in reviews, QuickCue addresses the following two main technical challenges, which primarily stem from the heterogeneous nature of customer reviews: (i) A review can contain information about multiple aspects; and (ii) A review can contain both positive and negative opinions about different aspects (e.g., *food was good, but the table and seats were not properly cleaned*) and even within the same aspect (e.g., *the staff were very friendly, but the wait was too long!*). To address these challenges, QuickCue performs the following core operations by adapting the state-of-the-art large language models (LLMs): (i) Joint classification of reviews – given a review, determine all the <aspect, sentiment> pairs that are applicable to that review based on its contained information; and (ii) Aspect-focused summarization of information in reviews – given a set of reviews, generate a summary that only focuses on a particular aspect and sentiment. The details of these operations are provided in the next.

4.2 Joint Classification of Reviews

QuickCue performs joint classification of reviews to group them according to the aspects and sentiments covered in their content. Specifically, for each review, QuickCue determines all the <aspect,sentiment> pairs that are applicable to that review. For example, for the review "The food was delicious, but the service was slow.", QuickCue's joint classifier will output "[["Food," "Positive"],["Customer Service," "Negative"]]." As justified earlier, QuickCue primarily looks for five aspects (food, ambiance, customer service, pricing, and hygiene) and two sentiments (positive and negative), while doing the classifications.

To do the joint classification, QuickCue leverages the GPT-4 large language model (LLM) [66], due to its proven ability to reason over complex contexts and generate tailored outputs [41, 80]. To instruct the LLM to accurately classify the reviews, we specifically adapted the Clue and Reasoning Prompting (CARP) strategy [76], given its suitability for this task. The CARP strategy enhances the classification performance by instructing the LLM to look for 'clues' and use that in the reasoning process while determining the class of the input text. The clues may refer to a keyword, phrase, or contextual element extracted from the input text that provides evidence for classification. Since the original CARP prompting [76] was intended for only sentiment classification, we modified it so as to make the LLM generate a set of aspect-sentiment pairs as output instead of a single classification label. The structure of our modified CARP prompt is shown below.

> **Task Description:** This is a joint aspect-sentiment classifier for restaurant reviews.
>
> **First,** present **CLUES** (i.e., keywords, phrases, contextual information, semantic meaning, semantic relations, tones, references) that support the joint aspect-sentiment determination of input (look for clues related to *Food, Ambiance, Customer Service, Pricing, Hygiene* for aspect, and clues related to *positive, negative* for sentiment).
>
> **Second,** deduce a diagnostic **REASONING** process from premises (i.e., clues, input) that supports the sentiment determination for each identified aspect. Note that an aspect can be identified multiple times in different locations of the input.
>
> **Third,** determine the list of aspect-sentiment pairs present in the **INPUT**, considering the **CLUES** and the **REASONING** process.
>
> Output all possible aspect-sentiment pairs after removing empty pairs if any.
> For **ASPECT**, choose from the following predefined set of words: `[Food, Ambiance, Hygiene, Customer Service, Pricing]`.
> For **SENTIMENT**, choose from the following two words: `[Positive,Negative]`
>
> **EXAMPLES:**
>
> **INPUT:** The ambiance was warm and inviting, but the pasta lacked seasoning and was undercooked.
>
> **CLUES:** [ambiance, warm], [ambiance, inviting], [pasta, lacked seasoning], [pasta, undercooked]
>
> **REASONING:** The terms "warm" and "inviting" suggest a welcoming and pleasant atmosphere, indicating a positive experience with ambiance.
> The phrases "lacked seasoning" and "undercooked" indicate dissatisfaction with the food quality, suggesting a negative sentiment for food.
>
> **ASPECT-SENTIMENT Pairs:** ["Ambiance", "Positive"] ["Food", "Negative"]
>
> **INPUT:** [Insert review text here]

As shown in the template, to further enhance the classification performance, we augmented the CARP prompt with few-shot examples. Based on the recommendation of prior work which advocated a minimum of 16 few-shot examples for the CARP prompt [76], we created 20 examples, i.e., 2 examples for each of the 10 aspect-sentiment pairs. We chose a diverse set of reviews for these examples, covering different restaurant cuisines and locations.

After classifying each of the reviews using the joint classifier, QuickCue categorizes the reviews based on the aspect-sentiment pairs, i.e., for each of the 10 aspect-sentiment pairs, QuickCue identifies the corresponding subset of matching reviews, based on the classifier output. Note that, given the heterogeneity of reviews, a review can possibly be included in multiple subsets corresponding to different aspect-sentiment pairs. Each subset then serves as the input for generating *focused* review summaries, as explained later in Section 4.3.

Evaluation. To evaluate the performance of the joint classifier, we created a ground truth dataset by manually annotating 50 reviews or examples which were randomly sampled across diverse restaurant cuisines and locations. The frequency breakdown for the 10 aspect-sentiment pairs in this dataset was as follows: [Food, Negative]: 24, [Food, Positive]: 27, [Customer Service, Negative]: 10, [Customer Service, Positive]: 15, [Pricing, Negative]: 9, [Pricing, Positive]: 8, [Ambiance, Negative]: 5, [Ambiance, Positive]: 12, [Hygiene, Negative]: 6, [Hygiene, Positive]: 5. The annotations were done manually by a research assistant and subsequently verified by two other research assistants.

We experimented with two prompt variations: zero-shot (no illustrative examples) and few-shot (20 demonstrative examples), and we used the standard precision, recall, and F-1 metrics to measure performance. The few-shot method yielded a significantly better performance of 0.8001 precision, 0.8201 recall, and 0.8099 F1-score (all averages) compared to the zero-shot method, which exhibited a much lower performance of 0.615 precision, 0.625 recall, and 0.6199 F1-score. Note that the reported values are averages, since each review was tagged with multiple aspect-sentiment pairs; precision, recall, and F1-scores were computed for each review in the dataset and then averaged across all the reviews. Upon careful analysis of the classifier outputs, the majority of classification errors occurred when the the reviews had the same aspect occurring twice, once for each sentiment (e.g., *pasta was excellent but the chicken tenders were cold*), thereby indicating that further experimentation with prompts is needed to differentiate between the two occurrences of the same aspect.

4.3 Focused Summarization

Once the subsets are identified for each aspect-sentiment pair, QuickCue next generates *focused* summaries pertaining to each pair. Recall that we need focused summarization due to the heterogeneous nature of information in reviews; general summarization techniques may omit salient information pertaining to the target aspect-sentiment pair and instead include information about other pairs in the generated summary. For this task, QuickCue employs the Directional Stimulus Prompting (DSP) technique for LLMs [55]. This prompting technique involves providing keywords as directional stimuli to tailor the summarization process via selective information prioritization. A snippet of our DSP prompt template is shown below.

> **Task Instructions:** Summarize the given reviews by focusing only on the specified main aspect and desired sentiment. Use the **Directional Stimuli (keywords)** for guidance. Ensure the generated summary **excludes unrelated aspects, redundant phrases, and undesired sentiments**, while keeping it concise and clear.
>
> **Reviews:** [Input reviews here]
> **Directional Stimuli:**
> **Main Aspect:** [Insert Desired Aspect]
> **Desired Sentiment:** [Insert Desired Sentiment]
>
> **Output Instruction:**
> *Generate the summary as a sequence of bullet points, with each point highlighting one salient feature uncovered about the specified aspect and desired sentiment.*
>
> **Examples:**
> **Reviews:** [Input reviews here]
> **Directional Stimuli:**
> **Main Topic:** Customer Service
> **Sentiment:** Negative
> **Output Summary:**
> - Many customers complained about slow service, stating that their orders took significantly longer than expected.
> - Some reviewers mentioned that the staff appeared inattentive and unresponsive, making it difficult to get assistance.
> - Several customers reported that the staff lacked knowledge about the menu, leading to confusion when placing orders.
> - Several reviews pointed out that employees lacked professionalism, often engaging in personal conversations rather than attending to customers.
> - Many visitors expressed frustration over the lack of courtesy from staff, mentioning that employees were often rude or dismissive.
>
> ...
>
> **INPUT:** [Input reviews here]

As seen in the above prompt, we included few-shot examples [5] within the prompt template to improve the quality of the summarization. We specifically crafted 10 few-shot examples (one for each aspect-sentiment pair), each of which included a set of randomly sampled reviews and the corresponding handcrafted summary.

Evaluation. To evaluate the efficacy of our DSP prompt in generating focused summaries, we built a test dataset comprising 50 examples – 5 for each aspect-sentiment pair. The ground truth summaries for these examples were carefully handcrafted and verified to ensure that they contained all the salient pieces of information pertaining to the corresponding aspect-sentiment pairs. To evaluate the quality of generated summaries, we relied on manual human evaluation, given the proven unreliability of automatic evaluation methods [75]. Specifically, 10 human annotators provided the following two metrics for each generated summary: (i) Factuality (1 for least accurate and 10 for highly accurate), which determines if all facts in the summary aligned with the reviews, and (ii) Noisiness (1 for highly noisy and 10 for least noisy), which evaluated the extent to which extraneous, off-topic information is present in a summary considering the target aspect-sentiment pair. For both of these metrics, the average score for each example (across the 10 annotators) was first computed, and then the overall average (i.e., average of averages) was computed across all the 50 examples in the test dataset.

The average factuality and noisiness scores for the few-shot DSP prompt were 7.9 and 8.3 respectively. Annotators noted that reduced scores for factuality often stemmed from incomplete information in generated summaries. A closer inspection of low-scoring examples revealed that this was mostly due to conflicting and vague information in reviews. For example, for a particular dish, some reviews expressed positive sentiment due to its spiciness whereas a few other reviews expressed a positive sentiment indicating that the same dish was not that spicy. The generated summary however only included the latter aspect of the dish as one of the salient points. This highlights the inherent ambiguity due to the subjectivity of the peoples' perceptions regarding the different aspects. Note that QuickCue follows a modular architecture, allowing the summarization module to be easily replaced with an improved version in future research.

4.4 User Interface

As mentioned earlier, QuickCue inserts its content into the existing Google Maps as an augmentation, so that the user does not have to shift focus to another page (see Figure 2). As shown in Figure 1, the content generated by QuickCue is arranged in a hierarchy comprising three layers (i.e., aspects, focused summaries, and original reviews). Specifically, QuickCue renders this hierarchy in HTML as an accordion, and automatically injects ARIA (Accessible Rich Internet Applications) attributes [87] to make it accessible. Moreover, QuickCue also adds tab-index attributes to relevant nodes in the accordion for enabling screen-reader users to easily navigate content at each layer using TAB and SHIFT+TAB hotkeys. QuickCue allows the users to navigate down the hierarchy via the ENTER key, and navigate up the hierarchy using the ESCAPE key. In sum, QuickCue simplifies interaction with information in reviews, limiting it to a few basic screen reader hotkeys.

4.5 Additional Implementation Details

The QuickCue was implemented as a Google Chrome browser extension, following open-source guidelines for browser extensions [29]. Extraction of reviews from Google Maps was done by leveraging pre-defined XPath information identifying the HTML DOM nodes corresponding to these reviews. The extracted data was structured into JSON objects and transferred to the backend through RESTful API calls [28] for further processing. Text preprocessing was performed using the NLTK [16] and spaCy [6] libraries, and noise elements such as emojis, out-of-vocabulary words, and excessive whitespace were filtered out using regular expressions [37]. All

ID	Age/ Gender	Age of Vision Loss	Occupation	Screen Reader	Web Experience
P1	43/M	Since birth	Teacher	JAWS	7 years
P2	36/M	Age 8	Unemployed	JAWS	4 years
P3	28/M	Since birth	Student	NVDA	11 years
P4	23/M	Age 10	Student	JAWS	8 years
P5	36/F	Since birth	Social Worker	JAWS	5 years
P6	32/M	Age 6	Teacher	JAWS	4 years
P7	25/F	NA	Unemployed	NVDA	2 years
P8	38/M	NA	Social Worker	JAWS	6 years
P9	31/F	Since birth	Teacher	JAWS	5 years
P10	22/F	Age 8	Student	JAWS	6 years

Table 1: Participant demographics. All information was self-reported by the participants.

data exchanges between QuickCue modules were in JSON for consistency and convenience. Integration of LLM into QuickCue was done using the well-known LangChain framework [22].

5 USER STUDY

We conducted an IRB-approved user study with screen reader users to evaluate the usability of QuickCue and compare it against the status quo. A total of 10 blind participants[1] (4 female, 6 male) were recruited through email lists and snowball sampling. The participants had an average age of 31.4 years (Median = 31.5, Range = 22–43). The inclusion criteria were: (i) familiarity with web screen reading and online review platforms; (ii) experience using the Chrome web browser and proficiency in JAWS; (iii) proficiency in using the standard QWERTY keyboard; and (iv) proficiency in communicating in English. To preserve external validity, we ensured that there was no overlap between the participant groups of this study and the prior interview study. All participants reported that they regularly engaged with customer reviews across various platforms, including shopping websites and food ordering services. The participant demographics are detailed in Table 1.

5.1 Design

In a within-subject experimental design, each participant was asked to freely explore and compare two restaurants on Google Maps based on their reviews, under the following two conditions:

- **Screen Reader**: The status quo baseline, where the participants used their screen reader to peruse reviews in the default Google Maps user interface.
- **QuickCue**: The participants used their screen reader to interact with the augmented Google Maps user interface, containing the accordion generated by our QuickCue assistive tool.

[1]This is the typical sample size for research in this area, due to the difficulty in recruiting participants belonging to this community [13, 67].

Influenced by the insights from the interview study, we chose a free-form comparison task to emulate real-world interaction scenarios in which users typically navigate reviews of multiple restaurants before making their decisions. To mitigate learning effects and avoid confounding variables, we selected four different restaurants for the two tasks. Additionally, we ensured that QuickCue accurately retrieved all reviews to prevent any confounding effects of retrieval accuracy. The assignment of restaurants to conditions and the ordering of conditions were counterbalanced across study participants using the well-known Latin Square method [17]. A maximum of 30 minutes was allotted for each task.

5.2 Procedure

At the beginning of the study, the experimenter explained the study objectives to the participant, and obtained an informed consent. This was followed by a practice session where the participant was allowed to familiarize with the QuickCue interface, refresh memories regarding screen reader hotkeys, and making any adjustments to the screen reader configuration (e.g., adjust speech rate). Note that all participants did the tasks on a Windows ThinkPad laptop provided by the experimenter, with all the necessary software installed, and also connected to an external standard QWERTY keyboard familiar to all the participants. The experimenter then asked the participant to complete the tasks in the pre-determined counterbalanced order. After the tasks, the experimenter administered the standard questionnaires, namely the System Usability Scale (SUS) [20] to assess usability, and the NASA Task Load Index (NASA-TLX) [42] to evaluate perceived workload. Lastly, the experimenter debriefed the participant in an exit interview, encouraging to provide subjective feedback, including the experience with QuickCue, difficulties while doing the tasks, and suggestions for improvement. All interactions were conducted in English. To prioritize participants' well-being, they were informed that they could take breaks or withdraw from the study at any time. The participants received $30 as compensation for their time.

5.3 Data Collection and Analysis

The experimenter documented participants' responses to the SUS and NASA-TLX questionnaires, recorded their think-aloud utterances during task execution, and observed their screen reader interaction behavior throughout the study. The SUS and NASA-TLX responses were analyzed using descriptive and inferential statistical methods. For qualitative data, we applied grounded theory methods [65], specifically the open coding and axial coding techniques [70] to systematically iterate over transcribed participant responses and uncover recurring themes and key insights. We present our findings next.

5.4 Results

5.4.1 Usability. As mentioned earlier, SUS questionnaire [20] was used to evaluate usability. Specifically, SUS asks the participants to rate ten statements on a 5-point Likert scale ranging from 1 to 5, with 1 indicating "strongly disagree" and 5 signifying "strongly agree." A final SUS score between 0 and 100 is then calculated by assimilating the individual ratings based on a predefined formula, and higher scores indicate better usability. The SUS scores for the

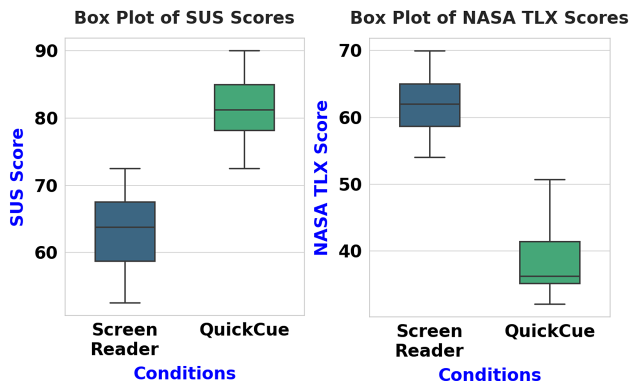

Figure 3: (Left) SUS scores, (right) NASA-TLX scores.

two experimental conditions are shown in Figure 3, where it can be clearly observed that the QuickCue condition received significantly higher scores compared to the default screen reader condition. Specifically, the Screen Reader condition received an average SUS score of 63.25 (median = 63.75, min = 52.5, max = 72.5), whereas the QuickCue condition received an average SUS score of 81.5 (median = 81.25, min = 72.5, max = 90.0). A one-way ANOVA test confirmed that the difference in usability scores between the two conditions was statistically significant $F = 45.03$, $p = 2.72 \times 10^{-6}$.

5.4.2 Task Workload. We employed the standard NASA-TLX questionnaire [42] to assess task workload. Like SUS, NASA-TLX also assimilates user's ratings into a score between 0 and 100, however, lower TLX scores indicate less workload and therefore better performance and user experience. The TLX score statistics for the two conditions are shown in Figure 3, where it can be observed that QuickCue significantly decreased the task workload, thereby substantially improving participants' user experience. Specifically, in the Screen Reader condition, the average TLX score was 62.09 (median = 62.0, min = 54.0, max = 69.93) and in the QuickCue condition, the average TLX score was 38.37 (median = 36.17, min = 32.0, max = 50.67). This difference in average TLX scores between conditions was also found to be statistically significant (One-way ANOVA, $F = 99.27$, $P = 9.45 \times 10^{-9}$). A closer inspection of individual ratings revealed that the *Effort* and *Temporal Demand* TLX sub-scales contributed relatively more to the significant difference between conditions, compared to the other four sub-scales.

5.4.3 Qualitative Feedback. The following core insights were uncovered from the qualitative analysis of the participants' subjective feedback during the exit interviews.

Simplistic design and ease of use. A majority (7) of participants attributed their higher usability perception of QuickCue to its simplistic design, which allowed them to navigate it using simple keyboard shortcuts that they were already familiar with. Additionally, they expressed appreciation for the *"summary feature"*, stating that it helped them to *"listen less and learn more"* and *"listen to only what they wanted"* about a restaurant. Regarding this, P3 said, " If I have these summaries, I will not at all listen to the reviews. It is extremely frustrating to listen to a lot of irrelevant and repeated feedback that barely tells me anything about what I would like to know about the food and the experience." The experimenter also noted such an interaction behavior during the study, where some of the participants did not bother going through the original reviews, and instead just listened to the summaries pertaining to a few aspect-sentiment pairs of interest. Five participants further stated that QuickCue would enable them to "explore more" and try ordering new dishes instead of the ordering the "same old tried-and-tested" items they are already familiar with at a given restaurant. For instance, P2 mentioned, *"This is very helpful...you know...I don't cook much, but when I order, it's always the same food because it takes a lot of time to sit and read reviews, and find something else...especially after a long day's work. If there's an easier way, of course, I'd use it to know what else is good."*

Extension of QuickCue to other review platforms. All participants appreciated how perusing user reviews on Google Maps was *"quite organized"* and *"less boring"* with QuickCue, enabling them to explore others' opinions with greater interest and helping them feel *"more confident"* in their dining decisions. Seven participants inquired if the system could be extended to other platforms, particularly e-commerce, where they prefer to read user reviews before making a purchase. For instance, P8 asked, *"Does this work on Amazon? I shop quite a lot there, and this would definitely help me make better purchase decisions. I'd like to know what others are saying about a product before I buy it...like its quality, whether it's worth the price, or if there are any updates or improvements."*

Repetitive search hinders restaurant comparison. Nearly all participants (9) mentioned that during the comparison task, they had to repeat the tedious process of searching for pieces of information pertaining to a specific aspect of interest, when then navigated to the reviews of the second restaurant from the reviews of the first restaurant. The experimenter also noted some of the think-aloud utterances that corroborate this statement, e.g., *"Okay, now I have to do it again. Let's find out where I can find comments about the taste and price of burgers."* Five participants further explained that such a repetitive search process was simply "too tiring" without additional support, and that QuickCue helped them reduce this effort to a "large extent". Nonetheless, this feedback highlights the need for personalized summarization to enable users to prioritize their favorite aspects across multiple restaurants, thereby facilitating convenient comparisons between restaurants.

6 DISCUSSION

6.1 Limitations

One limitation of our work is that we evaluated QuickCue exclusively with JAWS screen reader users. Although JAWS is the most popular screen reader, many blind users also use other screen readers such as NVDA and VoiceOver [85]. While QuickCue conceptually is screen reader-agnostic and will therefore likely produce similar results in the case of other screen reader users, the evaluation must nonetheless be conducted to validate our hypotheses.

Another limitation of our prototype is that it currently supports only desktop and laptop platforms. Research highlights a growing trend of users relying on smartphones to read and interact with online reviews, necessitating the adaptation of QuickCue for mobile platforms. However, this transition poses challenges, as mobile web browsers generally lack support for extensions. To overcome this limitation, we plan to explore alternative solutions, such as developing standalone service applications, to extend QuickCue's functionality to smartphone users.

A third limitation is that our few-shot learning and testing examples for joint classification and focused summarization were confined to English-language restaurant reviews, leaving the algorithm's effectiveness in other languages unexplored. The sizes of the samples too were relatively small due to the large amount of manual effort involved in building these test datasets. Moreover, QuickCue is currently restricted to the Chrome web browser. Although Chrome is the most commonly used browser among blind users, a significant portion of users still rely on alternatives such as Firefox [86] and Safari. Future work will therefore also focus on expanding QuickCue's compatibility to other languages and additional web browsers.

Another limitation of our work was that the technique to extract the reviews from the Google Maps webpage consisted of handcrafted rules based on predefined XPath patterns. For enabling QuickCue to function end-to-end, algorithms need to be devised to automatically detect and extract user reviews from webpages.

Furthermore, our evaluation was mostly qualitative, lacking quantitative metrics such as task completion time or error rates. While this approach provided valuable user experience insights, in a future study with a larger group of participants, we plan to design tasks that will facilitate quantitative analysis and therefore a more comprehensive assessment of QuickCue's effectiveness.

Also, as mentioned earlier, QuickCue was specifically designed and tested only for the review sections of restaurant menus, and its effectiveness on other types of websites is yet to be explored. However, given the modular and generalizable architecture of QuickCue, future research could focus on expanding its functionality to other website genres such as e-commerce platforms, classifieds, and entertainment sites, to evaluate its broader applicability.

6.2 Personalized User Preferences

In the subjective feedback, many participants highlighted the need for personalization in QuickCue, specifically the ability to store user preferences such as favorite aspect-sentiment pairs, and ensuring it automatically applies across all restaurant menus. Based on this feedback, we plan to extend QuickCue to support user-driven customization [14, 62]. To implement this, we will develop a pop-up interface that allows users to pre-select their preferences from predefined aspect-sentiment pair options. These selections will then be processed through a custom filtering algorithm, dynamically generating a user interface that reflects the chosen settings. In future work, we aim to fully integrate personalized support within QuickCue, enhancing usability and delivering a more tailored, user-centric experience.

6.3 Generalization Beyond Restaurant Reviews

Online customer reviews provide valuable, topical, and relevant feedback on service features and user experiences [82]. Numerous studies have investigated customer reviews in the restaurant sector, where the experiential nature of dining amplifies the impact of reviews and user comments [54]. Given this, our paper focused on developing QuickCue using restaurant reviews as a case study to enhance the reviews' usability and improve the overall user experience for blind users. However, QuickCue is not strictly limited to restaurant reviews, as its modular architecture was designed for inherent scalability across different domains. It consists of two core components: joint classification and focused summarization, enabling seamless adaptation to various review-based platforms, such as product reviews on e-commerce sites or classified advertisements. This flexibility requires minimal re-engineering, primarily involving data-driven modifications. For instance, in this study, we classify reviews based on five predefined aspects; transitioning to a different domain would require identifying and extracting relevant domain-specific aspects. Moreover, the prompt templates utilized in our approach are specifically curated for restaurant reviews and would also need to be adapted to align with the contextual requirements of other domains.

7 CONCLUSION

The current layout of customer reviews regarding restaurants primarily caters to the preferences and ease of sighted users. For blind users, however, this arrangement results in a tedious and frustrating content-consumption experience, requiring them to traverse large volumes of text while often encountering irrelevant content. To address this usability gap, we developed QuickCue, an intelligent assistive tool embodied as a browser extension specifically designed for blind screen reader users to conveniently access information in online customer reviews, specifically those pertaining to Google Maps. QuickCue streamlines access to review sections, allowing users to efficiently search for relevant information and compare restaurant menus more effectively, thereby enhancing decision-making. The QuickCue organizes the review section by breaking it down into selectable aspects (e.g., food, service, ambiance), followed by the presentation of positive and negative summaries for each aspect to provide a quick overview. In a user study with 10 blind participants, QuickCue significantly outperformed the status quo regarding usability and overall user experience.

ACKNOWLEDGMENTS

We sincerely thank Md Javedul Ferdous, Nithiya Venkatraman, and Abhinav Sai Choudary Panchumarthi for their assistance in preparing the seminal version of the manuscript.

REFERENCES

[1] Julio Abascal, Myriam Arrue, and Xabier Valencia. 2019. Tools for web accessibility evaluation. *Web accessibility: a foundation for research* (2019), 479–503.
[2] Shadi Abou-Zahra. 2008. Web accessibility evaluation. *Web accessibility: A foundation for research* (2008), 79–106.
[3] Hayfa Y Abuaddous, Mohd Zalisham Jali, and Nurlida Basir. 2016. Web accessibility challenges. *International Journal of Advanced Computer Science and Applications* 7, 10 (2016).
[4] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).
[5] Toufique Ahmed and Premkumar Devanbu. 2022. Few-shot training llms for project-specific code-summarization. In *Proceedings of the 37th IEEE/ACM international conference on automated software engineering*. 1–5.
[6] Explosion AI. 2024. spaCy: Industrial-Strength Natural Language Processing. https://spacy.io/ Accessed: March 3, 2025.

[7] Apple Inc. 2022. Accessibility - Vision - Apple. https://www.apple.com/accessibility/vision/.
[8] Jinat Ara, Md Toufique Hasan, Abdullah Al Omar, and Hanif Bhuiyan. 2020. Understanding customer sentiment: Lexical analysis of restaurant reviews. In *2020 IEEE Region 10 Symposium (TENSYMP)*. IEEE, 295–299.
[9] Vikas Ashok, Syed Masum Billah, Yevgen Borodin, and IV Ramakrishnan. 2019. Auto-suggesting browsing actions for personalized web screen reading. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. 252–260.
[10] Vikas Ashok, Yevgen Borodin, Svetlana Stoyanchev, Yuri Puzis, and IV Ramakrishnan. 2014. Wizard-of-Oz evaluation of speech-driven web browsing interface for people with vision impairments. In *Proceedings of the 11th Web for All Conference*. 1–9.
[11] Vikas Ashok, Yury Puzis, Yevgen Borodin, and IV Ramakrishnan. 2017. Web screen reading automation assistance using semantic abstraction. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*. 407–418.
[12] Ali Selman Aydin, Shirin Feiz, Vikas Ashok, and IV Ramakrishnan. 2020. Sail: Saliency-driven injection of aria landmarks. In *Proceedings of the 25th international conference on intelligent user interfaces*. 111–115.
[13] Jeffrey P Bigham, Anna C Cavender, Jeremy T Brudvik, Jacob O Wobbrock, and Richard E Ladner. 2007. WebinSitu: a comparative analysis of blind and sighted browsing behavior. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*. 51–58.
[14] Nilton Bila, Troy Ronda, Iqbal Mohomed, Khai N Truong, and Eyal De Lara. 2007. Pagetailor: reusable end-user customization for the mobile web. In *Proceedings of the 5th international conference on Mobile systems, applications and services*. 16–29.
[15] Syed Masum Billah, Vikas Ashok, Donald E Porter, and IV Ramakrishnan. 2017. Speed-dial: A surrogate mouse for non-visual web browsing. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. 110–119.
[16] Steven Bird, Ewan Klein, and Edward Loper. 2024. Natural Language Toolkit (NLTK). https://www.nltk.org/ Accessed: March 3, 2025.
[17] James V. Bradley. 1958. Complete Counterbalancing of Immediate Sequential Effects in a Latin Square Design. *J. Amer. Statist. Assoc.* 53, 282 (1958), 525–528. https://doi.org/10.1080/01621459.1958.10501456 arXiv:https://amstat.tandfonline.com/doi/pdf/10.1080/01621459.1958.10501456
[18] Giorgio Brajnik. 2004. Comparing accessibility evaluation tools: a method for tool effectiveness. *Universal access in the information society* 3 (2004), 252–263.
[19] Anke M Brock, Philippe Truillet, Bernard Oriola, Delphine Picard, and Christophe Jouffrais. 2015. Interactivity improves usability of geographic maps for visually impaired people. *Human–Computer Interaction* 30, 2 (2015), 156–194.
[20] John Brooke. 1996. Sus: a "quick and dirty'usability. *Usability evaluation in industry* 189, 3 (1996).
[21] Ben Caldwell, Michael Cooper, Loretta Guarino Reid, Gregg Vanderheiden, Wendy Chisholm, John Slatin, and Jason White. 2008. Web content accessibility guidelines (WCAG) 2.0. *WWW Consortium (W3C)* 290 (2008), 1–34.
[22] Harrison Chase. 2022. LangChain: Building Applications with LLMs through Composability. GitHub Repository. https://github.com/langchain-ai/langchain
[23] Garima Chhikara, Anurag Sharma, V Gurucharan, Kripabandhu Ghosh, and Abhijnan Chakraborty. 2024. LaMSUM: Amplifying Voices Against Harassment through LLM Guided Extractive Summarization of User Incident Reports. *arXiv preprint arXiv:2406.15809* (2024).
[24] Wendy Chisholm, Gregg Vanderheiden, and Ian Jacobs. 2001. Web content accessibility guidelines 1.0. *Interactions* 8, 4 (2001), 35–54.
[25] World Wide Web Consortium et al. 1999. Web content accessibility guidelines 1.0. (1999).
[26] Anupam Dash, Dongsong Zhang, and Lina Zhou. 2021. Personalized ranking of online reviews based on consumer preferences in product features. *International Journal of Electronic Commerce* 25, 1 (2021), 29–50.
[27] Kevin De Boeck, Jenno Verdonck, Michiel Willocx, Jorn Lapon, and Vincent Naessens. 2022. Reviewing review platforms: a privacy perspective. In *Proceedings of the 17th International Conference on Availability, Reliability and Security*. 1–10.
[28] Flask-RESTful Developers. 2024. Flask-RESTful Documentation. https://flask-restful.readthedocs.io/en/latest/ Accessed: March 3, 2025.
[29] Google Developers. 2024. Chrome Extensions Documentation. https://developer.chrome.com/docs/extensions/ Accessed: March 3, 2025.
[30] Tiago do Carmo Nogueira, Deller James Ferreira, Sérgio Teixeira de Carvalho, Luciana de Oliveira Berretta, and Mycke R Guntijo. 2019. Comparing sighted and blind users task performance in responsive and non-responsive web design. *Knowledge and Information Systems* 58 (2019), 319–339.
[31] Wafaa S El-Kassas, Cherif R Salama, Ahmed A Rafea, and Hoda K Mohamed. 2021. Automatic text summarization: A comprehensive survey. *Expert systems with applications* 165 (2021), 113679.
[32] Mohammed Fakrudeen, Sufian Yousef, and Mahdi H Miraz. 2017. Finger Based Technique (FBT): An Innovative System for Improved Usability for the Blind Users' Dynamic Interaction with Mobile Touch Screen Devices. *arXiv preprint arXiv:1708.05073* (2017).
[33] Limin Fang. 2022. The Effects of Online Review Platforms on Restaurant Revenue, Consumer Learning, and Welfare. *Manag. Sci.* 68 (2022), 8116–8143. https://api.semanticscholar.org/CorpusID:246595766
[34] Javedul Ferdous, Hae-Na Lee, Sampath Jayarathna, and Vikas Ashok. 2022. InSupport: Proxy Interface for Enabling Efficient Non-Visual Interaction with Web Data Records. In *27th International Conference on Intelligent User Interfaces*. 49–62.
[35] Javedul Ferdous, Sami Uddin, and Vikas Ashok. 2021. Semantic table-of-contents for efficient web screen reading. In *Proceedings of the 36th Annual ACM Symposium on Applied Computing*. 1941–1949.
[36] Raffaele Filieri. 2015. What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in e-WOM. *Journal of business research* 68, 6 (2015), 1261–1270.
[37] Python Software Foundation. 2024. re — Regular expression operations. https://docs.python.org/3/library/re.html Accessed: March 3, 2025.
[38] Freedom Scientific. 2022. JAWS ® – Freedom Scientific. https://www.freedomscientific.com/products/software/jaws/.
[39] Ioana Frîncu. 2023. In search of the perfect prompt. (2023).
[40] Guilherme Corredato Guerino and Natasha Malveira Costa Valentim. 2020. Usability and user experience evaluation of natural user interfaces: a systematic mapping study. *Iet Software* 14, 5 (2020), 451–467.
[41] Dorit Hadar-Shoval, Zohar Elyoseph, and Maya Lvovsky. 2023. The plasticity of ChatGPT's mentalizing abilities: personalization for personality structures. *Frontiers in Psychiatry* 14 (2023), 1234397.
[42] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139 – 183. https://doi.org/10.1016/S0166-4115(08)62386-9
[43] Xin Jin, Jonathan Larson, Weiwei Yang, and Zhiqiang Lin. 2023. Binary code summarization: Benchmarking chatgpt/gpt-4 and other large language models. *arXiv preprint arXiv:2312.09601* (2023).
[44] Elnaz Karamad. 2023. THE IMPACT OF ONLINE REVIEWS ON CONSUMER DECISION-MAKING: A SURVEY IN THE UKRAINIAN MARKET. *Market economy: modern management theory and practice* (2023). https://api.semanticscholar.org/CorpusID:268133863
[45] Madiha Khalid, Imran Ashraf, Arif Mehmood, Saleem Ullah, Maqsood Ahmad, and Gyu Sang Choi. 2020. GBSVM: sentiment classification from unstructured reviews using ensemble classifier. *Applied Sciences* 10, 8 (2020), 2788.
[46] Aurangzeb Khan, Umair Younis, Alam Sher Kundi, Muhammad Zubair Asghar, Irfan Ullah, Nida Aslam, and Imran Ahmed. 2020. Sentiment classification of user reviews using supervised learning techniques with comparative opinion mining perspective. In *Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC), Volume 2 1*. Springer, 23–29.
[47] Prerna Khanna, IV Ramakrishnan, Shubham Jain, Xiaojun Bi, and Aruna Balasubramanian. 2024. Hand Gesture Recognition for Blind Users by Tracking 3D Gesture Trajectory. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–15.
[48] Pooja Kherwa and Poonam Bansal. 2017. Latent semantic analysis: an approach to understand semantic of text. In *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*. IEEE, 870–874.
[49] Shashank Kumar, Jeevitha Shree DV, and Pradipta Biswas. 2021. Comparing ten WCAG tools for accessibility evaluation of websites. *Technology and Disability* 33, 3 (2021), 163–185.
[50] Jonathan Lazar, Aaron Allen, Jason Kleinman, and Chris Malarkey. 2007. What frustrates screen reader users on the web: A study of 100 blind users. *International Journal of human-computer interaction* 22, 3 (2007), 247–269.
[51] Hae-Na Lee and Vikas Ashok. 2021. Towards Enhancing Blind Users' Interaction Experience with Online Videos via Motion Gestures. In *Proceedings of the 32nd ACM Conference on Hypertext and Social Media*. 231–236.
[52] Hae-Na Lee and Vikas Ashok. 2022. Customizable Tabular Access to Web Data Records for Convenient Low-vision Screen Magnifier Interaction. *ACM Transactions on Accessible Computing (TACCESS)* 15, 2 (2022), 1–22.
[53] Hae-Na Lee, Vikas Ashok, and IV Ramakrishnan. 2020. Rotate-and-Press: A Non-visual Alternative to Point-and-Click?. In *HCI International 2020–Late Breaking Papers: Universal Access and Inclusive Design: 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings*. Springer, 291–305.
[54] Nao Li, Xiaoyu Yang, IpKin Anthony Wong, Rob Law, Jing Yang Xu, and Binru Zhang. 2023. Automating tourism online reviews: a neural network based aspect-oriented sentiment classification. *Journal of Hospitality and Tourism Technology* 14, 1 (2023), 1–20.
[55] Zekun Li, Baolin Peng, Pengcheng He, Michel Galley, Jianfeng Gao, and Xifeng Yan. 2024. Guiding large language models via directional stimulus prompting. *Advances in Neural Information Processing Systems* 36 (2024).
[56] Bing Liu. 2022. *Sentiment analysis and opinion mining*. Springer Nature.
[57] Christina Low, Emma McCamey, Cole Gleason, Patrick Carrington, Jeffrey P Bigham, and Amy Pavel. 2019. Twitter a11y: A browser extension to describe

images. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*. 551–553.

[58] Michael Luca. 2016. Reviews, reputation, and revenue: The case of Yelp. com. *Com (March 15, 2016). Harvard Business School NOM Unit Working Paper* 12-016 (2016).

[59] Michael Luca. 2016. Reviews, Reputation, and Revenue: The Case of Yelp.Com. https://api.semanticscholar.org/CorpusID:14511907

[60] Calvin Luy, Jeremy Law, Lily Ho, Richard Matheson, Tracey Cai, Anuradha Madugalla, and John Grundy. 2021. A toolkit for building more adaptable user interfaces for vision-impaired users. In *2021 IEEE symposium on visual languages and human-centric computing (VL/HCC)*. IEEE, 1–5.

[61] Valentyn Melnyk, Vikas Ashok, Valentyn Melnyk, Yury Puzis, Yevgen Borodin, Andrii Soviak, and IV Ramakrishnan. 2015. Look ma, no aria: generic accessible interfaces for web widgets. In *Proceedings of the 12th International Web for All Conference*. 1–4.

[62] Michael Nebeling, Maximilian Speicher, and Moira C Norrie. 2013. CrowdAdapt: enabling crowdsourced web page adaptation for individual viewing conditions and preferences. In *Proceedings of the 5th ACM SIGCHI symposium on Engineering interactive computing systems*. 23–32.

[63] NV Access. 2022. NV Access. https://www.nvaccess.org/.

[64] Uran Oh, Hwayeon Joh, and YunJung Lee. 2021. Image accessibility for screen reader users: A systematic review and a road map. *Electronics* 10, 8 (2021), 953.

[65] Julianne S Oktay. 2012. *Grounded theory*. Oxford University Press.

[66] OpenAI. 2024. GPT-4 Model Documentation. https://platform.openai.com/docs/models/gpt-4 Accessed: March 3, 2025.

[67] Yash Prakash, Akshay Kolgar Nayak, Mohan Sunkara, Sampath Jayarathna, Hae-Na Lee, and Vikas Ashok. 2024. All in One Place: Ensuring Usable Access to Online Shopping Items for Blind Users. *Proceedings of the ACM on Human-Computer Interaction* 8, EICS (2024), 1–25.

[68] Yash Prakash, Mohan Sunkara, Hae-Na Lee, Sampath Jayarathna, and Vikas Ashok. 2023. AutoDesc: facilitating convenient perusal of web data items for blind users. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. 32–45.

[69] Pradeep Racherla and Wesley Friske. 2012. Perceived 'usefulness' of online consumer reviews: An exploratory investigation across three services categories. *Electron. Commer. Res. Appl.* 11 (2012), 548–559. https://api.semanticscholar.org/CorpusID:28895996

[70] Johnny Saldaña. 2021. The coding manual for qualitative researchers. (2021).

[71] Ather Sharif, Sanjana Shivani Chintalapati, Jacob O Wobbrock, and Katharina Reinecke. 2021. Understanding screen-reader users' experiences with online data visualizations. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–16.

[72] Daniel Siahaan, Achmad Maududie, Agus Subekti, Hotniar Siringoringo, et al. 2017. A new approach for modeling researcher community based on scientific article metadata using natural language processing. In *2017 Seventh International Conference on Information Science and Technology (ICIST)*. IEEE, 386–390.

[73] Beverley A Sparks and Victoria Browning. 2011. The impact of online reviews on hotel booking intentions and perception of trust. *Tourism management* 32, 6 (2011), 1310–1323.

[74] Manjunath Srinivas, S Vamsi Krishna Reddy, Manoj NM, and Hatsuho Miyazawa. 2024. Evaluation of ChatGPT, Gemini and Llama-2 for E-commerce Product Attribute Extraction. In *Proceedings of the 2024 10th International Conference on e-Society, e-Learning and e-Technologies (ICSLT)*. 43–48.

[75] Amanda Stent, Matthew Marge, and Mohit Singhai. 2005. Evaluating evaluation methods for generation in the presence of variation. In *International conference on intelligent text processing and computational linguistics*. Springer, 341–351.

[76] Xiaofei Sun, Xiaoya Li, Jiwei Li, Fei Wu, Shangwei Guo, Tianwei Zhang, and Guoyin Wang. 2023. Text Classification via Large Language Models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, Houda Bouamor, Juan Pino, and Kalika Bali (Eds.). Association for Computational Linguistics, Singapore, 8990–9005. https://doi.org/10.18653/v1/2023.findings-emnlp.603

[77] Mohan Sunkara, Yash Prakash, Hae-Na Lee, Sampath Jayarathna, and Vikas Ashok. 2023. Enabling customization of discussion forums for blind users. *Proceedings of the ACM on Human-Computer Interaction* 7, EICS (2023), 1–20.

[78] Oguzhan Tas and Farzad Kiyani. 2007. A survey automatic text summarization. *PressAcademia Procedia* 5, 1 (2007), 205–213.

[79] Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805* (2023).

[80] Mike Thelwall. 2024. ChatGPT for complex text evaluation tasks. *Journal of the Association for Information Science and Technology* (2024).

[81] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* (2023).

[82] Ana Valdivia, M Victoria Luzón, and Francisco Herrera. 2017. Sentiment analysis on tripadvisor: Are there inconsistencies in user reviews?. In *Hybrid Artificial Intelligent Systems: 12th International Conference, HAIS 2017, La Rioja, Spain, June 21-23, 2017, Proceedings 12*. Springer, 15–25.

[83] Damir Vandic, Lennart J Nederstigt, Flavius Frasincar, Uzay Kaymak, and Enzo Ido. 2024. A framework for approximate product search using faceted navigation and user preference ranking. *Data & Knowledge Engineering* 149 (2024), 102241.

[84] Byron C Wallace, Sayantan Saha, Frank Soboczenski, and Iain J Marshall. 2021. Generating (factual?) narrative summaries of rcts: Experiments with neural multi-document summarization. *AMIA Summits on Translational Science Proceedings* 2021 (2021), 605.

[85] WebAIM. 2019. WebAIM: Screen Reader User Survey #8 Results. https://webaim.org/projects/screenreadersurvey8/

[86] Jared Smith WebAIM. 2015. Screen reader user survey# 5 results. *http://webaim.org/projects/screenreadersurvey6/*. (2015).

[87] World Wide Web Consortium (W3C). 2021. Accessible Rich Internet Applications (WAI-ARIA) 1.2. https://www.w3.org/TR/wai-aria-1.2/. Accessed: 2025-02-14.

[88] Shaomei Wu, Jeffrey Wieland, Omid Farivar, and Julie Schiller. 2017. Automatic alt-text: Computer-generated image descriptions for blind users on a social network service. In *proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 1180–1192.

[89] Jingnan Xu, Xiaolin Zheng, and Weifeng Ding. 2012. Personalized recommendation based on reviews and ratings alleviating the sparsity problem of collaborative filtering. In *2012 IEEE Ninth International Conference on e-Business Engineering*. IEEE, 9–16.

[90] Xianjun Yang, Yan Li, Xinlu Zhang, Haifeng Chen, and Wei Cheng. 2023. Exploring the limits of chatgpt for query or aspect-based text summarization. *arXiv preprint arXiv:2302.08081* (2023).

[91] Ali Yayli and Murat Bayram. 2012. e-WOM: the effects of online consumer reviews on purchasing decisions. *International Journal of Internet Marketing and Advertising* 7 (2012), 51. https://api.semanticscholar.org/CorpusID:168100918

[92] Jonathan Zong, Crystal Lee, Alan Lundgard, JiWoong Jang, Daniel Hajas, and Arvind Satyanarayan. 2022. Rich screen reader experiences for accessible data visualization. In *Computer Graphics Forum*, Vol. 41. Wiley Online Library, 15–27.

[93] Cristina Zuheros, Eugenio Martínez-Cámara, Enrique Herrera-Viedma, and Francisco Herrera. 2021. Sentiment analysis based multi-person multi-criteria decision making methodology using natural language processing and deep learning for smarter decision aid. Case study of restaurant choice using TripAdvisor reviews. *Information Fusion* 68 (2021), 22–36.