Improving Usability of Data Charts in Multimodal Documents for Low Vision Users

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

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

Shoaib Mohammed Alyaan
Old Dominion University
Department of Computer Science
Norfolk, Virginia, USA
salya006@odu.edu

Pathan Aseef Khan
Old Dominion University
Department of Computer Science
Norfolk, Virginia, USA
pkhan002@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

Data chart visualizations and text are often paired in news articles, online blogs, and academic publications to present complex data. While chart visualizations offer graphical summaries of the data, the accompanying text provides essential context and explanation. Associating information from text and charts is straightforward for sighted users but presents significant challenges for individuals with low vision, especially on small-screen devices such as smartphones. The visual nature of charts coupled with the layout of the text inherently makes it difficult for low vision users to mentally associate chart data with text and comprehend the content due to their dependence on screen magnifier assistive technology, which only displays a small portion of the screen at any instant due to content enlargement. To address this problem, in this paper, we present a smartphone-based multimodal mixed-initiative interface that transforms static data charts and the accompanying text into an interactive slide show featuring frames containing "magnified views" of relevant data point combinations. The interface also includes a narration component that delivers tailored information for each "magnified view". The design of our interface was informed by a user study with 10 low-vision participants, aimed at uncovering low vision interaction challenges and user-interface requirements with multimodal documents that integrate text and chart visualizations. Our interface was also evaluated in a subsequent study with 12 low-vision participants, where we observed significant improvements in chart usability compared to both status-quo screen magnifiers and a state-of-the-art solution.

CCS CONCEPTS

• **Human-centered computing** → **Accessibility technologies**; *Empirical study*.

ICMI '24, November 04–08, 2024, San Jose, Costa Rica

ACM ISBN 979-8-4007-0462-8/24/11
https://doi.org/10.1145/3678957.3685714

KEYWORDS

Low vision; Graph usability; Screen magnifier; Graph perception

ACM Reference Format:
Yash Prakash, Akshay Kolgar Nayak, Shoaib Mohammed Alyaan, Pathan Aseef Khan, Hae-Na Lee, and Vikas Ashok. 2024. Improving Usability of Data Charts in Multimodal Documents for Low Vision Users. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '24), November 04–08, 2024, San Jose, Costa Rica.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3678957.3685714

1 INTRODUCTION

Data visualizations like bar charts, line charts, and pie charts are commonly used to present quantitative data. These charts are increasingly presented with accompanying text and are featured in various data storytelling formats, including news magazines, slide presentations, and videos [51] and are consumed at large, specifically on smartphones [49]. When these charts are integrated with explanatory text, they create a 'synergistic effect,' i.e., charts capture the audience's attention and provide a perceptually effective way to represent data while accompanying text guides the viewers and adds necessary context [32]. While sighted individuals can effortlessly switch between text and charts quickly, low-vision users [26, 61] face significant challenges, especially on smartphones, even if they can technically "*see*" the charts and accompanying text. This is because they can only view a small portion of text or chart content at any given time due to content magnification, given the fixed and limited screen size of smartphones. To see the occluded parts, low-vision users must constantly move their screen magnifier lens all over the content, a process known as *panning*. Panning is well known to cause significant cognitive overhead and other usability issues for low vision users [34, 35, 64].

Existing solutions to improve accessibility of data charts have primarily focused on blind screen reader users [39]. These include providing alternative textual descriptions, natural sounds, and audio-haptic interfaces [20, 22, 54]. While these solutions can also be used by low vision users, they do not exploit the residual function vision of these users, which can be a powerful input modality while interacting with charts. Also, as these solutions do not offer direct interaction with the data [52] and instead provide only the author's interpretation of the data, they prevent users from forming their own analyses and insights [55].

To enhance the user experience of low-vision users with charts, specifically in multimodal documents, it is essential to first understand their preferences and requirements. To this end, we conducted a user study with 10 low-vision participants. In the study, the participants expressed a preference for viewing data points (e.g., bars in bar charts) within the same magnifier viewport to avoid issues associated with excessive panning and zooming. Also, the participants noted difficulties in shifting focus between text and charts, which was exacerbated by magnification-induced disorientation. They expressed that the process of sequentially navigating through text while repositioning the magnifier lens for subsequent lines, and then transitioning to related visual charts for a more comprehensive understanding, was both tedious and cumbersome. Specifically, they said that the continuous magnifier adjustment hindered their ability to integrate information seamlessly.

Building on the identified preferences and requirements of low-vision individuals, as well as previous research [51, 59], we devised ChartSync to enhance the low vision usability of charts in multimodal documents on smartphones. ChartSync transforms traditionally static charts on web pages into an *interactive slide show* (see Figure 1). Specifically, ChartSync identifies important combinations of data points in charts, selecting them based on their relevance to the accompanying text and significant data facts present within the chart. ChartSync then automatically generates "magnified views" of these data point combinations, which are presented to low-vision users as a slide show. Each slide also includes a voice-based component that delivers customized information for each magnified view, providing summaries of specific data points rather than traditional overall data chart summaries [20, 28].

We evaluated ChartSync in a user study with 12 low-vision participants. In the study, the subjective feedback for ChartSync was significantly more positive than the baseline methods – default screen magnifier and a state-of-the-art method [12]. All participants reported that ChartSync significantly reduced their cognitive load while interacting with data charts, allowing them to view the desired content simultaneously even with magnification. In sum, our contributions are: (i) The findings of a user study illuminating the preferences and requirements of low-vision users when interacting with charts on multimodal documents using screen magnifiers; and (ii) ChartSync – an assistive technology application that provides an alternative usable interaction mode for charts.

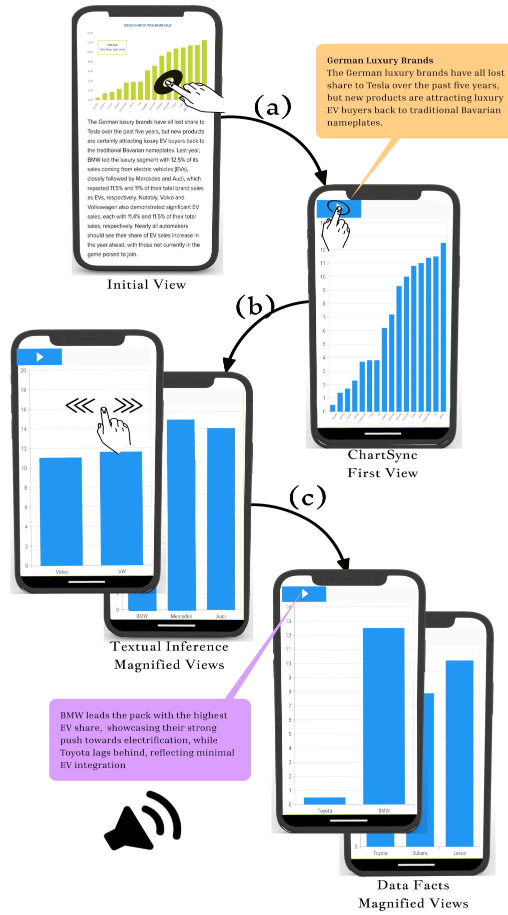

Figure 1: Illustration of ChartSync. When the user selects a chart in a document, ChartSync opens up an alternative interactive interface containing a slideshow of salient chart content. The default first view shows the full chart along with an audio summary. The user can swipe right/left to go through the remaining slides containing magnified views of select data points along with the corresponding tailored audio descriptions.

2 RELATED WORK

2.1 Low Vision Interaction with Smartphones

Extant research has predominantly focused on blind screen reader users [11, 25, 27, 29]. Prior research on the needs of low vision screen magnifier users is still in its infancy [13, 58]. Szpiro et al. [58] conducted a study examining the interaction behaviors of low-vision users utilizing screen magnifiers on touch devices like smartphones. The research highlighted several accessibility and usability challenges these individuals encountered while using smartphone applications. Key findings included: (i) Difficulty in panning back and forth following content enlargement; (ii) The need to remember and implement various multi-finger gestures; (iii) Challenges with content-agnostic screen enlargement that complicated navigation and comprehension of the app content; and (iv) The time-consuming nature of constantly changing zoom levels to properly view the content. Although this study provides insight into the general interaction challenges faced by low-vision users on smartphones, the specific impact of these issues on their interaction with multimodal documents, especially those with charts, is yet to be explored – a knowledge gap that will be addressed in our work.

There also exist works that have proposed solutions to improve usability for low-vision screen magnifier users [3, 6, 33, 35, 36, 41]. However, these solutions were primarily tailored for desktop interaction scenarios; very few works that have concentrated on improving smartphone interaction for low vision users [24, 41]. Moreover, these solutions have primarily focused on *local context preservation* in general-purpose mobile and desktop interaction. Therefore, they are less helpful for global contextual preservation, i.e., situations where content is spatially distributed as in the case of multimodal documents with chart visualizations.

2.2 Data Charts in Multimodal Documents

Prior work has shown that users often struggle to integrate information effectively across text and data visualizations, thereby indicating a need for enhanced interaction support while consuming such content [42]. Therefore, many works have focused on improving the relationship between charts and text to facilitate better information assimilation for sighted users [31, 32, 45, 57, 70]. For instance Latif et al. [32] developed *Kori*, a system that comprises three main components: a chart gallery, an editing area, and a link setting panel. As users type using a keyboard, *Kori* automatically suggests potential links between text and charts, highlighting possible connections with dotted gray underlines. Users can accept these automatic suggestions or manually create links through simple interactions, enabling a dynamic and interactive storytelling environment. There also exist a few other solutions exclusively designed for people who are blind [20, 63]. For example, Greenbacker et al. [20] developed a system for enhancing comprehension of multimodal documents containing line graphs. Their system pinpoints the most relevant paragraph within the article and delivers a summary of the graph at that point, thereby improving the coherence of the presentation.

As above solutions are primarily designed for either sighted or blind users, they lack the necessary usability features needed for catering specifically to low-vision users. Although low-vision users can technically use these systems with screen magnifiers, they are likely to face significant challenges, e.g., extensive panning and precise visual/motor skills, in tasks such as identifying automatic suggestions, previewing options, and interacting with visual elements. These solutions also fail to consider the residual visual capabilities that many low-vision users rely on when interacting with digital devices. This paper therefore introduces a novel system, ChartSync, specifically designed to enhance the usability of charts in multimodal documents for low-vision users who depend on screen magnifiers.

3 UNCOVERING LOW-VISION USER NEEDS

We conducted an Institutional Review Board (IRB)-approved study with 10 low-vision participants to identify their interaction challenges as well as explore their needs and preferences while interacting with multimodal documents on smartphones. The average age of participants was 26.1 (SD: 1.58) years and the gender representation was equal (5 male, 5 female). The participants had diverse eye conditions including optic atrophy, glaucoma, retinitis pigmentosa, LCA, nystagmus, and cataracts. All participants stated that they browsed the web daily for at least 2 hours. The visual acuity of the participants too varied between 20/100 and 20/500 (good eye).

In the study, participants were presented with multimodal documents and were instructed to interpret data from charts and their corresponding textual content. After the interpretation was complete, participants were asked a series of questions related to the chart they had just reviewed and its associated textual content. Examples of these questions included: *What difficulties did you encounter when interpreting data charts and accompanying text?*, *What strategies did you employ to overcome these challenges?*, *How satisfied were you with the information gained from the multimodal representation?*. The feedback gathered from the user study was qualitatively analyzed using an open coding technique [50], which involved an iterative review and annotation of user responses to discover recurring key insights, pain points, and themes.

Findings: Some of the notable themes from the qualitative analysis are presented next.

Visual clarity preferred in chart interpretation. Most participants (7) preferred simpler bar and line charts over more complex multi-bar and multi-line charts, as they found the latter visually cluttered and confusing. Additionally, three participants engaged in estimating the height of bars by using an adjacent bar as a reference point. For instance, in a multi-bar chart featuring three bars, they focused on the tallest bar and gauged the height of the other two bars by comparing them to the tallest one, which led to errors in estimating the height of the bars in the bar chart.

Preference for context over clarity. Most of the participants (6) expressed a preference for viewing bars in bar charts or trends in line charts within the same frame or viewport of the smartphone. Although they had the option to enhance their visual perception by zooming into specific chart details, they preferred to trade off magnification-induced clarity for context preservation.

Text overload reduces chart engagement. Four participants noted an abundance of redundant text in multimodal documents. The time required to read this textual content via screen magnification led to cognitive overload, prompting them to entirely skip reviewing the charts associated with the text.

Problem assimilating text and charts. All participants expressed frustration over having to navigate through extensive text and spend additional time examining charts using screen magnifiers, highlighting the difficulty of integrating these two forms of information effectively. Synthesizing information across two distinct modalities (text and charts) presents a significant challenge when these elements are spatially separated. This spatial separation requires readers to continually switch their attention between the textual content and the visual cues on the charts that encode data (such as bars, lines, or points). This cognitive load phenomenon is known as the split-attention effect [4] as defined in cognitive load theory [43], and corresponds to the contiguity principle in the cognitive theory of multimedia learning [40]. Cognitive overload due to split attention is naturally exacerbated for individuals with low vision when using screen magnifiers compared to sighted individuals.

The results of the preliminary user study provided the needs and requirements of low-vision users interacting with multimodal documents and reinforced the design of ChartSync.

4 CHARTSYNC INTERFACE

Figure 2 presents the design schematic of ChartSync mobile assistive technology. Upon loading a webpage (e.g., blog, news article) in ChartSync – a browser-based mobile application, it leverages a custom-trained Inception-V3 model [68] to automatically identify and classify the types of charts (e.g., bar chart, line chart) on the page. Following the classification of charts, ChartSync leverages the ChartOCR model [38] to extract detailed elements like labels, legends, and data values from the charts. Next, ChartSync extracts

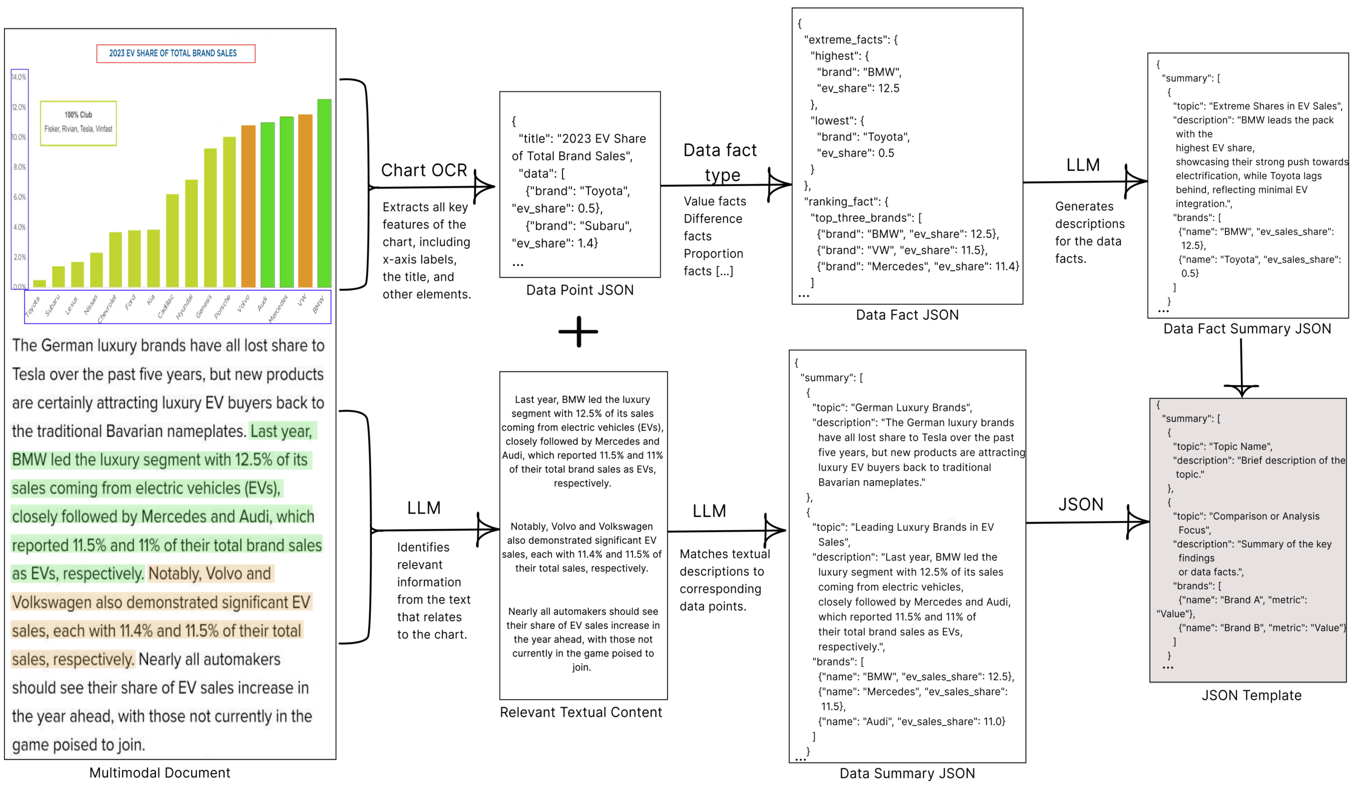

Figure 2: ChartSync architectural workflow: This figure illustrates the backend operations of ChartSync from the moment a user selects a chart. The data summary JSON lists relevant data points from associated text, and the data fact summary JSON details key facts within the chart.

the text paragraphs relevant to each identified chart by measuring the similarity between the chart's textual components and the content of each paragraph in the document. ChartSync then leverages a pre-trained LLaMA Model (LLM) [60], guided through custom prompt engineering, to accurately identify and extract key combinations of data points from charts, linking them to their corresponding textual content (see Section 4.2). Additionally, ChartSync identifies key data facts within the chart, e.g., Ranking fact – ranking the top three elements (see Section 4.3).

When a user selects a data chart on the loaded webpage using a single tap gesture [9], ChartSync activates a multimodal mixed-initiative interface tailored for low-vision users (see Figure 1). Utilizing a swipe gesture [18], the user can then explore an *interactive slideshow* showcasing tailored "magnified views" of the chart comprising select combinations of important data points in each slide. Furthermore, each slide is equipped with a voice button; when activated, it provides a customized *data narration* that focuses on the specific set of data points highlighted in that slide (see Figure 1).

4.1 ChartSync Interface Design

Informed by prior research [9], the user interface of ChartSync was specifically designed for ease of navigation using simple one-finger gestures, as opposed to the traditional two-finger slide gestures required by standard OS accessibility services. To engage with the ChartSync interface, a user has to simply tap on a chart in the current webpage. This action opens the interface where the first "view" presented is an interactive version of the selected chart. Upon pressing the voice button with a one-finger 'Tap' gesture, users can listen to an audio narrative that provides an overview of the chart, including its title, the names of the axes, and any captions associated with the figure (see Figure 1).

When a user performs a side-ward swipe gesture, they can navigate through a set of "views" as in an interactive slideshow. Each slide showcases a magnified view of crucial data points from the chart, organized into two sequential sets on the interface. The first set presents combinations of data points that align with the accompanying text of the chart, allowing users to listen to an audio narration of this text while simultaneously viewing the data points. This setup helps illustrate the author's perspective and provides a correlation between the visual data and textual explanation. The second set focuses on significant data facts from the chart (see Section 4.3). At any point during the interaction, the user can execute a upward swipe gesture to invoke the 'Customizer' tool for personalizing the appearance of the chart (e.g., color and contrast).

Design Choices. Our design choices were informed by our preliminary user study and supported by existing research. Traditionally, readers interact with news articles through a "Martini Glass" approach – initially following the narrative in text and later exploring the visual elements that interest them [51]. This method is unimodal, requiring users to process text and visualizations separately. Such unimodal navigation is not optimal for low-vision users, given their dependence on screen magnifiers. Instead, they benefit more from a cross-modal congruent approach [59]. Cross-modal congruence involves delivering stimuli from different senses in a coordinated manner, enhancing the consistency and compatibility between them. An example of this is watching a video of a dog barking while simultaneously hearing the bark, which aligns the auditory and visual senses and fosters a more integrated perception. This alignment can significantly improve cognitive processing and reaction times. In ChartSync, we apply this principle by simultaneously presenting visual content highlighting relevant data points in charts and corresponding audio narrations.

User-Centered Design. In our work, we adopted a user-centered design approach [10], involving 5 participants throughout the development process. The idea behind engaging these users directly in the development was to enhance satisfaction and acceptance of ChartSync. We provided them with a high-fidelity prototype [62], designed as a web app interface on a smartphone. This prototype was developed based on insights gathered from our preliminary user study, allowing us to test design concepts and gather feedback on the functionality and flow of the design.

Overall, the feedback from all participants was positive; however, they identified several notable limitations in the earlier stages of the design process. Notable issues included the absence of an introductory slide providing an overview of the chart, the absence of a button to easily navigate to the full view of the chart from the slideshow, and the one-dimensional flow of the slideshow, lacking the functionality to revisit previously viewed charts. In response to this constructive feedback, we implemented substantial enhancements to address these concerns and refine the design. Furthermore, we integrated customization features as suggested by the participants.

Chart customization. ChartSync features a ‘Customizer’ tool that enables users to adjust the color and contrast of data charts through a simple upward swipe gesture. In designing our Customizer, we adhered to the WCAG 2.1 guidelines [1], specifically criteria 1.4.3 and 1.4.11, to ensure the color and contrast settings are optimal for accessibility and user experience.

4.2 Generating Data Narratives

We employed prompt engineering to instruct a pre-trained LLaMA Model (LLM) [60] to identify and extract relevant data points from text paragraphs associated with the chart data. This prompt provides the contextual cues and delineates the steps involved in accurately identifying all data points referenced in the text. Specifically, we leveraged Chain-of-Thought (CoT) [65] and ReAct [69] prompting techniques, which instill reasoning abilities into LLMs and help formulate responses through a series of logical steps. The CoT prompt directs the LLM to discern two types of chart references [32] in each phrase of the text: (i) Point-level Matching, where discrete data points directly mentioned in the text are pinpointed, and (ii) Interval-level Matching, where data points across described ranges or intervals in the text are identified. The ReAct prompt guides the LLM on subsequent actions, including the validation and grouping of data points based on their references in the text. An example

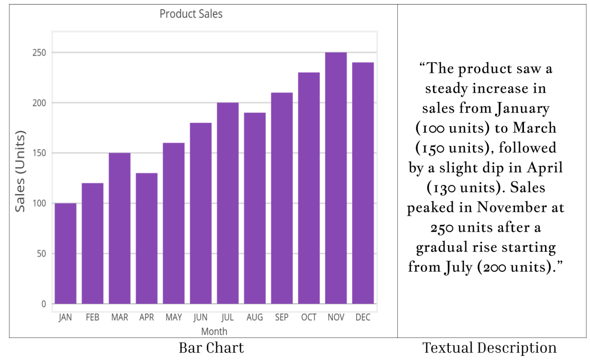

Figure 3: Example bar chart and accompanying text to illustrate the CoT and ReAct prompting used in ChartSync.

of this methodology, as applied to the chart and text presented in Figure 3, is detailed below:

i. **Identification Using CoT:**
  - **For Sentence 1:**
    - **Point-level Matching:** Jan: 100 units, Mar: 150 units, Apr: 130 units
    - **Interval-level Matching:** Increase from Jan to Mar, Dip in Apr
    - **Data Range:** Encompasses Jan, Feb, Mar, and Apr
  - **For Sentence 2:**
    - **Point-level Matching:** Jul: 200 units, Nov: 250 units
    - **Interval-level Matching:** Gradual rise from Jul to Nov
    - **Data Range:** Encompasses Jul, Aug, Sep, Oct, and Nov.

ii. **Grouping Using ReAct:**
  - **For Group 1: Early Year Sales Trend (Jan to Apr):**
    - **Reference Grouping:** Include Sentence 1.
    - **Data Points:** Jan (100 units), Feb (120 units), Mar (150 units), Apr (130 units).
    - **Identified Ranges:** Increase from Jan to Mar and a dip in Apr.
  - **For Group 2: Mid to Late Year Sales Build-Up (Jul to Nov)**
    - **Reference Grouping:** Include Sentence 2.
    - **Data Points:** Jul (200 units), Aug (190 units), Sep (210 units), Oct (230 units), Nov (250 units).
    - **Identified Range:** Gradual rise from Jul to Nov.

Demonstrative examples of reasoning (Chain of Thought, CoT) and actions (ReAct) were manually constructed for a sampled set of data-text pairs and were provided to the LLM using a few-shot prompt template [8, 47].

Evaluation. To evaluate the accuracy of identifying data points from charts corresponding to sentences in the associated textual paragraphs, we calculated the F1 score [15]. Ground truth data was manually generated by annotating relevant data points in a sample of 25 chart-text pairs, which included various chart types supported by ChartSync, such as simple and multi-bar charts and line charts. We then compared ChartSync’s output to this ground truth, resulting in a precision of 0.75, a recall of 0.78, and an F1 score of 0.764.

4.3 Data Fact Narratives

ChartSync includes supplementary data facts that go beyond the information outlined in the text. However, the potential range of data facts is extensive due to the myriad combinations of data cells and types of data facts available, making it difficult to compile all of them in real time. To overcome this problem, we selected only a few salient data facts that can provide a detailed overview to the low-vision user. For this, we first collected nine types of data facts based on prior work by Shi et al. [53], and then selected only those data facts that were not only computationally feasible (e.g., Extreme facts that highlight the smallest and largest values of bars in a bar chart), but also aligned with the Gestalt laws of perceptual organization [66] – principles that explain how humans instinctively perceive visual elements as organized patterns and objects. Examples include:

- Categorization Facts: These align with the Law of Similarity, which states that elements that are similar tend to be perceived as more related than those that are dissimilar. For example, in a survey about job satisfaction, you might categorize responses into 'Satisfied' and 'Dissatisfied.'
- Ranking Facts: These correlate with the Law of Proximity, suggesting that physically close elements are perceived as more related. For example, "The countries with the highest rates of literacy are Norway, Australia, and Canada."
- Trend Facts: These are associated with the Law of Continuity, which implies that elements arranged on a line are perceived to be more related than those arranged randomly. For example, "The rate of deforestation in the Amazon has steadily increased from 2015 to 2020."

We utilized Calliope [53] to identify potential facts from the data extracted via chart OCR [38]. We employed natural language generation templates from prior research [53] and refined these templates to enhance the naturalness of the generated sentences.

4.4 Implementation Details

We implemented ChartSync[1] as an Android mobile browser application developed using an open source framework, namely Flutter [18]. When the user loads a webpage, ChartSync leverages in-built Dart functions [14] to extract the entire HTML DOM of the webpage and send it to the backend server via a POST request. We used the Beautiful Soup [48] Python package to extract all images in the DOM, labeling them with positional IDs. These images are then sent to a custom-trained Inception-V3 model [19] trained on CHARTEX dataset [2] and annotated with flags (True, False) to indicate if they are data charts. Subsequently, all chart images are processed using ChartOCR [56] to extract data attributes such as labels, legends, and data values. Then, to identify textual paragraphs associated with charts, we implemented the P-KLA algorithm [20] in Python and extended it with an LLM [23] for enhanced text analysis.

The chart data points and the corresponding textual paragraph pairs are then passed to a pre-trained LLaMA Model [23] along with a well-engineered prompt to accurately match data points with the corresponding sentences in the textual paragraph, including demonstrative examples from the Kori dataset [32]. The prompt template was constructed using Chain-of-Thought (CoT) [65] and ReAct prompting techniques [69] as explained before. ChartSync simultaneously utilizes Calliope [53] to identify existing data facts within the charts, which are then sent to the LLM along with a natural language template to construct corresponding summaries. ChartSync then packages the relevant sentences and associated data points from textual paragraphs, identifies data facts for all the charts as a JSON object, and ships it to the Flutter module.

When the user taps on a chart, ChartSync creates an interactive slideshow using Flutter UI modules. Each slide in the slideshow showcases data points corresponding to a data fact present in the aforementioned JSON object, using the FLChart package [16]. Additionally, each slide includes an audio button powered by the Flutter TTS package [17] that reads out the text associated with the data facts. Other functionalities, such as customization features, were implemented using Dart [14].

[1]https://github.com/accessodu/ChartSync.git

5 EVALUATION

To assess the efficacy of ChartSync, we conducted an IRB-approved user study with low-vision screen magnifier users.

5.1 Participants

We recruited a total of 12 low vision users[2]. The average age of the participants was 34.1 years (Median = 34.5, Minimum = 28, Maximum = 40) and the gender representation was balanced (5 female and 7 male). The included participants were proficient screen magnifier users who did not rely on screen readers. All participants indicated that they browsed the web daily for at least 2 hours and had visual acuity ranging from 20/100 to 20/500 (good eye). The eye conditions included cataracts, glaucoma, pigmentosa, cancer, and diabetic retinopathy.

5.2 Design

In a within-subject setup, the participants did a "free task" where they were asked to draw inferences from charts and the associated text associated under three distinct conditions:

- Screen Magnifier (*SM*) – Participants used their preferred screen magnifier to complete the task.
- Data Chart Summary (*DS*) – The participants could listen to the audio narration of the chart summary to perform the task [37].
- ChartSync (*CS*) – The participants leveraged ChartSync to complete the task.

During the study, no time restrictions were imposed, allowing the participants ample opportunity and time to interpret the charts and accompanying textual descriptions. At the end of each study condition, the participants were asked: "What inferences did you derive from the chart and associated text?". To avoid confounds, we compiled a collection of 12 multimodal documents in which ChartSync demonstrated optimal performance, achieving 100% accuracy in extracting data points from charts and correctly identifying associated text paragraphs. The charts in these documents contained an average of 20 data points and at least three references to these data points in the accompanying texts. To ensure unbiased results, the assignment and ordering of documents for the task were counterbalanced using the Latin square method [30].

[2]This is the typical sample size for research in this area due to the difficulty in recruiting participants belonging to this disadvantaged community.

5.3 Procedure

The experimenter first obtained formal consent from participants and briefly outlined the study's objectives. Participants were introduced to the mobile app, followed by a 15-minute practice session to familiarize themselves with the interface and adjust settings as needed. They then completed the study tasks in a predetermined order. After completing the tasks, a questionnaire was administered to assess usability and perceived effort. The session concluded with a brief exit interview for subjective feedback. With consent, the session was recorded, and any notable interaction behaviors were documented. Participants received an Amazon gift card, and all interactions were conducted in English. Additional details regarding the apparatus used in the study are available on GitHub[1].

5.4 Data Collection and Analysis

We collected a comprehensive set of metrics and data, which included: (i) task completion times; (ii) chart narrations, a summary of the inferences derived by the participant from the provided chart; (iii) responses to the System Usability Scale (SUS) questionnaire [7] to assess perceived usability; (iv) responses to the NASA Task Load Index (NASA-TLX) questionnaire [21] for evaluating perceived user effort; and (v) qualitative feedback from participants as well as observations made by the experimenter. The subjective feedback from the participants, coupled with the experimenter's notes, were analyzed using the open coding technique [67]. We report our findings in the following subsection.

5.5 Results

Table 1: Participant task performance data including study condition, task completion time (TC), chart narration (CN), SUS, and NASA-TLX.

Condition	TC (min)	CN (BLEU)	SUS	NASA-TLX
SM	6.5	0.45	47.25	77.63
DS	4.3	0.54	57.75	48.1
CS	3.4	0.71	73.25	29.77

5.5.1 Task Completion times. On average, the participants spent 3.4 minutes (Median = 3.1, Minimum = 2.3, Maximum = 4.8) completing tasks in the ChartSync condition, which was significantly lower than the average time spent in the screen magnifier condition (Mean = 6.5 minutes; Median = 6.8, Minimum = 5.5, Maximum = 8.1) and the chart summary condition (Mean = 4.3 minutes; Median = 4.1, Minimum = 3.7, Maximum = 5.6). This difference in task completion times was statistically significant, as indicated by the one-way ANOVA test[3] ($F = 72.82$, $p < 0.001$), demonstrating the efficiency of the ChartSync condition in comparison to the other conditions. The faster completion times observed with ChartSync can be attributed to its inherent design, which facilitates quicker navigation and information inference through interactive slide shows, unlike the baseline screen magnifier condition, where participants had to first read and memorize text before moving to the chart to draw inferences; ChartSync streamlined this process. In the audio summary (DS) condition, the participants often chose to listen to the entire summary while simultaneously using the screen magnifier to pinpoint important points.

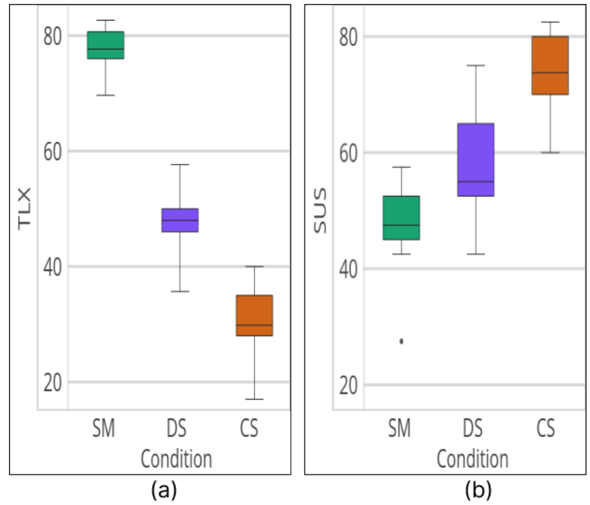

Figure 4: (a) Task workload (NASA-TLX) and (b) Perceived usability (SUS) for all three study conditions.

5.5.2 Chart Narration. During the user study, the participants were asked to provide a detailed summary of their inferences from the chart to understand their comprehension. We assessed the effectiveness of this chart comprehension using the BLEU score [44, 46]. This metric quantifies the lexical similarity between the participant-expressed narrative summaries and a ground truth baseline established by annotators from academic backgrounds. The BLEU score for ChartSync was 0.71, markedly surpassing the scores in other conditions – 0.54 for chart summary and 0.45 for screen magnifier. This indicates that ChartSync significantly enhanced users' understanding of the charts. In the baseline screen magnifier condition, the participants' inferences about the chart were typically vague, offering only a general overview without delving into specific factual details, which explains the lower observed BLEU scores. In the chart summary (DS) condition, although the inferences included more factual details compared to the baseline, they appeared somewhat scattered and did not convincingly demonstrate a comprehensive understanding of the chart. On the other hand, with ChartSync, the sequence and quality of facts provided by the users indicated a deeper and more structured understanding of the chart; this was evidenced by higher BLEU scores. Additionally, the experimenter noticed an increase in the confidence of participants' responses when discussing their inferences with ChartSync, a level of assurance that was absent in the other two conditions.

5.5.3 Usability and Perceived Workload. To evaluate the usability of ChartSync, we employed the System Usability Scale (SUS) questionnaire [5, 7]. The SUS questionnaire consists of 10 alternating positive and negative 5-scale Likert statements, where a rating of 5 corresponds to strongly agree, 1 represents strongly disagree, and

[3]We ensured that all statistical tests adhered to necessary assumptions. We verified the independence of observations within each group, confirmed the normal distribution of data within groups via the Shapiro-Wilk test, and verified the homogeneity of variances across groups.

3 represents a neutral rating. The responses to these 10 statements are then assimilated into a single score between 0 and 100, with higher scores indicating higher usability of the system. Figure 4(b) displays the SUS statistics for the three study conditions. The average SUS scores for ChartSync was $\mu = 73.25$ ($\sigma = 6.61$), which was statistically higher than those for both screen magnifiers ($\mu = 47.25$, $\sigma = 7.86$) and chart summary. ($\mu = 57.75$ $\sigma = 9.58$) condition (one-way Anova test, $F = 23.39$, $p < 0.001$).

We employed the NASA-TLX (NASA Task Load Index) questionnaire [21] to measure the perceived workload of participants. The questionnaire collects subjective ratings across six subscales: Mental Demand, Physical Demand, Temporal Demand, Overall Performance, Effort, and Frustration Level. Responses are aggregated into a single score ranging from 0 to 100, where lower scores indicate better performance, contrasting with the SUS scoring system. Our analysis revealed a significant impact of the study conditions on the NASA-TLX scores, as evidenced by the results of the one-way ANOVA test ($F = 195.55$, $p < 0.001$). Figure 4 (a) displays the NASA-TLX statistics for the three study conditions. Notably, the TLX scores for the ChartSync condition ($\mu = 29.77$, $\sigma = 6.04$) were significantly lower than those for the screen magnifier ($\mu = 77.63$, $\sigma = 3.5$) and chart summary ($\mu = 48.1$, $\sigma = 5.62$) conditions. This significant difference in TLX scores was further validated through pairwise comparisons using the post-hoc Tukey's HSD test, which confirmed that ChartSync significantly outperformed both the screen magnifier ($Q = 27.72$, $p < 0.001$) and chart summary conditions ($Q = 10.62$, $p = 0.016$). The reasons behind the participants' SUS and NASA-TLX ratings across the different study conditions were elucidated through the analysis of feedback collected during the open-ended exit interviews.

5.5.4 Qualitative Feedback. The analysis of the subjective feedback from the exit interviews revealed the following insights:

ChartSync is user-friendly and straightforward. A majority of the participants (8) attributed ChartSyncs' high usability ratings to the intuitive and easily navigable interface, which facilitated a smooth user experience, even for individuals who were new to multimodal documents. The participants also noted the ease with which they became familiar with the system's features. Furthermore, participants reported no notable latency in retrieving relevant data points and their associated text.

Need to enhance navigational access. While the majority of participants (10) were satisfied with the information provided by ChartSync, few participants (4) expressed a desire for broader access to additional paragraphs present in the multimodal documents. This feedback underscores a common need to seek more comprehensive interaction with content beyond the specific data points and charts initially highlighted by the system. Participants indicated that having the ability to explore and interact with other sections of the document would enrich their understanding and allow for a more holistic grasp of the presented information.

6 DISCUSSION

The user study demonstrated that ChartSync improves multimodal graphical understanding and usability for low-vision individuals who use screen magnifiers on smartphones (see Table 1). Nonetheless, our approach had its limitations, which highlight potential areas for future research. We will discuss notable ones next.

Limitations. In the evaluation study, we focused on simple bar charts, ensuring both data extraction and chart-text pairing were 100% accurate to eliminate confounding variables. Therefore, the performance of ChartSync must be evaluated "in the wild" on random charts and text pairs on multimodal documents. For this purpose, a separate user study is necessary to evaluate how the accuracy of data extraction and identification of relevant text paragraphs influence ChartSync's usability.

Another limitation was the reduced accuracy in charts with densely packed data points. Moreover, the accuracy of identifying relevant paragraphs associated with charts was not extensively tested. However, recent advancements in multimodal large language models (MLLM) for chart interpretation offer significant opportunities to enhance extraction performance. Given the modular architecture of ChartSync, it would be straightforward to replace the current algorithm with a more robust MLLM-based solution in the future, thereby improving overall system accuracy and reliability.

Lastly, ChartSync was only implemented for smartphone chart interaction, and therefore it is completely untested in the desktop/laptop environments. However, we do believe that the benefits of ChartSync will carry over to the desktop/laptop environments and help improve low-vision chart interaction on desktops. Deployment of ChartSync on desktops/laptops will also be easier in the form of browser extensions, which can provide support real-time support on arbitrary web pages containing data charts.

Screen magnifier skimming. Based on the participants' feedback, we aim to develop a system that allows users to input a specific topic of interest in a multimodal document. This system would then identify all relevant sentences and data points from charts associated with that topic and automatically facilitate panning, moving the screen magnifier across to showcase important information. Furthermore, the system will be designed to allow users to interact dynamically by adjusting the zoom level and panning speed according to their preferences to facilitate natural skimming.

7 CONCLUSION

In this paper, we first investigated how low-vision users interact with multimodal documents containing data charts and uncovered their pain points as well as their interaction requirements and preferences. Informed by these findings, we then developed ChartSync, a smartphone browser-based assistive technology for low-vision screen magnifier users to interact with multimodal chart data via an alternative interactive slideshow interface that enables them to obtain quick magnified 'views' of salient information in the charts. Evaluation of ChartSync in a user study with 12 low vision screen magnifier users demonstrated the effectiveness of ChartSync over status-quo solutions while also illuminating the limitations of our current prototype. Future improvements include large-scale evaluation, deeper algorithm testing, and support for a diverse set of multimodal documents with data visualizations.

ACKNOWLEDGMENTS

We sincerely thank Amanda Kelly Dcosta and Aditya Vishal for their invaluable contributions to this work.

REFERENCES

[1] Nancy Alajarmeh. 2022. The extent of mobile accessibility coverage in WCAG 2.1: sufficiency of success criteria and appropriateness of relevant conformance levels pertaining to accessibility problems encountered by users who are visually impaired. *Universal Access in the Information Society* 21, 2 (2022), 507–532.

[2] asbljy. 2024. DeepRuleDataset. https://huggingface.co/datasets/asbljy/DeepRuleDataset/tree/main.

[3] Ali Selman Aydin, Shirin Feiz, Vikas Ashok, and IV Ramakrishnan. 2020. Towards making videos accessible for low vision screen magnifier users. In *Proceedings of the 25th international conference on intelligent user interfaces*. 10–21.

[4] Paul Ayres and John Sweller. 2005. The split-attention principle in multimedia learning. *The Cambridge handbook of multimedia learning* 2 (2005), 135–146.

[5] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3 (2009), 114–123.

[6] Syed Masum Billah, Vikas Ashok, Donald E Porter, and IV Ramakrishnan. 2018. SteeringWheel: a locality-preserving magnification interface for low vision web browsing. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

[7] John Brooke. 1996. Sus: a "quick and dirty'usability. *Usability evaluation in industry* 189, 3 (1996).

[8] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.

[9] Maria Claudia Buzzi, Marina Buzzi, Barbara Leporini, and Amaury Trujillo. 2015. Exploring visually impaired people's gesture preferences for smartphones. In *Proceedings of the 11th Biannual Conference of the Italian SIGCHI Chapter*. 94–101.

[10] Adriana Chammas, Manuela Quaresma, and Cláudia Mont'Alvão. 2015. A closer look on the user centred design. *Procedia Manufacturing* 3 (2015), 5397–5404.

[11] Sarah Chiti and Barbara Leporini. 2012. Accessibility of android-based mobile devices: a prototype to investigate interaction with blind users. In *International Conference on Computers for Handicapped Persons*. Springer, 607–614.

[12] Jinho Choi, Sanghun Jung, Deok Gun Park, Jaegul Choo, and Niklas Elmqvist. 2019. Visualizing for the non-visual: Enabling the visually impaired to use visualization. In *Computer Graphics Forum*, Vol. 38. Wiley Online Library, 249–260.

[13] World Wide Web Consortium et al. 2016. Accessibility Requirements for People with Low Vision. https://www.w3.org/TR/low-vision-needs/.

[14] Dart Team. 2024. Dart Programming Language. https://dart.dev/.

[15] Leon Derczynski. 2016. Complementarity, F-score, and NLP Evaluation. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. 261–266.

[16] FL Chart Contributors. 2024. Flutter Package. https://pub.dev/packages/fl_chart.

[17] Flutter Team. 2024. Flutter Package for TTS. https://pub.dev/packages/flutter_tts.

[18] Flutter Team. 2024. Write Your First Flutter App. https://docs.flutter.dev/get-started/codelab.

[19] Google Cloud Team. 2024. Advanced Guide to Inception v3 on Cloud TPU. https://cloud.google.com/tpu/docs/inception-v3-advanced.

[20] Charles F Greenbacker, Peng Wu, Sandra Carberry, Kathleen F McCoy, Stephanie Elzer, David D McDonald, Daniel Chester, and Seniz Demir. 2011. Improving the accessibility of line graphs in multimodal documents. In *Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies*. 52–62.

[21] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[22] Md Naimul Hoque, Md Ehtesham-Ul-Haque, Niklas Elmqvist, and Syed Masum Billah. 2023. Accessible Data Representation with Natural Sound. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[23] Hugging Face. 2024. LLaMA Model Documentation. https://huggingface.co/docs/transformers/main/en/model_doc/llama.

[24] 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. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 2 (2023), 1–36.

[25] Mohit Jain, Nirmalendu Diwakar, and Manohar Swaminathan. 2021. Smartphone usage by expert blind users. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[26] Randall T Jose. 1983. *Understanding low vision*. American Foundation for the Blind.

[27] Shaun K Kane, Jeffrey P Bigham, and Jacob O Wobbrock. 2008. Slide rule: making mobile touch screens accessible to blind people using multi-touch interaction techniques. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*. 73–80.

[28] Shankar Kanthara, Rixie Tiffany Ko Leong, Xiang Lin, Ahmed Masry, Megh Thakkar, Enamul Hoque, and Shafiq Joty. 2022. Chart-to-Text: A Large-Scale Benchmark for Chart Summarization. *arXiv preprint arXiv:2203.06486* (2022).

[29] Akif Khan and Shah Khusro. 2019. Blind-friendly user interfaces–a pilot study on improving the accessibility of touchscreen interfaces. *Multimedia Tools and Applications* 78, 13 (2019), 17495–17519.

[30] Beob G Kim and Taemin Kim. 2010. A program for making completely balanced Latin Square designs employing a systemic method. *Revista Colombiana de Ciencias Pecuarias* 23, 3 (2010), 277–282.

[31] Nicholas Kong, Marti A Hearst, and Maneesh Agrawala. 2014. Extracting references between text and charts via crowdsourcing. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 31–40.

[32] Shahid Latif, Zheng Zhou, Yoon Kim, Fabian Beck, and Nam Wook Kim. 2021. Kori: Interactive synthesis of text and charts in data documents. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 184–194.

[33] 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.

[34] Hae-Na Lee, Vikas Ashok, and IV Ramakrishnan. 2021. Bringing things closer: Enhancing low-vision interaction experience with office productivity applications. *Proceedings of the ACM on Human-computer Interaction* 5, EICS (2021), 1–18.

[35] Hae-Na Lee, Yash Prakash, Mohan Sunkara, IV Ramakrishnan, and Vikas Ashok. 2022. Enabling Convenient Online Collaborative Writing for Low Vision Screen Magnifier Users. In *Proceedings of the 33rd ACM Conference on Hypertext and Social Media*. 143–153.

[36] Hae-Na Lee, Sami Uddin, and Vikas Ashok. 2020. TableView: Enabling efficient access to web data records for screen-magnifier users. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–12.

[37] Mengsha Liu, Daoyuan Chen, Yaliang Li, Guian Fang, and Ying Shen. 2024. ChartThinker: A Contextual Chain-of-Thought Approach to Optimized Chart Summarization. *arXiv preprint arXiv:2403.11236* (2024).

[38] Junyu Luo, Zekun Li, Jinpeng Wang, and Chin-Yew Lin. 2021. ChartOCR: data extraction from charts images via a deep hybrid framework. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 1917–1925.

[39] Rubén Alcaraz Martínez, Mireia Ribera Turró, and Toni Granollers Saltiveri. 2019. Accessible statistical charts for people with low vision and colour vision deficiency. In *Proceedings of the XX International Conference on Human Computer Interaction*. 1–2.

[40] Richard E Mayer and Logan Fiorella. 2014. 12 principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In *The Cambridge handbook of multimedia learning*. Vol. 279. Cambridge University Press New York, NY, 279–315.

[41] Farhani Momotaz and Syed Masum Billah. 2021. Tilt-Explore: Making Tilt Gestures Usable for Low-Vision Smartphone Users. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 1154–1168.

[42] Alvitta Ottley, Aleksandra Kaszowska, R Jordan Crouser, and Evan M Peck. 2019. The curious case of combining text and visualization. *EuroVis 2019-Short Papers* (2019).

[43] Fred Paas and John Sweller. 2014. Implications of cognitive load theory for multimedia learning. *The Cambridge handbook of multimedia learning* 27 (2014), 27–42.

[44] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*. 311–318.

[45] Joao Pinheiro and Jorge Poco. 2022. ChartText: Linking Text with Charts in Documents. *arXiv preprint arXiv:2201.05043* (2022).

[46] 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.

[47] Jack W Rae, Sebastian Borgeaud, Trevor Cai, Katie Millican, Jordan Hoffmann, Francis Song, John Aslanides, Sarah Henderson, Roman Ring, Susannah Young, et al. 2021. Scaling language models: Methods, analysis & insights from training gopher. *arXiv preprint arXiv:2112.11446* (2021).

[48] Leonard Richardson. 2015. Beautiful Soup Documentation. https://beautiful-soup-4.readthedocs.io/en/latest/.

[49] María-Carmen Ricoy, Sara Martínez-Carrera, and Isabel Martínez-Carrera. 2022. Social overview of smartphone use by teenagers. *International journal of environmental research and public health* 19, 22 (2022), 15068.

[50] Johnny Saldaña. 2021. *The coding manual for qualitative researchers*. sage.

[51] Edward Segel and Jeffrey Heer. 2010. Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics* 16, 6 (2010), 1139–1148.

[52] 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* (Virtual Event, USA) *(ASSETS '21)*.

Association for Computing Machinery, New York, NY, USA, Article 14, 16 pages. https://doi.org/10.1145/3441852.3471202
[53] Danqing Shi, Xinyue Xu, Fuling Sun, Yang Shi, and Nan Cao. 2020. Calliope: Automatic visual data story generation from a spreadsheet. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 453–463.
[54] Alexa Siu, Gene SH Kim, Sile O'Modhrain, and Sean Follmer. 2022. Supporting accessible data visualization through audio data narratives. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19.
[55] Alexa F Siu, Danyang Fan, Gene SH Kim, Hrishikesh V Rao, Xavier Vazquez, Sile O'Modhrain, and Sean Follmer. 2021. COVID-19 highlights the issues facing blind and visually impaired people in accessing data on the web. In *Proceedings of the 18th International Web for All Conference*. 1–15.
[56] soap117. 2023. DeepRule. https://github.com/soap117/DeepRule.
[57] Arjun Srinivasan, Bongshin Lee, and John Stasko. 2020. Interweaving multimodal interaction with flexible unit visualizations for data exploration. *IEEE Transactions on Visualization and Computer Graphics* 27, 8 (2020), 3519–3533.
[58] Sarit Felicia Anais Szpiro, Shafeka Hashash, Yuhang Zhao, and Shiri Azenkot. 2016. How people with low vision access computing devices: Understanding challenges and opportunities. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*. 171–180.
[59] Stefano Targher, Valeria Occelli, and Massimiliano Zampini. 2012. Audiovisual integration in low vision individuals. *Neuropsychologia* 50, 5 (2012), 576–582.
[60] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288* (2023).
[61] W3C. 2023. 2023. Low vision definition. https://www.w3.org/WAI/GL/low-vision-a11y-tf/wiki/Definitions_of_Low_Vision
[62] Miriam Walker, Leila Takayama, and James A Landay. 2002. High-fidelity or low-fidelity, paper or computer? Choosing attributes when testing web prototypes. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 46. Sage Publications Sage CA: Los Angeles, CA, 661–665.
[63] Yanan Wang and Yea-Seul Kim. 2023. Making data-driven articles more accessible: An active preference learning approach to data fact personalization. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 1353–1366.
[64] Yanan Wang, Yuhang Zhao, and Yea-Seul Kim. 2024. How Do Low-Vision Individuals Experience Information Visualization?. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–15.
[65] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems* 35 (2022), 24824–24837.
[66] Max Wertheimer. 1938. Laws of organization in perceptual forms. (1938).
[67] David Wicks. 2017. The coding manual for qualitative researchers. *Qualitative research in organizations and management: an international journal* 12, 2 (2017), 169–170.
[68] Xiaoling Xia, Cui Xu, and Bing Nan. 2017. Inception-v3 for flower classification. In *2017 2nd international conference on image, vision and computing (ICIVC)*. IEEE, 783–787.
[69] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2022. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629* (2022).
[70] Qiyu Zhi, Alvitta Ottley, and Ronald Metoyer. 2019. Linking and layout: Exploring the integration of text and visualization in storytelling. In *Computer Graphics Forum*, Vol. 38. Wiley Online Library, 675–685.