A11yPDF: A Comprehensive Evaluation Tool Designed to Improve PDF Accessibility Awareness and Compliance

ANONYMOUS AUTHOR(S)

Portable Document Format (PDF) files are widely used for sharing documents that need to be viewed or printed exactly as intended by the author due to their platform-independent nature. However, like any digital content, PDFs can present significant accessibility challenges, preventing individuals with disabilities from fully accessing or understanding the information. Although there are several accessibility evaluation tools available for PDFs, many of these tools have limitations, including the inability to assess certain document elements, high costs associated with paid licenses, and complex interfaces that make them difficult to use. In response to these challenges, this paper introduces A11yPDF, a free and comprehensive tool designed to evaluate and enhance the accessibility of PDF documents. A11yPDF offers several advantages over existing tools: it provides a more extensive evaluation of a wide range of accessibility elements, features an intuitive user interface that simplifies the evaluation process, and generates clear, actionable reports that guide users in making their content more accessible.

Additional Key Words and Phrases: Accessibility, PDF accessibility, Tools, WCAG.

ACM Reference Format:
Anonymous Author(s). 2024. A11yPDF: A Comprehensive Evaluation Tool Designed to Improve PDF Accessibility Awareness and Compliance. In *Proceedings of CHI Conference on Human Factors in Computing Systems (CHI '25).* ACM, New York, NY, USA, 31 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Portable Document Format (PDF) files are widely used in various sectors, including education. They are known for their versatility and widespread adoption as a file format for documents, ensuring consistent presentation and formatting across various devices and platforms. Nevertheless, they may present accessibility issues for people with diverse abilities using assistive technologies [21]. Navigation and understanding information on PDF files are hindered by the frequent lack of essential accessibility features, including structured tags [72, 78], labeled headers [97], and alternate text for image [54, 92, 110]. Additionally, complex PDF layouts and non-linear structures can further disrupt screen readers and magnification tools [76, 106].

Despite established guidelines like PDF/UA-1 [34] and the Matterhorn Protocol [69], numerous studies reveal a low compliance rate for accessibility standards in PDF documents. For instance, Nazemi [76] reviewed 200 research articles in PDF format from 2009 to 2013 and revealed that 97% of the journals lacked alternate text for images and only 4.5% were tagged PDF documents. Similarly, Fernandes et al. [39] identified a concerning trend of non-compliance with accessibility guidelines in healthcare resources distributed as PDF files. Many of these materials lack essential accessibility features, making them difficult or impossible for individuals with diverse abilities and assistive technologies to access. WANG et al. [106] further emphasized this issue, revealing that only 2.4% of academic PDFs meet accessibility

Manuscript submitted to ACM

standards, demonstrating a significant gap between awareness and implementation. Addressing this gap through effective accessibility testing is crucial for improving information access for individuals with diverse needs.

While numerous PDF accessibility evaluation and improvement tools exist, widespread adoption faces several limitations. Tools like PDFA Inspector [1], PAVE [2], ABBYY FineReader [3], Common Look [4], and Adobe Acrobat Pro Accessibility Checker [5] can be cost-prohibitive due to paid licenses and require technical expertise for effective use, creating barriers for many users [36, 106]. Additionally, these tools often prioritize text and interactive elements, with less emphasis on non-text content like images, links, and navigational elements. This focus on specific aspects neglects broader accessibility considerations and hinders significant improvements in the overall accessibility of PDF documents.

Therefore, this research aims to address accessibility concerns in PDF and identify user needs by employing a two-part survey study. First, we conducted a pre-survey to explore participants' awareness of PDF accessibility challenges and their current practices for evaluating accessibility. The findings from this pre-survey motivated the development of a user-centered PDF Accessibility Analyzer tool named A11yPDF. This tool aims to enhance the accuracy, efficiency, and intuitiveness of accessibility evaluation. Following the development and implementation of A11yPDF, a post-survey was conducted to assess ease of use, interpretation, and effectiveness in evaluating PDF accessibility for participants who interact with the tool.

Accordingly, this research investigates user awareness of PDF accessibility through the following research question (RQ):

RQ1: What is the level of user awareness regarding PDF accessibility challenges? The objective of this research question is to assess and quantify the current level of user awareness about the challenges and requirements associated with making PDFs accessible. This includes evaluating users' understanding of the specific aspects necessary to achieve PDF accessibility, their familiarity with relevant accessibility guidelines and laws, and their knowledge of where to find resources or requirements related to PDF accessibility.

RQ2: What is the extent to which users adopt accessibility requirements for PDF in practice?

This research question seeks to explore the practical application of accessibility requirements by users when creating PDFs. It aims to understand whether users consider accessibility aspects during PDF creation and whether they have previously used any tools to evaluate PDF accessibility. The findings will provide insight into current practices and the adoption rate of accessibility requirements.

RQ3: What are the most important accessibility requirements according to users?

The purpose of this research question is to identify which accessibility features users consider most important when evaluating PDF documents. By analyzing user preferences and priorities, this question aims to inform the development and refinement of tools like A11YPDF to better align with user needs and expectations.

Furthermore, the research explores the potential impact of the A11yPDF tool on user behavior through four additional research questions:

RQ4: To what extent do users perceive the A11yPDF tool as helpful in evaluating PDF accessibility?

This research question aims to evaluate user perceptions of the A11YPDF tool's effectiveness in assessing PDF accessibility. It focuses on understanding user experiences with the tool, particularly in terms of accuracy, usability, and

[1] https://github.com/pdfae/PDFAInspector
[2] https://pdf.abbyy.com/
[3] https://pdf.abbyy.com/
[4] https://monsido.com/monsido-commonlook-partnership
[5] https://www.adobe.com/accessibility/products/acrobat/using-acrobat-pro-accessibility-checker.html

the perceived value of the simulations provided. This information will help determine the tool's practical utility and identify areas for improvement.

RQ5: Does the use of the A11yPDF tool have a measurable impact on user awareness of PDF accessibility?

The goal of this research question is to measure the impact of using the A11YPDF tool on users' awareness of PDF accessibility. It examines whether exposure to the tool and its features has enhanced users' understanding of accessibility principles, influenced their attitudes toward designing accessible PDFs, and improved their confidence in applying these principles in practice.

RQ6: How does our tool compare to other tools that evaluate PDF accessibility?

This research question seeks to compare the A11YPDF tool with other existing PDF accessibility evaluation tools from the perspective of users who have experience with both. The objective is to gather comparative insights on usability, functionality, and overall effectiveness, which will help position A11YPDF within the broader landscape of accessibility tools.

RQ7: Is A11yPDF a functional tool for evaluating PDF accessibility?

The purpose of this research question is to assess the overall functionality and practicality of the A11YPDF tool. It evaluates user satisfaction, ease of navigation, and clarity of information. Additionally, it seeks user feedback and gauges the likelihood of future use, ensuring the tool meets practical needs in real-world settings.

The contributions of this study are:

- We provide an in-depth analysis of user awareness regarding the challenges of PDF accessibility. We also examine the effectiveness and limitations of current tools used for evaluating and ensuring PDF accessibility. This analysis uncovers significant gaps in both awareness and tool functionality that need to be addressed to improve accessibility outcomes.
- We introduce A11YPDF, a novel web application specifically designed to comprehensively evaluate PDF documents against a broad spectrum of accessibility criteria, covering a wide array of features critical for ensuring accessibility across various disability types and needs.
- We explore the effectiveness of A11YPDF in enhancing user understanding and practices related to PDF accessibility via pre- and post-survey analysis. The pre-survey establishes baseline user awareness, while the post-survey measures the tool's impact on improving knowledge and changing behavior. This dual-phase evaluation provides empirical evidence of A11YPDF's contribution to advancing PDF accessibility awareness and practices among users.

The remainder of this article is organized as follows: Section 2 presents the literature relevant to the underlying research area. Section 3 details the survey design and the specific procedures used to implement the tool in this study. Section 4 presents the survey findings, which serve as the foundation for the development of the PDF accessibility evaluation tool and also evaluates the tool's performance through user experiences. Section 6 addresses the potential threats to the validity of this study. Finally, Section 7 concludes this research.

2 Related Work

This section outlines a review of existing literature on PDF accessibility which covers the challenges associated with making PDFs accessible and the current landscape of tools available for PDF accessibility evaluation. A comparison of existing tools' features is included to highlight the potential significance of the proposed tool in addressing current limitations.

2.1 State of PDF Accessibility

Web accessibility has become a well-established field, with many researchers and practitioners adhering to the WCAG guidelines [39, 48, 104]. However, a related topic, PDF accessibility, has received considerably less attention [32, 49, 93]. A study by Youngblood and Brooks [112] indicated that, on average, nearly 30% of PDFs on the websites of Veterans Affairs Medical Centers lacked accessibility. This finding aligns with observations by Cooper [26], where students frequently reported issues with PDFs not adhering to accessibility standards. Concerns about PDF accessibility among students were also confirmed by Lazar's interviews with eighteen university directors of digital accessibility [62, 63]. This issue extends even to conferences dedicated to accessibility, as evidenced byBrady et al. [22], who identified low accessibility at accessibility-related conferences. Their findings highlight the inherent challenges associated with creating accessible PDFs.

2.2 Challenges in PDF Accessibility

Several factors contribute to the lack of accessibility in PDFs. Zhang et al. [115] point to a scarcity of research focused on accessible authoring tools. This limited research output translates to a shortage of accessible educational resources for users with disabilities [115]. Accessibility features within authoring tools themselves can be restricted, as demonstrated by Rajkumar et al. [85] who identified an intricate user interface (UI) and limited undo options for accessibility features, with accessibility options often restricted to the Pro edition [85]. The actions of content creators also play a role in accessibility issues. For instance, Ahmetovic et al. [12] observed that mathematical formulas in PDFs are inaccessible to screen reader users due to the absence of alternative text descriptions by authors. Similarly, Nganji [79] emphasizes that inaccessible PDFs often fail to adhere to PDF and WCAG 2.0 accessibility standards. These limitations create significant barriers for users with disabilities. Stewart et al. [99] highlighted the challenges faced by users with visual impairments, dyslexia, and motor limitations in accessing online libraries. Bianchetti et al. [20] acknowledges the existence of PDF tagging tools but points out limitations due to unclear specifications and laborious creation processes. A survey by Fichten et al. [40] further confirms these difficulties, with e-learning students reporting struggles with inaccessible course PDFs and a lack of adaptive technologies. Drümmer and Chang [35] emphasize the low number of currently tagged PDFs, further restricting accessibility for users with disabilities. Baule [16] presents data indicating that only 25% of special education cooperatives meet minimum accessibility standards for learners.

2.3 PDF Accessibility Evaluation Tools

In response to these challenges, researchers have proposed various solutions, including software prototypes and accessibility tools. One example is a software prototype developed by Darvishy et al. [30] to evaluate PDF accessibility. This tool identifies existing tags and proposes missing tags based on the ISO 32000-1 accessibility framework, creating a report for accessibility evaluation. Another approach targets the accessibility of mathematical formulas within PDFs. Ahmetovic et al. [12] developed a LaTeX package named "axessibility.sty" that addresses this issue. This package embeds hidden comments with the LaTeX code to allow screen readers and braille displays to accurately interpret mathematical formulas.

Fayyaz et al. [37] created a prototype that renders extracted semantics into a format understandable by Braille displays, specifically targeting the challenge of understanding PDF tables. Darvishy et al. [31] have also explored accessibility plugins for authoring tools, examining a PowerPoint accessibility plugin that offers free accessibility support in multiple languages; positive results from accessibility testing suggest the potential for broader application of such plugins across different authoring tools. Zulfiqar et al. [117] introduced AGAP, an extension for LaTeX that

provides an accessibility mode for PDF creation, which guides users in creating accessible PDFs by warning them of potential accessibility issues during document composition. AGAP integrates features to simplify accessible PDF creation for both sighted and visually impaired creators.

Uckun et al. [102] investigated accessibility challenges faced by screen reader users when filling out PDF forms and identified issues with inaccessible form fields and illogical narration sequences. To address these problems, they recommend converting inaccessible PDF forms into HTML forms, improving the user experience for screen reader users [102]. Their approach also involves examining the PDF to identify logical passages for a more coherent narration sequence. Another tool, AxesPDF, converts Microsoft Word documents to accessible PDFs that meet the PDF/UA-1 standard, automating processes like font embedding and table conversion to ensure accessibility compliance [91] .

Despite ongoing research efforts, a significant number of PDFs remain inaccessible to users with disabilities. There is a need for more comprehensive evaluation tools. This research is motivated by the development of an accessibility evaluation tool that encompasses a wider range of accessibility features, catering to a wider range of user needs.

Table 1. Feature Comparison of PDF Accessibility Evaluation Tools

Tool Name	Year	Header	Footer	Page Number	Link	Font Size	Figure Caption	Table Caption	Dyslexia Friendly	Image Contrast	Page Contrast	Color Blindness
FixRep [51]	2012	✔	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖
ZHAW [28]	2013	✔	✖	✖	✖	✔	✖	✔	✖	✖	✖	✖
PAC 2 [103]	2014	✔	✖	✖	✖	✔	✖	✖	✖	✖	✖	✖
ver-aPDF [70]	2017	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖
AGAP [28]	2020	✔	✖	✖	✖	✖	✖	✔	✖	✖	✖	✖
SciA11y [105]	2021	✔	✔	✖	✖	✖	✔	✔	✖	✖	✖	✖
A11yPDF	2024	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔

2.4 Feature Comparison of PDF Accessibility Evaluation Tools

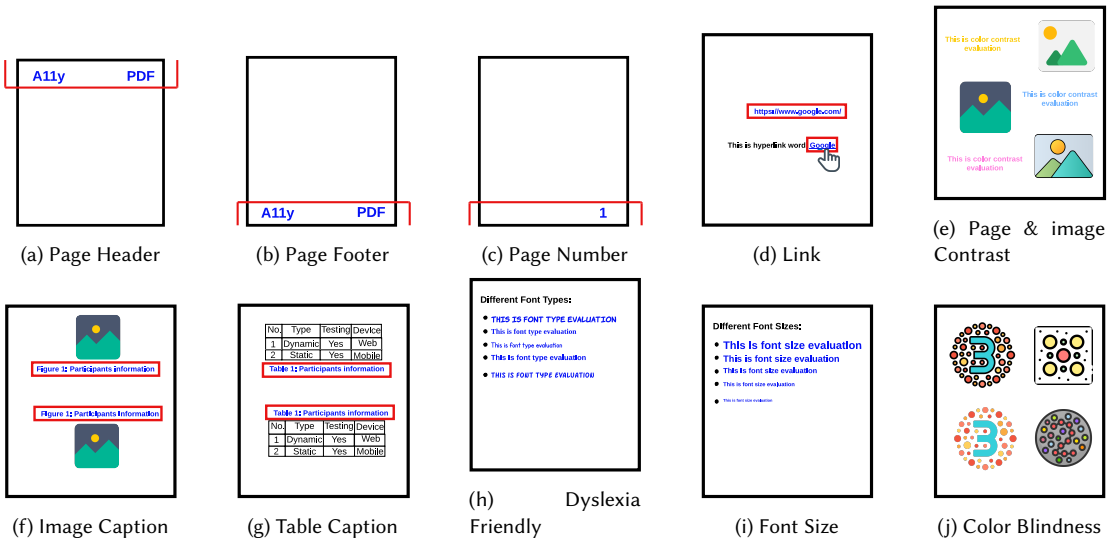
Several software tools have been developed to assess the accessibility of PDF documents, ensuring they are usable by people with disabilities. Table 1 compares these tools based on the specific features they evaluate within a PDF. FixRep [51] captures and stores accessibility information and only evaluates the header of a PDF document, limiting its scope in ensuring overall accessibility. ZHAW Accessibility Plugins [28] provided a comprehensive suite of tools for accessibility checking, offering style suggestions, template creation for accessible PDFs, and real-time violation warnings. They evaluate core elements like headers, font size, and captions. However, HAW Accessibility Plugins likely function as add-ons or extensions for specific PDF editing software limiting its access to wider range of people. Akin to this, PAC 2 [103] is a free, user-friendly tool that promotes PDF accessibility for a wider audience. It focuses on core accessibility features like headers, font size, and link functionality. However, it's important to note that PAC 2 primarily targets technical aspects outlined in accessibility standards and may not evaluate user experience or address complex layouts. ver-aPDF [70] analyzes PDFs against PDF/A standards, which lays a foundation for accessibility but doesn't directly assess accessibility features.

AGAP [28], a LaTeX editor, offers real-time accessibility feedback for visually impaired users through speech and keyboard shortcuts. While it doesn't evaluate the full range of PDF features (as seen in Table 1), it addresses the specific needs of users with vision impairments. SciA11y [105] leverages machine learning to generate accessible HTML from scientific PDFs, specifically focusing on headers, footers, and figure captions. This focus limits its applicability to other document types and potentially overlooks accessibility needs beyond those addressed by the converted HTML format.

This research aims to develop a web application that can evaluate the accessibility of any PDF document. This focus on accessibility extends beyond the core technical aspects to consider the needs of users with specific disabilities. The

application will assess features like headers, footers, page numbers, links, font size, figure and table captions, image and page contrast, all with the goal of ensuring the PDF is accessible to a wider audience, including users with color blindness and dyslexia. By evaluating these elements alongside traditional accessibility checks, the application will provide a more comprehensive assessment of a PDF's usability for individuals with diverse needs.

3 Study Design



(a) Page Header (b) Page Footer (c) Page Number (d) Link (e) Page & image Contrast

(f) Image Caption (g) Table Caption (h) Dyslexia Friendly (i) Font Size (j) Color Blindness

Fig. 1. Illustration of various accessibility considerations for digital content, including (a) Page Header for clear document identification, (b) Page Footer for consistent navigation aids, (c) Page Numbering for easy reference, (d) Links with clear visual indicators, (e) Page and Image Contrast to accommodate visual impairments, (f) Image Captions for content description, (g) Table Captions for enhanced data understanding, (h) Dyslexia-friendly font types for readability, (i) Appropriate Font Sizes for legibility, and (j) Visual adjustments for Color Blindness, ensuring inclusive design for diverse users.

This research utilizes a two-phase survey approach: 1) to explore user awareness of PDF accessibility and 2) to assess the effectiveness of a newly developed evaluation tool. The first phase involves a pre-survey aimed at gauging participants' knowledge and current practices related to PDF accessibility. Motivated by insights from the pre-survey and accessibility guidelines, we developed A11yPDF, a web application specifically designed to address the identified accessibility challenges in PDF documents. A11yPDF is intended to improve the accuracy, efficiency, and user-friendliness of accessibility evaluations. The second phase involves a post-survey administered to participants who used A11yPDF. This post-survey collected feedback on the tool's usability, clarity, and overall effectiveness in enhancing PDF accessibility. The subsequent sections detail the design and implementation of the surveys and the development of A11yPDF.

3.1 Surveys

Table 2 details the questions administered to participants in both the pre- and post-surveys.

3.1.1 Pre-survey. The pre-survey questions targeted participants' current understanding and use of PDF accessibility evaluation tools. The survey initially included 18 number of questions, which were then revised to eliminate redundancy and irrelevant questions. This revision process resulted in 12 Likert-scale questions and 6 closed-ended questions.

Table 2. Set of questions in pre-survey and post-survey.

Survey	Question	Type	RQ
Pre-Survey	Are you aware of the aspects required to make a PDF accessible?	Close-ended	RQ1
Pre-Survey	Do you have familiarity with any of the following accessibility guidelines/laws?	Close-ended	RQ1
Pre-Survey	Do you know where to find accessibility requirements for PDFs?	Close-ended	RQ1
Pre-Survey	Do you consider any accessibility aspect or requirements when you create a PDF?	Likert	RQ2
Pre-Survey	Have you ever used any accessibility evaluation tools for PDF before?	Close-ended	RQ2
Pre-Survey	How important are the following features in a PDF evaluation tool for you?	Likert	RQ3
Post-Survey	In your experience, how accurate were the evaluations provided by the tool?	Likert	RQ4
Post-Survey	Did the A11yPDF simulation tool available on the website help you in evaluating the accessibility of your PDF?	Likert	RQ4
Post-Survey	To what extent do you think the A11yPDF website raised your awareness and its impact on accessibility?	Likert	RQ5
Post-Survey	Indicate if you agree or desagree with the following statement "The A11yPDf website positively influenced my perspective on designing an accessible PDF"	Likert	RQ5
Post-Survey	To what extent has using the website influenced your approach to design and development PDF content?	Likert	RQ5
Post-Survey	After using the website, do you feel better equipped to design or develop accessibility PDF?	Likert	RQ5
Post-Survey	If you have used any accessibility PDF related accessibility tools before, how would you compare the A11yPDF tool with it?	Close-ended	RQ6
Post-Survey	Would you use A11yPDF tool to help you in your future PDF evaluation?	Open-ended	RQ7
Post-Survey	Overall, how satisfied are you with the PDF evaluation tool?	Likert	RQ7
Post-Survey	Rate your overall experience with the A11yPDF website.	Likert	RQ7
Post-Survey	Was the website easy to navigate?	Likert	RQ7
Post-Survey	How clear and easy was it to understand the information provided on the website?	Likert	RQ7
Post-Survey	How did you find the website's content useful in comprehending the accessibility in PDF?	Likert	RQ7
Post-Survey	Please elaborate how your experience with the website has affected your approach to design and develop PDFs.	Open-ended	RQ7

Subsequently, we conducted a pilot study with 12 participants from diverse backgrounds to evaluate the effectiveness and clarity of the survey. The feedback indicated that the initial version was too lengthy and contained some repetition. Based on this information, the number of questions in the survey was reduced from 18 to 10 questions. The final survey was developed using Google Forms, which facilitated data collection.

3.1.2 Post-survey. The post-survey design initially included a set of 22 questions. To ensure clarity and relevance, a revision process was conducted. This process eliminated repetitive, irrelevant, or overly personal questions. A pilot study with 12 participants further informed the survey refinement. Based on the pilot study findings, 6 questions were identified as unnecessary for the intended analysis and were subsequently removed. The final post-survey consisted of 16 questions. The post-survey employed a combination of 11 Likert-scale questions and 5 open-ended questions to gather user feedback on the accessibility and functionality of A11yPDF. This approach aimed to capture both structured responses and unfiltered user insights.

3.1.3 Participants. The survey link was distributed through social media platforms. The initial target number of participants was 100. We received a total of 159 responses, representing a high response rate. Participants came from various backgrounds, with the majority (55.6%) from the technology/IT field. Engineering (19.0%) and education/teaching (14.3%) were also well-represented, with the remaining respondents belonging to diverse fields. The pre-survey and post-survey completion rates differed. A total of 159 participants completed the pre-survey, while 144 participants participated in the post-survey. To ensure data accuracy and reliability, responses with a completion rate below 70% were excluded from the analysis. This resulted in a final sample size of 139 participants.

3.2 Tool Implementation

A11yPDF adheres to the WCAG 2.0 standards to evaluate PDF documents for conformance. Users can directly access A11yPDF within a web browser, eliminating the need for additional software installations. The process entails uploading a PDF document. A11yPDF then analyzes the document for potential accessibility issues based on WCAG 2.0 guidelines.

The user receives a report highlighting these issues, allowing them to make informed decisions about correcting the PDF. If no corrections are desired, the user gains insight into the document's accessibility level.

This section details our methodology for evaluating and analyzing the accessibility of PDF documents based on ten features. Figure 1 visually depicts the comprehensive design and scope of our study. A11yPDF analyzes PDF documents for accessibility across ten key areas. These encompass page contrast, font size, presence of navigational elements (headers, footers, and page numbers), and table/image captions for alternative descriptions. Additionally, the tool assesses features relevant to colorblindness and dyslexia for a more inclusive user experience. Accessibility ratings are computed and assigned to each feature, as discussed in subsequent sections. These ratings culminate in an aggregated score reflecting the overall accessibility of the PDF document. Details of the guidelines implemented for the underlying accessibility features are presented in Table 3. Subsequent sections provide detailed explanations of each element and their significance in enhancing the accessibility of PDF documents.

3.2.1 Page Contrast PDF accessibility guidelines, such as WCAG 2.2, prioritize clear and measurable contrast ratios to guarantee the information is readily accessible to all viewers [3, 4]. Darvishy et al. [29] highlight that creators often lack awareness of insufficient contrast issues, creating accessibility barriers. Many researchers have emphasized the need for an optimum contrast ratio between content and background for inclusive user experience [11, 33, 72, 84, 113, 114]. A11yPDF addresses this need by analyzing contrast ratios within PDFs using image processing techniques. As illustrated in Figure 1e, the tool first converts each page into an image representation. Next, it calculates the luminance-based contrast ratio for each image. These calculated ratios are then compared against the WCAG AA standards, which define minimum thresholds for optimal accessibility.

A minimum of 3:1 contrast is required for non-text content [1]. For text and images containing text, a threshold of 7:1 is mandated for optimal accessibility. Moreover, large-scale text and images containing large-scale text demand a minimum contrast ratio of 4.5:1 [2]. Images meeting or exceeding these thresholds are deemed accessible, while others are flagged for improvement. By quantifying overall PDF image accessibility, this tool pinpoints areas requiring enhancement. This improves user experience for all audiences, regardless of visual abilities, by ensuring equal access to information.

The process begins by converting each PDF page as an image I_i, creating a collection denoted as I_{page}. Then, RGB values of the top-left and bottom-right pixels are extracted, represented as $P_{\text{top-left}}$ and $P_{\text{bottom-right}}$, respectively. Next, the contrast ratio CR_i is computed for each image I_i. This computation involves transitioning the RGB values to an alternative color space to ascertain the luminance values, which are pivotal in determining the contrast ratio. Subsequently, we determine whether the image I_i and its associated page comply with the WCAG AA requirements using the computed contrast ratio CR_i. An image I_i (and its page) is deemed accessible if $CR_i \geq 4.5$ for larger text or $CR_i \geq 3$ for smaller text; otherwise, it is considered not accessible.

$$I_{page} = \text{extract\_images}(PDF) \tag{1}$$

$$\forall I_i \in I_{page} : \tag{2}$$

$$P_{top-left} = \text{RGB values of top-left pixel of } I_i \tag{3}$$

$$P_{bottom-right} = \text{RGB values of bottom-right pixel of } I_i \tag{4}$$

$$CR_i = \text{get\_contrast\_ratio}(P_{top-left}, P_{bottom-right}) \tag{5}$$

$$\text{Accessibility of } I_i = \begin{cases} \text{Accessible}, & \text{if } CR_i > 4.5 \\ \text{Not Accessible}, & \text{otherwise} \end{cases} \tag{6}$$

3.2.2 Font Size Font size plays a crucial role in ensuring document accessibility for individuals with diverse abilities, particularly those with visual impairments [41, 89, 102, 114]. A larger font size enhances readability and legibility, promoting inclusivity and accessibility [53]. Various font sizes can be employed in documents As illustrated in Figure 1i. WCAG standard recommends a minimum font size of 14px for documents to be accessible to a broader range of PDF reader software [5]. This standard ensures the text can be magnified up to 200% without distortion, maintaining readability. A11y tool anlayzes font sizes in PDF documents against a 14px threshold. The tool compiles the frequencies of each unique font size within the document and categorizes them based on their compliance with the threshold. Let B be a row matrix of size $1 \times m$ representing the frequency of each unique font size encountered in the PDF document, where m is the number of unique font sizes.

$$B = \begin{bmatrix} b_1 & b_2 & \cdots & b_m \end{bmatrix} \tag{7}$$

Where b_i represents the count of occurrences of a specific font size in the document. Next, B is partitioned into two matrices $B_{\geq 14}$ and $B_{<14}$ using indicator matrices $I_{\geq 14}$ and $I_{<14}$ of size $1 \times m$ to distinguish between font sizes based on the WCAG accessibility guidelines.

$$I_{\geq 14} = \begin{bmatrix} 1 & 1 & \cdots & 0 \end{bmatrix} \tag{8}$$

$$I_{<14} = \begin{bmatrix} 0 & 0 & \cdots & 1 \end{bmatrix} \tag{9}$$

Where $I_{\geq 14}$ contains 1 at indices corresponding to font sizes greater than or equal to 14px, and $I_{<14}$ contains 1 at indices corresponding to font sizes less than 14. These partitioned matrices are obtained by element-wise multiplication with B.

$$B_{\geq 14} = B \cdot I_{\geq 14} \tag{10}$$

$$B_{<14} = B \cdot I_{<14} \tag{11}$$

Subsequently, the total counts of font occurrences greater than or equal to 14 ($T_{\geq 14}$) and less than 14 ($T_{<14}$) are computed.

$$T_{\geq 14} = \sum_{i=1}^{m} B_{\geq 14}[i] \tag{12}$$

$$T_{<14} = \sum_{i=1}^{m} B_{<14}[i] \tag{13}$$

Then, the overall total count of font size occurrences (T) is computed.

$$T = T_{\geq 14} + T_{<14} \tag{14}$$

Finally, the percentage of font sizes greater or less than 14px are computed.

$$P_{\geq 14} = \left(\frac{T_{\geq 14}}{T} \right) \times 100 \tag{15}$$

$$P_{<14} = \left(\frac{T_{<14}}{T}\right) \times 100 \tag{16}$$

The tool iterates through individual text elements and gathers their font sizes for further processing. This approach exemplifies the potential of A11yPDF in promoting document accessibility by evaluating font size compliance with WCAG guidelines.

3.2.3 Links The inclusion of web-based references, or URLs, plays a significant role in PDFs [65]. However, research suggests that a substantial portion of these references are inaccessible to users with disabilities [66, 109]. Wren's study found a 39% inaccessibility rate among URLs in PDFs [109]. Similarly, Loan et al. reported that over 32% of URLs extracted from scientific articles exhibited accessibility issues [66]. Recognizing this gap, Teixeira da Silva and Nazarovets [100] proposed a tool for evaluating URL accessibility in PDFs, but its scope was limited to scientific publications. A11yPDF addresses this limitation by offering a broader solution. The tool extracts and analyzes web links within PDFs to assess their accessibility for users with disabilities.

Initially, the tool identifies and catalogs all embedded URLs within the PDF. Each extracted URL undergoes a thorough accessibility verification process. This assesses whether the linked web resource is available and functional, ensuring the relevance and utility of the reference. The process culminates in a detailed report that not only lists the total number of URLs but also categorizes them as accessible or inaccessible. This provides a clear understanding of the document's reliance on external resources and their current accessibility status, adhering to WCAG standards [6]. The combined operation can be formalized as follows:

$$\mathbf{S} = \text{Summarize}\left(\text{CheckAccessibility}\left(\text{Extract}(\mathbf{PDF})\right)\right) \tag{17}$$

Initially, the function Extract() is applied to the input **PDF**, which processes the document and generates a Document-URL Binary Matrix, **D**. This matrix indicates the presence or absence of URLs across the document's pages. Subsequently, the function CheckAccessibility() is employed to evaluate each URL identified within **D**. This evaluation yields an Accessibility Status Vector, **A**, where accessible URLs are marked as 1 and inaccessible URLs as 0. Finally, the function Summarize() is utilized to compile insights derived from **D** and **A**. It produces a summary vector or set, **S**, containing key metrics such as the total number of URLs (T), the count of accessible URLs (Y), and the count of inaccessible URLs (N). Through this process, the accessibility status of URLs within the PDF document is effectively summarized for further analysis and interpretation. Furthermore, the summary metrics are calculated as follows:

$$T = \sum_{i,j} D_{ij}, \quad Y = \sum_{j} A_j, \quad N = T - Y \tag{18}$$

Where, T is the total number of URLs detected in the PDF, Y is the number of accessible URLs, and N is the count of inaccessible URLs, calculated by subtracting the number of accessible URLs from the total number of URLs.

This analytic summary serves as a foundation for informed decision-making. Users can identify and prioritize updating or correcting inaccessible links, ultimately enhancing the document's reliability and accessibility for all users.

3.2.4 Header/Footer/Page Number Headers, footers, and page numbers play a pivotal role in enhancing the accessibility of PDF files for people with diverse abilities. Well-structured headers and footers establish navigational consistency enabling assistive technologies like screen readers to maintain user orientation within the document, particularly for those with visual impairments [7]. Additionally, these elements often contain metadata such as chapter titles or dates, providing valuable contextual information for understanding the document's structure. Clear headers and footers also assist screen readers in determining the correct reading order, ensuring a logical content flow and

preventing confusion for users [39]. Furthermore, page numbers act as essential reference points when discussing specific content, assisting individuals who rely on assistive technologies to easily navigate and locate information.

A11yPDF offers functionalities to extract headers, footers, and page numbers from PDF files, adhering to accessibility standard WCAG 2.4.6 [7]. For this purpose, the PDF document is conceptualized as a collection of pages $\mathcal{P}$, where each page p_i comprises a set of text blocks $B(p_i)$. Each text block b_j includes spans $S(b_j)$ characterized by their textual content s_{text} and bounding box s_{bbox}. This conceptualization allows for the application of matrix-like operations to identify headers, footers, and page numbers based on their spatial properties.

Header extraction is formalized through a function

$$\mathcal{H}(p_i) = \{s_{\text{text}} \mid s \in S(b_j) \land b_j \in B(p_i) \land y_0 < \lambda\} \tag{19}$$

Where, (p_i) represents the i^{th} page of a document, b_j is the text block from the set of text blocks $B(p_i)$ on page p_i. Text blocks are the distinct sections or paragraphs on a page, containing one or more spans (s) of text. A span, s, is a continuous segment of text that shares common formatting. Each span s has associated properties, including the text itself (s_{text}) and a bounding box (s_{bbox}). he bounding box of a span is given by $[x_0, y_0, x_1, y_1]$, where (x_0, y_0) are the coordinates of the lower-left corner, and (x_1, y_1) are the coordinates of the upper-right corner of the bounding box. In a 2D Cartesian coordinate system typical for page layouts, y_0 represents the vertical position of the lower-left corner of the text span. Only text spans with the lower-left corner's y-coordinate less than the predefined threshold λ are considered. Finally, $H(p_i)$ filters and extracts text based on spatial criteria. If a header is present, the page is considered accessible. Otherwise, it is counted as inaccessible and contributes to an accessibility percentage reflecting the portion of pages lacking headers.

Footer extraction employs a function $\mathcal{F}$, with a nuanced condition `FooterCond` to identify footer text based on spatial properties and page dimensions $D(p_i) = (\text{width}_i, \text{height}_i)$. Where, $D(p_i)$ represents the dimensions of page p_i, specifically its width and height and $\delta_1, \delta_2, \delta_3, \delta_4$ are the predefined threshold values for determining the position of the footer text. The footer extraction function (Equation 20) leverages the footer condition (Equation 21) to identify and extract footer regions within a document.

$$F(p_i) = \{s_{\text{text}} \mid s \in S(b_j) \land b_j \in B(p_i) \land \mathit{FooterCond}(s_{\text{bbox}}, D(p_i))\} \tag{20}$$

$$\begin{cases} \text{true} & \text{if } y_0 > \text{height}_i - \delta_1 \land x_0 > \delta_2 \\ \text{true} & \text{if } y_0 > \text{height}_i - \delta_3 \land x_0 > \delta_4 \\ \text{false} & \text{otherwise} \end{cases} \tag{21}$$

In this context, y_0 denotes the y-coordinate of the lower-left corner of (s_{bbox}), used for identifying text at the bottom of the page. x_0 represents the x-coordinate of the lower-left corner of (s_{bbox}), employed to filter spans based on horizontal position. height_i signifies the height of page p_i, establishing a vertical cutoff from the bottom of the page for footer text.

This representation uses the bounding box's coordinates, x_0 and y_0, in conjunction with the page's dimensions and predefined thresholds to identify footer text dynamically. Similar to headers, the presence of footers and page numbers are used to calculate the overall accessibility percentage of the document.

For identifying page numbers, a function P examines text spans against specified spatial thresholds to determine potential page number locations, using regular expressions for pattern matching. The criterion for a text span B_{ijk} to be considered a potential page number location is defined as:

$$C_{ijk} = \begin{cases} 1 & \text{if } \Big(B_{ijk}[y_0] > \lambda_1 \wedge B_{ijk}[x_0] > \lambda_2\Big) \vee \\ & \Big(B_{ijk}[y_0] > \lambda_3 \wedge B_{ijk}[x_0] > \lambda_4\Big) \vee \\ & \Big(B_{ijk}[y_0] > \lambda_5\Big) \vee \Big(B_{ijk}[x_0] > \lambda_1 \wedge B_{ijk}[y_0] > \lambda_2\Big), \\ 0 & \text{otherwise.} \end{cases}$$

This definition incorporates multiple conditions to accurately identify page numbers based on their spatial characteristics.

Where:

- $s_{\text{bbox}}[0]$ refers to the x coordinate of the left boundary of the bounding box, indicating horizontal placement.
- $\delta_1, \delta_2, \delta_3, \delta_4$ are threshold values specific to the document's layout, designed to capture text positioned in the expected footer area based on empirical observations or document standards.
- λ is a threshold denoting the maximum allowable distance from the top of the page for text spans to be classified as headers.
- s_{bbox}: The bounding box of text span s, defined by coordinates $[x_0, y_0, x_1, y_1]$, where (x_0, y_0) is the lower-left corner, and (x_1, y_1) is the upper-right corner.
- The selection of λ is crucial, as it should reflect the typical vertical spacing of headers within the document's layout.

This strategy is flexible across different document layouts and formats, offering a robust solution for automated document management. Our framework focuses on the extraction of headers and footers from PDF documents, highlighting the importance of spatial characteristics in automating text extraction processes. This approach delivers a systematic way to segregate distinct text elements, improving both the precision and speed of data extraction from PDF files.

3.2.5 Table/Image Caption A11yPDF assesses the accessibility of captions within PDF documents, specifically focusing on tables and images. This evaluation adheres to the WCAG 2.0 standards, particularly Success Criterion 1.3.1 [10]. This criterion mandates that all non-text content, such as images and tables, must have equivalent text alternatives (captions) available, ensuring users who rely on assistive technologies like screen readers can access the information they convey.

Captions play a crucial role in ensuring PDF accessibility for users with visual impairments who rely on screen readers to access information [25, 73]. These textual alternatives provide a description of the content within tables and images, allowing screen readers to convey the information effectively [95]. The absence of captions can be frustrating for visually impaired users, as screen readers cannot interpret the content of tables or images without this additional information, hindering their ability to grasp the full context [38, 64]. A study by Fayyaz et al. [37] further highlights that accessibility considerations for tables within PDFs were often neglected. Under these considerations, A11yPDF leverages spatial analysis based on bounding box coordinates to evaluate the proximity between tables/figures and their corresponding captions.

The spatial relationships between tables and their captions, as well as figures and their captions, are evaluated using specific proximity functions within a document layout. For tables, denoted as T_j, and potential captions, denoted as C_k, the proximity function $P(S_{T_j}, S_{C_k})$ is defined as follows:

$$P(S_{T_j}, S_{C_k}) = \begin{cases} 1 & \text{if } S_{T_j}^{\text{top}} - S_{C_k}^{\text{bottom}} < \theta_{\text{top}} \\ & \text{or } S_{T_j}^{\text{bottom}} - S_{C_k}^{\text{top}} < \theta_{\text{bottom}} \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

Here, θ_{top} and θ_{bottom} are predefined thresholds determining the acceptable vertical distance for a caption to be considered related to a table, either above or below it.

Similarly, for figures (F_j) and their captions (C_k), the proximity function $P(S_{F_j}, S_{C_k})$ is employed, considering both horizontal and vertical thresholds:

$$P(S_{F_j}, S_{C_k}) = \begin{cases} 1 & \text{if } S_{F_j}^{\text{right}} - S_{C_k}^{\text{left}} < \theta_{\text{right}} \\ & \text{and } S_{F_j}^{\text{top}} - S_{C_k}^{\text{bottom}} < \theta_{\text{vertical}} \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

Where θ_{right} and θ_{vertical} denote the horizontal and vertical distances within which a caption is considered associated with a figure.

S_{T_j}, S_{C_k}, and S_{F_j} denote the spatial regions or bounding boxes of tables, captions, and figures, respectively. The subscripts T_j, C_k, and F_j specify the indexed positions of the tables, captions, and figures, respectively. $S_{T_j}^{\text{top}}$, $S_{T_j}^{\text{bottom}}$, $S_{C_k}^{\text{top}}$, $S_{C_k}^{\text{bottom}}$, $S_{F_j}^{\text{right}}$, and $S_{F_j}^{\text{left}}$ represent top and bottom vertical coordinates for tables and captions, and right and left horizontal coordinates for figures.

3.2.6 Dyslexia Friendly The A11yPDF tool assesses the accessibility of PDF documents for users with dyslexia. Dyslexia, a cognitive impairment, presents challenges in accurate and fluent word recognition [77, 94]. Improving PDF accessibility for dyslexic people is motivated by two main factors: (1) the prevalence of dyslexia is believed to be between 5% and 10% of the world's population [45], and (2) implementing accessibility features not only benefits users with dyslexia but also extends to individuals with other reading challenges [67]. Research suggests that people with dyslexia process visual information differently than those without the condition [13, 58]. This difference necessitates a focus on document typography during accessibility evaluation [42, 87]. One commonly explored aspect in this context is the use of fonts without serifs, also known as sans-serif fonts [44, 61, 80, 107, 108]. These fonts are characterized by the absence of small decorative strokes at the ends of letterforms, potentially making them easier to distinguish for individuals with dyslexia [19, 88, 90]. Due to the potential benefits for individuals with dyslexia, we have chosen to prioritize sans-serif fonts such as Arial, Comic Sans, Verdana, Tahoma, Century Gothic, Trebuchet, Calibri, and Open Sans in A11yPDF. This selection aligns with the WCAG 3.1.5 success criterion on reading, which emphasizes the importance of ensuring text content is presented in a way that is perceivable and understandable [9].

To assess the font type, we establish a systematic computational framework. We begin by defining a set of fonts under scrutiny, where each f_i denotes a distinct font.

$$\mathcal{F} = f_1, f_2, \ldots, f_n \tag{24}$$

Correspondingly, we define a collection of readable fonts delineated by specific criteria, such as sans serif attributes (e.g., Arial, Comic Sans, etc.).

$$\mathcal{R} = r_1, r_2, \ldots, r_m \tag{25}$$

$$R \subseteq \mathcal{F} \times \mathcal{R} \tag{26}$$

The foundation of our assessment lies in the establishment of a binary relation, wherein $(f_i, r_j) \in R$ if font f_i bears resemblance or conforms to readable font r_j based on predetermined criteria, which may include characteristics like sans serif property and letter spacing.

For representation, we construct an $n \times m$ adjacency matrix A derived from R.

$$A_{ij} = \begin{cases} 1 & \text{if } (f_i, r_j) \in R \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

Additionally, we introduce a usage vector v of dimensions $n \times 1$, with each v_i representing the usage count of font f_i. In the analysis phase, we compute

$$v_{\text{match}} = A \cdot v \tag{28}$$

resulting in an $m \times 1$ vector. Here, each element of v_{match} encapsulates the cumulative usage of fonts that match each readable font r_j.

$$s = \frac{1}{|v|_1} A^T \cdot v \tag{29}$$

where $|v|_1$ denotes the L1 norm of v, representing the total usage of all fonts. This computation furnishes a vector s indicative of the normalized usage scores of readable fonts relative to the total usage. Finally, the overall readability is quantified which offers a percentage-based assessment of readability grounded in the predefined criteria.

$$P = 100 \cdot |s|_1 \tag{30}$$

3.2.7 Colorblindness Friendly Color plays a vital role in digital content, but 10% of the global population experiences colorblindness, hindering their ability to interpret certain colors [23, 24, 27, 56, 81, 83]. Several tools exist for the evaluation of the accessibility of digital content for individuals with colorblindness, [46, 57, 57, 71, 74, 116] however, none currently address PDFs. This creates a gap in evaluating PDF accessibility for individuals with color blindness. A11yPDF addresses this research gap by evaluating PDF accessibility for colorblindness. Our tool employs k-means clustering and incorporates accessibility thresholds for different types of color blindness, including protanopia [43], deuteranopia [98], and tritanopia [52] against WCAG success criterion 1.4.1 [8]. The developed tool can identify the type of colorblindness a PDF might pose challenges for and pinpoint the specific pages with accessibility issues. This evaluation would allow content creators to identify and address potential barriers, ensuring more inclusive and accessible PDFs.

The initial step involves separating each page p_i within the PDF into individual image files I_i. These images are then converted to the RGB color space, ensuring consistent color representation across all pages. This conversion process results in a set of matrices C_i, where each entry represents the RGB values of a pixel in the corresponding page image.

For each page image I_i, the k-means clustering algorithm is applied with a predefined number of clusters K. This technique groups pixels with similar RGB values into clusters, effectively reducing the overall number of unique colors present in the image. The selection of the optimal value for K can be informed by the desired level of detail preservation and computational efficiency considerations. By simplifying the color palette, this step aims to capture essential visual information while reducing the complexity of the data.

Next, the luminance L_{ij} of each pixel c_{ij} in the processed image I_i is calculated using a weighted sum of its RGB components.

$$L_{ij} = 0.2126 * R_{ij} + 0.7152 * G_{ij} + 0.0722 * B_{ij} \tag{31}$$

. Luminance represents the perceived brightness of a color and plays a crucial role in analyzing visual contrast and readability within the document.

Then, minimum contrast ratio thresholds CR^t_{th} are established for each type of color blindness t, such as Protanopia, Deuteranopia, and Tritanopia. These thresholds are derived from WCAG accessibility guidelines [8] and ensure sufficient color contrast for individuals with specific color vision deficiencies.

Furthermore, for each page p_i, the tool iterates through all pairs of adjacent pixels c_{ij} and c_{kl} are iterated. Subsequently, contrast ratio CR_{ijkl} is calculated between each pixel pair using their respective luminance values.

$$CR_{ijkl} = (max(L_{ij}, L_{kl}) + 0.05)/(min(L_{ij}, L_{kl}) + 0.05) \tag{32}$$

Afterward, the average contrast ratio ACR_i is computed for the entire page by averaging the contrast ratios of all pixel pairs. Finally, each page undergoes evaluation against the pre-defined contrast ratio thresholds for different color blindness types.

$$p_i = \begin{cases} \text{Accessible}^t_i = \text{True} & \text{if } ACR_i \geq CR^t_{th} \\ \text{Accessible}^t_i = \text{False} & \text{otherwise} \end{cases} \tag{33}$$

3.2.8 Text Summarization Text summarization has been explored as a potential method for improving the accessibility of written content for people with diverse abilities [15, 50, 75]. A study by Balasuriya et al. [14] found that participants with disabilities found summarized articles easier to read. This has motivated the development of text summarization tools using text mining techniques. These techniques include weighted term frequency-inverse document frequency [15], cluster-based topic modeling [18], Latent Dirichlet Allocation [60], genetic algorithms [55], k-means [86], Latent semantic Analysis (LSA) [17, 47, 68]. Research by Siddiqui et al. [96] suggests that LSA-based summaries may be more accurate compared to other extractive techniques, supporting its potential application in PDF text summarization.

A11yPDF employs LSA for text summarization. The process commences by extracting text from the target PDF file P. This file contains individual pages P_i, where i indicates the page number. Each page's text is extracted and combined to form the complete extracted text X. Subsequently, tokenization is conducted to split the extracted text X into individual meaningful units, such as words or phrases, resulting in T. Finally, the LSA summarizer utilizes the tokenized text T and a predetermined number of desired summary sentences k as inputs. The output is the summarized text, represented as S.

$$\mathbf{X} = \sum_{i=1}^{n} \mathbf{P}_i \tag{34}$$

$$\mathbf{T} = \text{Tokenize}(\mathbf{X}) \tag{35}$$

$$\mathbf{S} = \mathbf{LSA}(\mathbf{T}, k) \tag{36}$$

A11yPDF generates concise summaries of information within PDF documents, aiming to enhance accessibility for individuals with various reading needs.

3.2.9 Overall Accessibility Score A11yPDF calculates an overall accessibility score to provide a quantitative assessment of a PDF document's accessibility for various user groups. This score incorporates the individual metric scores derived from the different features analyzed by the tool. This score integrates the individual metric scores obtained from the evaluation of various accessibility features within the document. The overall accessibility score $Score_{accessibility}$ is computed as:

Table 3. Mapping of WCAG Success Criteria to PDF Elements: A breakdown of accessibility principles, success levels, and guidelines associated with specific PDF components to ensure document accessibility.

No.	Principle	Success	Level	Guideline	PDF Element
1	Operable	2.4.9	AAA	Link Purpose (Link Only)	link
2	Perceivable	1.4.11	AA	Non-text Contrast	Contrast
3	Operable	2.4.6	AA	Headings and Labels	Header
4	Operable	2.4.6	AA	Headings and Labels	Footer
5	Operable	2.4.6	AA	Headings and Labels	Page Number
6	Perceivable	1.4.4	AA	Resize Text	Font Size
7	Perceivable	1.3.1	AA	Info and Relationships	Image Caption
8	Perceivable	1.3.1	AA	Info and Relationships	Table Caption
9	Understandable	3.1.5	AAA	Reading Level	Dyslexia
10	Perceivable	1.4.1	A	Use of Color	Color Blindness

$$Score_{accessibility} = \frac{1}{N} \sum_{i=1}^{N} w_i \cdot m_i \tag{37}$$

where, N is the total number of accessibility metrics evaluated, w_i is the Weight assigned to the $i-th$ accessibility metric, and m_i is the score obtained for the $i-th$ accessibility metric. This formula essentially calculates the weighted average of the individual metric scores m_i, where each score is weighted according to its relative significance w_i in contributing to the overall accessibility of the document.

For the accessibility and content quality of PDF, we define a comprehensive set of criteria encompassing both structural and content aspects. Each criterion is quantitatively assessed through a percentage score, p_i, reflecting the extent to which the document meets specific accessibility or content standards. The overall performance, $P_{\text{overall, rounded}}$, is then calculated as the arithmetic mean of these scores, providing a holistic measure of the document's quality and accessibility.

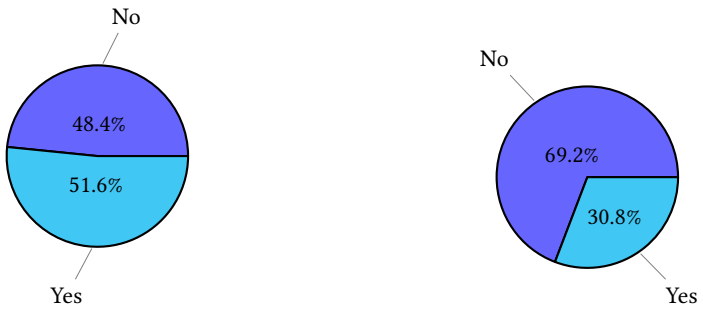
4 Survey Results

The results of a two-phase survey study are presented in this section.

4.1 RQ1: What is the level of user awareness regarding PDF accessibility challenges?

Objective. With this research question, we aimed to assess users' understanding and awareness of aspects required to make PDF accessible and accessibility guidelines. The results reveal significant insights into the current state of user awareness and highlight areas where knowledge and practice are lacking.

Results. The majority of the participants (51.6%) reported being aware of the aspects required to make a PDF accessible, as shown in Figure 2 (a). However, this also reveals that nearly half (48.4%) of the users lack essential knowledge about PDF accessibility. Moreover, Figure 2 (b) underscores a significant knowledge gap, with 69.2% of participants admitting they do not know where to find accessibility requirements for PDF documents. Further analysis, presented in Figure 3 shows that a majority (58.5%) of the participants indicated that they were unfamiliar with any of the listed guidelines or laws, highlighting a widespread lack of knowledge in this area. Among those aware of specific guidelines, the Web Content Accessibility Guidelines (WCAG) were the most recognized, with 32.7% of participants indicating familiarity. While this is encouraging, given WCAG's critical role in ensuring web content accessibility, the low awareness of other important guidelines, such as Accessible Rich Internet Applications (ARIA) and Inclusive Design Principles (each recognized by only 11.9% of participants), is concerning. Additionally, there was almost negligible recognition of Section 508 (0.6%) and Sugamya Bharat IS-17802 (0.6%).



(a) Awareness of PDF Accessibility Requirements

(b) Knowledge of Accessibility Requirements Location

Fig. 2. (a) Are you aware of the aspects required to make a PDF accessible? (b) Do you know where to find accessibility requirements for PDFs?

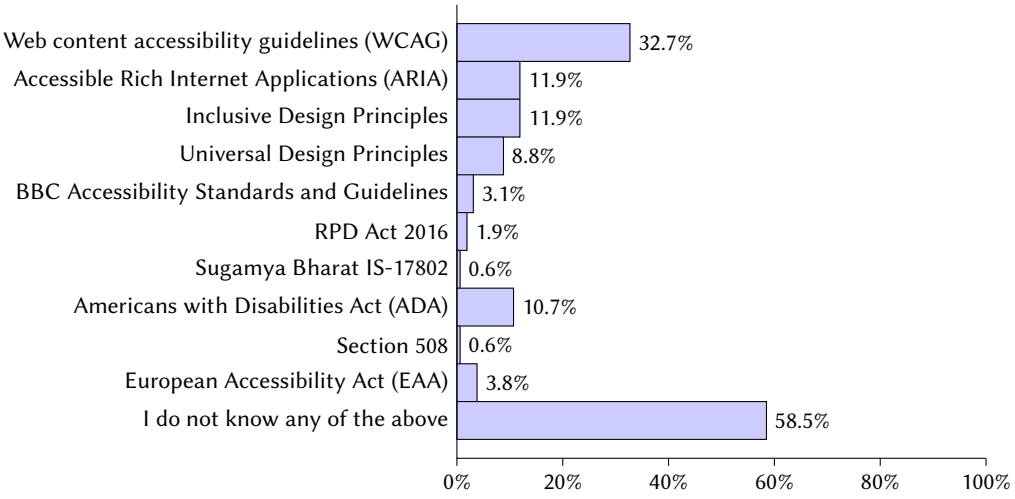


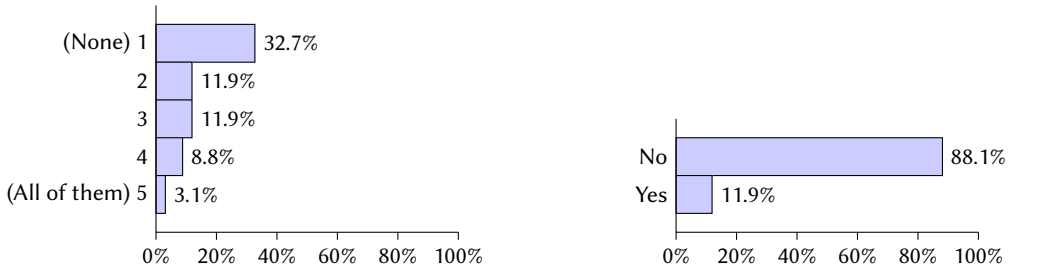
Fig. 3. Do you have familiarity with any of the following accessibility guidelines/laws?

Discussion. These results underscore significant gaps in user awareness and knowledge regarding PDF accessibility. Despite a slight majority being aware of the need for PDF accessibility, a substantial portion of users lack the necessary knowledge to locate and apply accessibility requirements effectively. This suggests that current educational and informational resources on PDF accessibility are insufficient or not reaching the intended audience. Moreover, the widespread unfamiliarity with critical accessibility guidelines, as revealed in the analysis, points to a broader issue of inadequate dissemination of information. While the recognition of WCAG is somewhat reassuring, the low awareness of other essential guidelines, such as ARIA and Section 508, indicates that many users may not be fully equipped to ensure the accessibility of their PDF documents. This lack of knowledge not only affects the accessibility of individual documents but also contributes to broader systemic barriers for individuals with disabilities who rely on accessible digital content.

4.2 RQ2: What is the extent to which users adopt accessibility requirements for PDF in practice?

Objective. The second research question aims to analyze the extent to which participants incorporate accessibility requirements when creating PDF documents. To explore this, we asked participants two key questions: (a) "Do you consider any accessibility aspects or requirements when you create a PDF?" and (b) "Have you ever used any accessibility evaluation tools for PDF before?" Additionally, we sought to identify which specific tools participants had used if they had any experience with accessibility evaluation.

Results. The survey results indicate a significant gap in the adoption of accessibility requirements during the PDF creation process. Figure 4(a) shows that a substantial majority of participants reported that they do not consider any accessibility aspects when creating PDFs, with 32.7% selecting '1' on the Likert scale (indicating no consideration). The data also reveals a sharp decline in the percentages as the Likert scale progresses from 'None' (1) to 'All of them' (5). Only 11.9% of participants indicated moderate consideration (selecting '2' or '3'), and even fewer selected higher levels (8.8% for '4' and 3.1% for '5'). Additionally, the results show that an overwhelming majority (88.9%) of participants have never used any accessibility evaluation tools for PDFs, as depicted in Figure 4(b). Among the small percentage (11.9%) who have used accessibility evaluation tools, WAVE and WCAG were the most commonly cited, followed by Microsoft Office Accessibility Checker, Adobe Acrobat Pro Accessibility Checker, and Acrobat Pro DC.



(a) PDF Accessibility Requirement Consideration (b) Accessibility Evaluation Tool Usage

Fig. 4. (a) Do you consider any accessibility aspect or requirements when you create a PDF? (b) Have you ever used any accessibility evaluation tools for PDF before?

Discussion. The results of RQ2 indicate a significant discrepancy between the perceived importance of accessibility requirements in PDF creation and their implementation. This finding aligns with previous researches [62, 63, 112]. Despite some level of awareness about accessibility, the majority of participants do not integrate these requirements into their workflow. This suggests that many PDF documents may not meet the necessary accessibility standards, potentially excluding users with disabilities from accessing important information [26]. The minimal use of accessibility evaluation tools further exacerbates this issue. Without these tools, document creators may lack the necessary resources to properly assess and address accessibility issues in their PDFs. There is a clear need for encouraging the use of accessibility evaluation tools through more intuitive interfaces and comprehensive user guides which can enhance the accessibility of PDF documents.

4.3 RQ3: What are the most important accessibility requirements according to users?

Objective. In this research question, we sought to identify the accessibility requirements that participants consider most important for making a PDF accessible. To achieve this, participants were asked to rate the importance of various

PDF features, including header and footer clarity, page number visibility, page contrast, colorblindness-friendly design, font size, dyslexia-friendly fonts, image captions, table captions, and URL links.

Results. The survey results, as illustrated in Figure 5, reveal that the majority of participants rated these features as either "Very Important" or "Important," highlighting a general consensus on the critical nature of these elements in ensuring PDF accessibility.

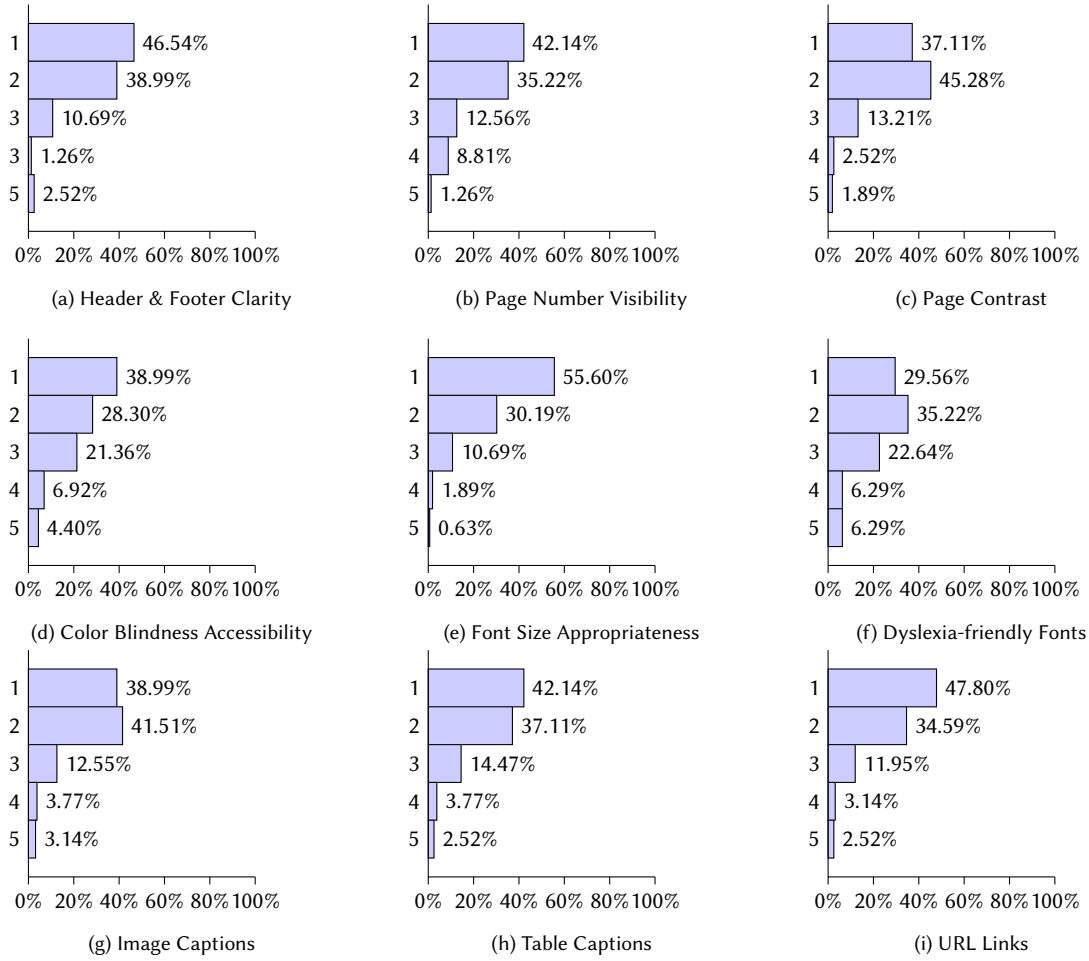


Fig. 5. How important are the following features in a PDF evaluation tool for you? (Very Important:1, Important:2, Moderately Important:3, Slightly Important:4, Not Important:5)

A significant portion of participants (46.54%) rated header and footer clarity as "Very Important," with an additional 38.99% rating it as "Important" as shown in Figure 5 (a). Page number visibility was similarly prioritized as evidenced by Figure 5 (b), with 42.14% of participants marking it as "Very Important" and 35.22% as "Important." Figure 5 (c) shows that appropriate page contrast emerged as a highly valued feature, with 45.28% of participants considering it "Important" and 37.11% rating it as "Very Important." In addition, when it comes to designing for colorblind users, 38.99% of participants viewed colorblind-friendly design as "Very Important," and 28.30% deemed it "Important." This response

as shown in Figure 5 (d), reflects an understanding of the need for color choices that are inclusive of users with color vision deficiencies, ensuring that information conveyed through color is accessible to all. Figure 5 (e) showcased that the font size was rated as "Very Important" by 55.60% of participants which is the highest rating among all features, indicating a strong emphasis on text readability. Additionally, dyslexia-friendly fonts were recognized as important by 35.22% of participants, with 29.56% rating them as "Very Important" as evidenced by Figure 5 (f). Image captions (Figure 5 (g)) and table captions (Figure 5 (h)) were also highly rated, with 38.99% and 42.14% of participants, respectively, marking these features as "Very Important." The importance of URL links was also emphasized as presented by the participants responses in Figure 5 (i). Major portion (47.80%) of participants considered them "Very Important" and 34.59% rated them as "Important."

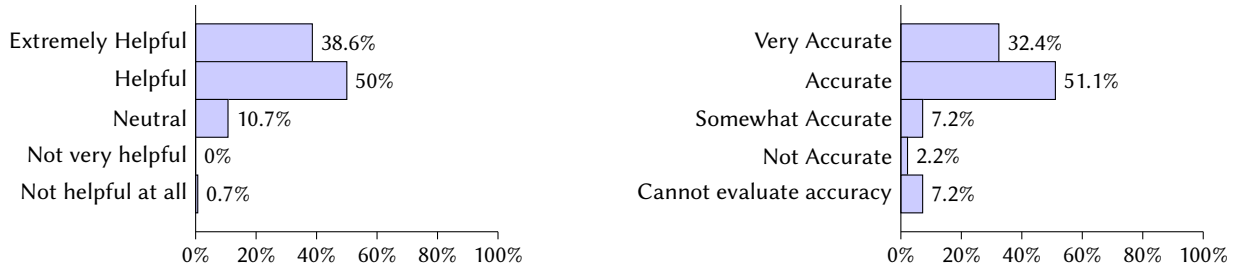
Discussion. The results from RQ3 indicate a strong consensus among users regarding the importance of various accessibility features in PDFs. The high ratings for features such as font size, page contrast, and URL links highlight the recognition of these elements as critical for creating accessible digital documents. Users' emphasis on font size and dyslexia-friendly fonts suggests a heightened awareness of the challenges faced by individuals with reading disabilities, and the importance of designing documents that accommodate these needs. Furthermore, the prioritization of header and footer clarity, page number visibility, and image/table captions points to a broader understanding of the role that document structure and descriptive content play in accessibility. These elements are essential not only for users with visual impairments but also for those relying on assistive technologies, underscoring the need for documents that are both well-structured and richly annotated. Notably, the focus on colorblind-friendly design and page contrast further reflects an awareness of the visual accessibility challenges that some users face. This suggests that users are cognizant of the need for inclusive design practices that ensure readability and comprehension for all users, regardless of visual ability.

4.4 RQ4: To what extent do users perceive the A11yPDF tool as helpful in evaluating PDF accessibility?

Objective. After the development and implementation of the A11yPDF tool, we wanted to evaluate its effectiveness in assessing PDF accessibility. To achieve this, participants were asked to provide feedback on two key aspects: (a) the overall helpfulness of the A11yPDF tool in evaluating the accessibility of their PDFs, and (b) the perceived accuracy of the evaluations provided by the tool.

Results. The survey results demonstrate a strong positive reception of the A11yPDF tool among participants. As illustrated in Figure 6(a), 50% of participants rated the tool as "Helpful" and 38.6% rated it as "Extremely Helpful" in evaluating the accessibility of their PDFs. Only 10.7% of participants were neutral, and a mere 0.7% found the tool "Not Helpful at All," indicating the tool's general effectiveness. Moreover, regarding the perceived accuracy of the A11yPDF tool's evaluations, the results in Figure 6(b) further affirm the tool's effectiveness. A substantial majority of participants rated the evaluations as either "Accurate" (51.1%) or "Very Accurate" (32.4%). This indicates a high level of trust among users in the tool's assessments. A smaller portion of participants found the tool to be "Somewhat Accurate" (7.2%) or "Not Accurate" (2.2%), while 7.2% of participants were unable to evaluate the accuracy of the tool's assessments.

Discussion The overwhelmingly positive feedback on the A11yPDF tool underscores its significant impact on improving PDF accessibility. The high percentage of users who found the tool helpful suggests that A11yPDF addresses a critical need for accessible and user-friendly tools in the domain of PDF accessibility evaluation. The strong ratings for both helpfulness and accuracy highlight the tool's reliability and effectiveness, which are crucial for users who may rely on these assessments to make their documents accessible to a broader audience. The accuracy ratings, in particular, suggest that users perceive A11yPDF as a trustworthy resource for evaluating PDF accessibility. This is a critical finding,



(a) Helpfulness of A11yPDF in Accessibility Evaluation

(b) A11yPDF Evaluation Accuracy

Fig. 6. (a) Did the A11yPDF simulation tool available on the website help you in evaluating the accessibility of your PDF? (b) In your experience, how accurate were the evaluations provided by the tool?

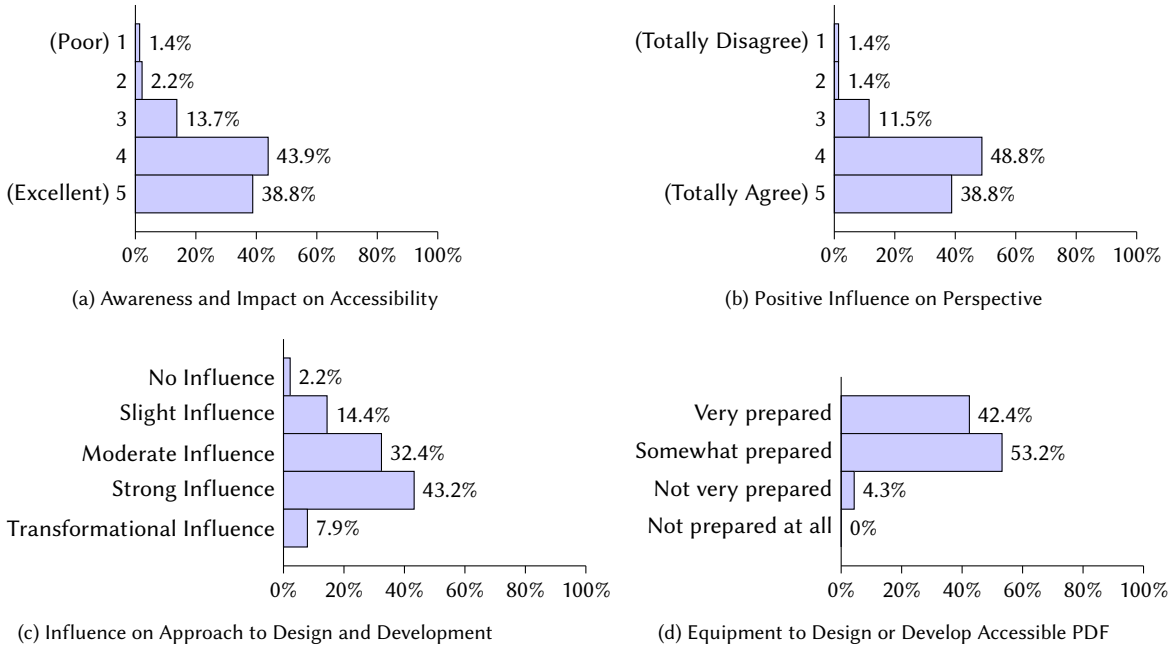
as it aligns with existing literature emphasizing the importance of accurate and reliable tools in accessibility evaluations [91]. Moreover, the minimal negative feedback regarding the tool's helpfulness and accuracy indicates that A11yPDF has successfully minimized common usability issues that often plague accessibility tools. This is particularly important in the context of accessibility, where the precision and ease of use are paramount. The success of A11yPDF in these areas suggests that it could serve as a model for future developments in accessibility evaluation software.

4.5 RQ5: Does the use of the A11yPDF tool have a measurable impact on user awareness of PDF accessibility?

Objective. This research question aims to evaluate the impact of the A11yPDF tool on user awareness of PDF accessibility. To assess this, participants were asked a series of questions after using the tool, focusing on how it influenced their understanding of accessibility, their approach to designing accessible PDFs, and their confidence in applying these principles.

Results. Figure 7(a) shows that the A11yPDF tool significantly impacted user awareness of PDF accessibility. A majority of participants rated the tool's impact on their awareness as either "4" (43.9%) or "5" (Excellent, 38.8%). Only a small fraction of participants (1.4%) rated the impact of tool as "1" (Poor). Further evidence of the positive influence of the tool is shown in Figure 7(b), where 48.8% of participants agreed and 38.8% totally agreed that A11yPDF positively influenced their perspective on designing accessible PDFs. The minimal disagreement (1.4% totally disagreed) was also observed among the participants. Furthermore, Figure 7(c) shows that most of the participants (43.2%) reported that A11yPDF had a "Strong Influence" on their design practices, and 32.4% noted a "Moderate Influence." Notably, 7.9% of participants indicated that the tool had a "Transformational Influence" on their approach and only 2.2% reported "No Influence." Regarding preparedness, Figure 7(d) shows that after using A11yPDF, a vast majority of participants felt better equipped to design or develop accessible PDFs, with 42.4% feeling "Very Prepared" and 53.2% feeling "Somewhat Prepared." Only 4.3% felt "Not Very Prepared," and no participants reported feeling "Not Prepared at All".

Discussion. The results demonstrate that A11yPDF not only enhances technical ability of the users to evaluate PDF accessibility but also significantly raises their awareness and shifts their attitudes toward accessible design. The high ratings for the tool's impact on accessibility awareness and design perspectives suggest that A11yPDF effectively addresses a critical gap in accessibility education, providing users with the knowledge and confidence needed to prioritize accessibility in their work. Moreover, the strong influence of A11yPDF on design and development practices is particularly noteworthy. The fact that a substantial percentage of participants reported a "Transformational Influence" indicates that A11yPDF has the potential to fundamentally change how users approach PDF creation, aligning with



(a) Awareness and Impact on Accessibility

(b) Positive Influence on Perspective

(c) Influence on Approach to Design and Development

(d) Equipment to Design or Develop Accessible PDF

Fig. 7. (a) To what extent do you think the A11yPDF website raised your awareness and its impact on accessibility? (b) Indicate if you agree or desagree with the following statement "The A11yPDf website positively influenced my perspective on designing an accessible PDF". (c) To what extent has using the website influenced your approach to design and development PDF content? (d) After using the website, do you feel better equipped to design or develop accessibility PDF?

literature that emphasizes the importance of tools that not only educate but also inspire lasting change in behavior. In addition. the significant boost in preparedness reported by users underscores A11yPDF's role as an effective educational resource. This aligns with studies highlighting the need for practical tools that can bridge the gap between accessibility knowledge and its application in real-world contexts [12, 91, 117].

4.6 RQ6: How does our tool compare to other tools that evaluate PDF accessibility?

Objective. This research question sought to evaluate the significance of A11yPDF in real-world settings by comparing it to other PDF accessibility evaluation tools that participants had previously used. Specifically, we aimed to understand how A11yPDF performs in practice relative to existing tools, based on user experiences.

Results. Figure 8 demonstrate a favorable reception of A11yPDF among participants with prior experience using other PDF accessibility evaluation tools. The survey results show that 20.1% of the participants rated A11yPDF as "much better," and 15.8% rated it as "somewhat better" than the tools they had previously used. Additionally, 10.1% of participants found A11yPDF to be similar to other tools they had used. However, only a small fraction of participants rated A11yPDF as "somewhat inferior" (1.4%) or "much inferior" (1.4%), suggesting minimal dissatisfaction among those familiar with other tools. Notably, a significant majority of participants (51.1%) reported that they had not used any PDF accessibility evaluation tools before A11yPDF.

Discussion. The results indicate that A11yPDF is perceived as a competitive and, in many cases, superior tool for evaluating PDF accessibility when compared to other tools. The fact that a significant portion of participants rated

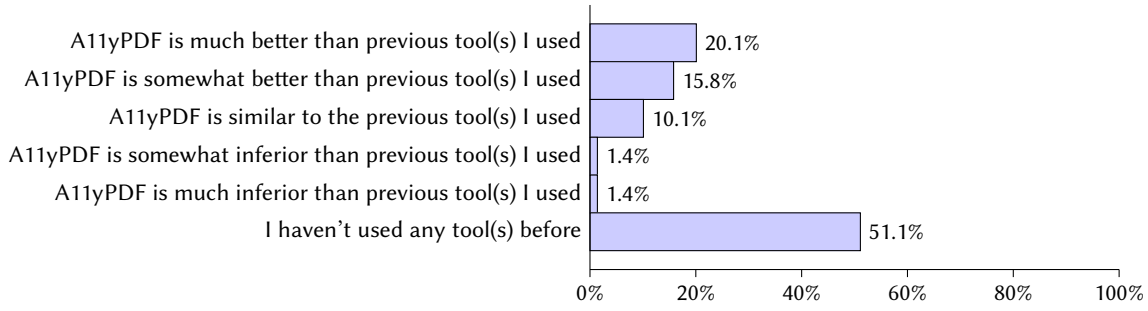


Fig. 8. If you have used any accessibility PDF related accessibility tools before, how would you compare the A11yPDF tool with it?

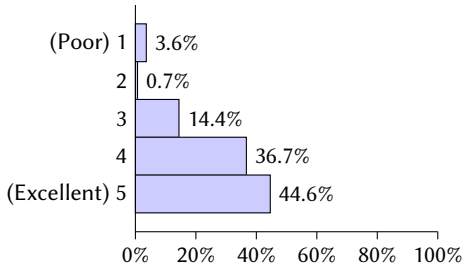
A11yPDF as better than their previous tools suggests that it offers enhanced features or a more user-friendly experience. This aligns with literature emphasizing the importance of usability and functionality in accessibility tools, where superior user experience can drive adoption and satisfaction [111]. We found that over half of the participants had not used any other accessibility tools before A11yPDF. It suggests that A11yPDF is filling a critical void for users who are either new to accessibility evaluation or have not previously engaged with such tools. This finding is significant in the context of accessibility tool adoption, as it highlights the tool's ability to attract and serve a new user base, potentially broadening the scope of accessibility practices across different user demographics. Moreover, the strong performance of the tool in comparison to other tools supports the idea that A11yPDF is well-positioned to become a preferred solution in the field. The favorable comparisons made by experienced users reinforce the notion that A11yPDF is not only competitive but may also offer advantages that other tools lack, such as improved usability, more comprehensive evaluations, or more intuitive interfaces.

4.7 RQ7: Is A11yPDF a functional tool for evaluating PDF accessibility?

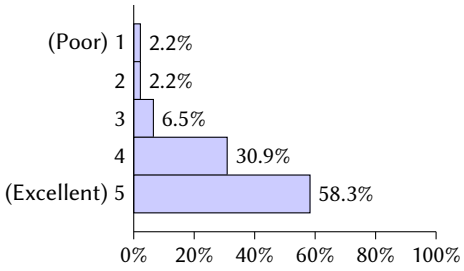
Objective. This research question aimed to evaluate the overall functionality and user experience of the A11yPDF tool, focusing on key aspects such as ease of navigation, clarity of content, usefulness in understanding PDF accessibility, and overall user satisfaction. In addition to quantitative data, participants were also invited to provide open-ended feedback about their experiences with A11yPDF.

Results. The survey results indicate a highly positive reception of the A11yPDF tool, with the majority of participants rating their overall experience favorably. Figure 9(a) shows that 44.6% of participants rated their experience as "Excellent," and 36.7% gave it a rating of "4" on a 5-point scale. Only a small percentage (3.6%) rated their experience as "Poor." In terms of navigation, Figure 9(b) shows that 58.3% of participants found the website "Excellent" in ease of navigation, and 30.9% rated it as "4." The clarity of the content provided on the A11yPDF website was also rated highly by participants, as depicted in Figure 9(c). A significant portion of users (38.8%) rated the clarity as "Excellent," with an additional 43.9% rating it as "4." In addition, the usefulness of A11yPDF in enhancing users' understanding of PDF accessibility is further emphasized in Figure 9(d). Nearly half of the participants (49.6%) rated the tool's usefulness as "Excellent," with 38.8% rating it as "4." Finally, overall user satisfaction with A11yPDF was notably high, as shown in Figure 9(e). A combined 90.7% of participants reported being either "Very Satisfied" (43.2%) or "Satisfied" (47.5%) with the tool, while no participants reported being "Dissatisfied" or "Very Dissatisfied."

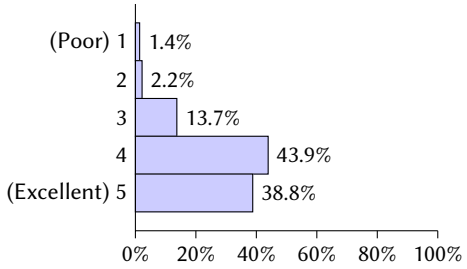
Discussion. The results underscore the effectiveness of A11yPDF in providing a user-friendly, informative, and highly satisfactory experience for users seeking to evaluate and improve the accessibility of their PDFs. The tool's high



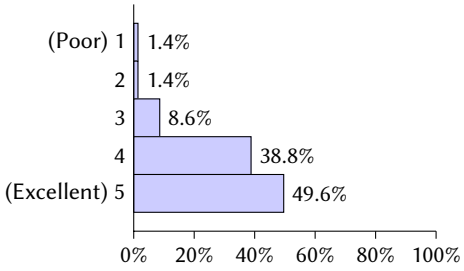
(a) Overall User Experience of A11yPDF



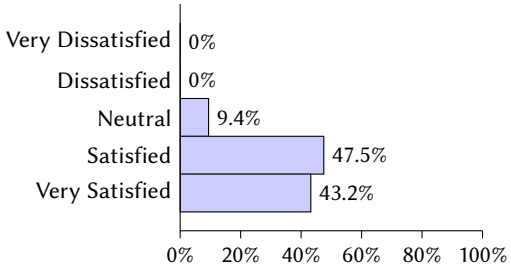
(b) Ease in Navigating the Website



(c) Clarity in Understanding Website Content



(d) Usefulness of A11yPDF in PDF accessibility Understanding



(e) User Satisfaction with A11yPDF

Fig. 9. (a) Rate your overall experience with the A11yPDF website. (b) Was the website easy to navigate? (c) How clear and easy was it to understand the information provided on the website? (d) - How did you find the website's content useful in comprehending the accessibility in PDF? (e) Overall, how satisfied are you with the PDF evaluation tool?

ratings for overall user experience and ease of navigation indicate that A11yPDF excels in delivering an intuitive and accessible interface. Users specifically praised the simplicity and clarity of the site's design, with one user noting, *"User interface is good and website is easy to use as well."* Another user highlighted the ease with which they could navigate the tool, commenting on how the *"easy-to-use UI made it easy to upload and receive evaluations,"* further affirming the importance of usability in digital tools.

The clarity of the content is another area where A11yPDF has been particularly successful. Users appreciated the clear and accessible presentation of information, which helped them understand complex accessibility concepts. One participant remarked, *"I was not familiar with PDF accessibility previously, but now I am a bit familiar with it,"* while another shared, *"When I found this website, I understood where I made the mistake in making my PDFs accessible, and the detailed report it generated was very useful to me."* These comments highlight A11yPDF's effectiveness not only as a functional tool but also as an educational resource that empowers users to improve their digital content by making accessibility concepts easier to grasp.

The tool's usefulness in enhancing users' understanding of PDF accessibility was also highly valued, with feedback indicating that it served as a critical learning tool. One user shared, *"It was very useful. This website usage was new for me, so I will definitely try to learn new things like accessibility,"* suggesting that A11yPDF is playing a significant role in educating users who are new to the field of accessibility.

Overall user satisfaction was high, with users expressing contentment with the tool's performance and its impact on their work. The fact that no participants reported dissatisfaction indicates that A11yPDF meets the needs of its users effectively, making it a reliable and valued resource in the domain of PDF accessibility evaluation.

5 Takeaways

This section outlines the takeaway of this study.

Takeaway 1. PDF remains a prevalent file format for document sharing due to its ability to maintain layout and formatting. However, a lack of awareness among creators regarding PDF accessibility standards can hinder the usability of these documents for individuals with disabilities [32, 49, 93]. Implementing accessibility best practices, such as utilizing WCAG guidelines [39, 48, 104] is essential for creating inclusive PDFs [115].

Takeaway 2. PDF creators should put a strong emphasis on elements like font size [91], contrast, alternative text [12], and clear navigation aids [102] in their design, as these features are fundamental to ensuring accessibility for a diverse range of users, including those with visual impairments, cognitive disabilities, and other challenges. By prioritizing these elements, PDF creators can significantly enhance the readability and usability of their documents, thereby promoting inclusivity and compliance with accessibility standards. This approach not only meets the immediate needs of users with disabilities but also contributes to the broader goal of creating universally accessible digital content, which is increasingly recognized as a critical component of responsible and ethical content creation in today's digital landscape.

Takeaway 3. While raising awareness of PDF accessibility standards is crucial, implementing processes for evaluating both existing PDFs and future content creation is equally important. There are fewer tools available for assessing PDF accessibility compared to web content, and some existing tools may not be fully accessible, require purchase, or have complex interfaces that exclude users with disabilities. Therefore, it is essential to develop user-friendly, accessible, and ideally free evaluation tools to remove barriers for creators and streamline the process of creating accessible PDFs.

Takeaway 4. Traditional PDFs often lack features that cater to diverse learning needs. For example, individuals with color blindness might struggle with documents that rely heavily on color coding for information [101]. Similarly, users with dyslexia might face challenges due to poorly formatted text or lack of alternative text descriptions for images [19]. Individuals with visual impairments, motor skill limitations, and other disabilities also require specific accessibility considerations. Current accessibility checking tools for PDFs often have limitations. It's crucial to expand the scope of these tools to encompass a wider range of disabilities. This would allow creators to identify and address accessibility issues that might hinder comprehension for various user groups.

Takeaway 5. The low adoption rate of accessibility guidelines [79] and the minimal use of evaluation tools [85] suggest that many PDF documents are likely being produced without adequate consideration for users with disabilities. To address this issue, it is essential to promote greater awareness of accessibility requirements among PDF creators and encourage the use of accessibility evaluation tools [12, 30, 31, 37]. A11yPDF has the potential to play a pivotal role in bridging this gap. It offers intuitive, easy-to-use evaluations and guides users through the accessibility process. By doing so, our tool can increase both the adoption of accessibility practices and the use of evaluation tools.

Takeaway 6. A11yPDF is a powerful tool not only for evaluating PDF accessibility but also for fostering a deeper understanding of inclusive design principles. By providing actionable accessibility evaluations, A11yPDF can significantly raise awareness, influence design practices, and boost preparedness. This aligns with emerging trends in accessibility education that emphasize the importance of tools that both educate and inspire lasting change in user behavior [59, 82]. To further maximize its impact, A11yPDF could be integrated into broader educational initiatives, such as workshops and online courses, thereby reaching a wider audience and fostering a more widespread commitment to digital accessibility.

Takeaway 7. A11yPDF's success as a user-centered and effective accessibility evaluation tool underscores the paramount importance of these factors in developing such tools. By prioritizing ease of use and feature-richness, tool developers can significantly enhance user adoption and satisfaction. This is particularly crucial as the demand for accessible digital content continues to grow, ensuring that accessibility tools are both powerful and user-friendly to drive widespread implementation of best practices.

Takeaway 8. The overall positive user experience with A11yPDF highlights the significance of integrating detailed and clear feedback mechanisms in accessibility tools. Providing users with understandable and actionable evaluations can greatly enhance their ability to create accessible content, ultimately contributing to a more inclusive digital environment.

Takeaway 9. The ability of A11yPDF to educate users on accessibility best practices through its clear and detailed feedback mechanisms reflects the growing need for tools that not only perform evaluations but also serve as educational resources. Such tools can bridge the knowledge gap in accessibility, especially for those who are new to the field, ensuring that more content creators are equipped to produce accessible PDFs.

Takeaway 10. A11yPDF's ability to raise awareness and influence design practices suggests that accessibility tools should be designed with the dual purpose of evaluation and education. By integrating these functions, such tools can play a significant role in transforming how digital content is created, making it more accessible to all users.

Takeaway 11. Ensuring accessible educational materials is critical for fostering inclusive learning environments, especially in institutions catering to students with disabilities. To achieve this, educational institutions can prioritize the creation and distribution of accessible PDF materials by establishing accessibility as an institutional focus. This might involve raising awareness and providing resources at all levels of the organization. Workshops, training programs, or readily available resources on PDF accessibility standards and tools could equip educators and administrators with the necessary knowledge. Furthermore, integrating accessibility best practices into pre-service educator preparation programs offers a long-term solution. Equipping future teachers and administrators with the knowledge and skills to develop accessible PDFs from the outset is crucial. This training could encompass understanding PDF accessibility criteria, utilizing tools and resources for creating accessible PDFs, and evaluating the accessibility of existing materials.

6 Threats to Validity

This section details the threats to the validity of this study.

Self-Reported Data Bias. We acknowledge that our study can have a threat of reliance on user-reported data, which may introduce biases such as overestimation of compliance with accessibility guidelines or misunderstanding of specific accessibility features. To mitigate this, we employed a mixed-methods approach by combining quantitative survey data with qualitative insights from user interactions with A11yPDF. This approach allowed us to validate the self-reported data against observable user behaviors and tool usage patterns, providing a more robust understanding of user practices and challenges in PDF accessibility. Additionally, the post-survey analysis helped in identifying discrepancies between user perceptions and actual tool performance, which we addressed through iterative improvements to A11yPDF's feedback mechanisms, enhancing its educational impact on accessibility standards.

Sample Bias. The diversity of the participant sample was predominantly composed of individuals from technology and IT backgrounds, which limits the generalizability of the study findings across broader user demographics, such as individuals from non-technical fields who may have different levels of familiarity with accessibility tools and guidelines. To address this, we involved teachers in different fields than technology and engaged to have more diverse participants. By broadening the participant base, we captured a broader range of perspectives and challenges, ultimately refining A11yPDF to better cater to the needs of all users, regardless of their technical expertise.

Limited Scope of Document Types. A potential threat is that the study primarily focused on evaluating static PDF documents, which may not fully represent the accessibility challenges encountered in dynamic or interactive PDFs that include elements like forms, multimedia, or interactive content. To mitigate this, our plan for future iterations of A11yPDF will incorporate features that assess and provide feedback on these interactive elements. By expanding the tool's capabilities to include dynamic content, we aim to provide a more comprehensive evaluation of PDF accessibility, ensuring that all document types, including those with interactive features, are accessible to users with diverse needs

7 Conclusion

In light of the results obtained from analyzing our quantitative data, we observe that awareness and practices related to PDF accessibility vary significantly among users, underscoring the importance of tools that can effectively bridge these gaps. A11yPDF was developed in response to these challenges, integrating features that go beyond those found in existing accessibility tools. The tool provides a comprehensive evaluation of PDF documents, focusing on a wide range of accessibility elements essential for creating inclusive content. The post-survey results demonstrate that A11yPDF significantly enhanced users' understanding of accessibility requirements and positively influenced their approach to designing accessible PDFs. Moroever, the user-friendly interface of A11yPDF and detailed feedback were particularly well-received, with many users reporting a better grasp of where their documents fell short and how to make improvements. A11yPDF's ability to deliver actionable insights and support users in creating more accessible documents marks a meaningful contribution to the field of digital accessibility.

References

[1] [n. d.]. G145: Ensuring that a contrast ratio of at least 3:1 exists between text (and images of text) and background behind the text | WAI | W3C — w3.org. https://www.w3.org/WAI/WCAG22/Techniques/general/G145.html. [Accessed 20-02-2024].
[2] [n. d.]. G18: Ensuring that a contrast ratio of at least 4.5:1 exists between text (and images of text) and background behind the text | WAI | W3C — w3.org. https://www.w3.org/WAI/WCAG22/Techniques/general/G18.html. [Accessed 20-02-2024].
[3] [n. d.]. How to Meet WCAG (Quick Reference). https://www.w3.org/WAI/WCAG22/Understanding/non-text-contrast.html. [Accessed 20-02-2024].
[4] [n. d.]. How to Meet WCAG (Quick Reference). https://www.w3.org/WAI/WCAG22/Understanding/contrast-enhanced.html. [Accessed 20-02-2024].
[5] [n. d.]. How to Meet WCAG (Quickref Reference) — w3.org. https://www.w3.org/WAI/WCAG22/Understanding/resize-text.html. [Accessed 20-02-2024].
[6] [n. d.]. How to Meet WCAG (Quickref Reference) — w3.org. https://www.w3.org/WAI/WCAG22/Understanding/link-purpose-in-context.html. [Accessed 20-02-2024].

[7] [n. d.]. PDF14: Providing running headers and footers in PDF documents | WAI | W3C — w3.org. https://www.w3.org/WAI/WCAG21/Techniques/pdf/PDF14.html. [Accessed 28-02-2024].

[8] [n. d.]. Understanding Success Criterion 1.4.1: Use of Color | WAI | W3C — w3.org. https://www.w3.org/WAI/WCAG21/Understanding/use-of-color.html#adaptable. [Accessed 03-03-2024].

[9] [n. d.]. Understanding Success Criterion 3.1.5 | Understanding WCAG 2.0 — w3.org. https://www.w3.org/TR/UNDERSTANDING-WCAG20/meaning-supplements.html. [Accessed 02-03-2024].

[10] [n. d.]. Web Content Accessibility Guidelines (WCAG) 2.2 — w3.org. https://www.w3.org/TR/WCAG22/#info-and-relationships. [Accessed 02-03-2024].

[11] Patricia Acosta-Vargas, Mario Gonzalez, Maria Rosa Zambrano, Ana Medina, Noah Zweig, and Luis Salvador-Ullauri. 2020. The portable document format: an analysis of PDF accessibility. In *Advances in Human Factors and Systems Interaction: Proceedings of the AHFE 2020 Virtual Conference on Human Factors and Systems Interaction, July 16-20, 2020, USA*. Springer, 206–214.

[12] Dragan Ahmetovic, Tiziana Armano, Cristian Bernareggi, Anna Capietto, Sandro Coriasco, Nadir Murru, et al. 2019. Axessibility 2.0: creating tagged PDF documents with accessible formulae. *Ars Texnica* 28 (2019), 138–145.

[13] Moncayo Arias Mónica Alexandra, Bastidas Vera Eduardo Alberto, Cabezas Macías Paola María, Ledesma Espín Carmen del Roció, Bayas Guevara Bolívar Israel, Saúl Rogelio Jiménez Bajaña, Marianita De Jesús Izquierdo Boza, and María Luisa Bazán Guzmán. 2024. Innovative and Inclusive Digital Applications to Enhance Literacy in Students with Dyslexia. *Migration Letters* 21, S1 (2024), 362–378.

[14] Saminda Sundeepa Balasuriya, Laurianne Sitbon, Jinglan Zhang, and Khairi Anuar. 2021. Summary and Prejudice: Online Reading Preferences of Users with Intellectual Disability. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. 285–289.

[15] Shakila Basheer, M Anbarasi, Darpan Garg Sakshi, and V Vinoth Kumar. 2020. Efficient text summarization method for blind people using text mining techniques. *International Journal of Speech Technology* 23 (2020), 713–725.

[16] Steven M Baule. 2020. Evaluating the accessibility of special education cooperative websites for individuals with disabilities. *TechTrends* 64, 1 (2020), 50–56.

[17] Ramesh Chandra Belwal, Sawan Rai, and Atul Gupta. 2021. Text summarization using topic-based vector space model and semantic measure. *Information Processing & Management* 58, 3 (2021), 102536.

[18] Ramesh Chandra Belwal, Sawan Rai, and Atul Gupta. 2023. Extractive text summarization using clustering-based topic modeling. *Soft Computing* 27, 7 (2023), 3965–3982.

[19] Gerd Berget and Siri Fagernes. 2021. Reading experiences and reading efficiency among adults with dyslexia: An accessibility study. In *International Conference on Human-Computer Interaction*. Springer, 221–240.

[20] Roberto Bianchetti, Markus Erle, and Samuel Hofer. 2012. Mainstreaming the creation of accessible PDF documents by a rule-based transformation from word to PDF. In *Computers Helping People with Special Needs: 13th International Conference, ICCHP 2012, Linz, Austria, July 11-13, 2012, Proceedings, Part I 13*. Springer, 595–601.

[21] Jeffrey P Bigham, Erin L Brady, Cole Gleason, Anhong Guo, and David A Shamma. 2016. An uninteresting tour through why our research papers aren't accessible. In *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*. 621–631.

[22] Erin Brady, Yu Zhong, and Jeffrey P Bigham. 2015. Creating accessible PDFs for conference proceedings. In *Proceedings of the 12th International Web for All Conference*. 1–4.

[23] Alex Chaparro, C FIII Stromeyer, EP Huang, RE Kronauer, and Rhea T Eskew. 1993. Colour is what the eye sees best. *Nature* 361, 6410 (1993), 348–350.

[24] Yu-Chieh Chen and Tai-Shan Liao. 2011. Hardware digital color enhancement for color vision deficiencies. *ETRI Journal* 33, 1 (2011), 71–77.

[25] Amanda Coolidge, Sue Doner, Tara Robertson, and Josie Gray. 2018. *Accessibility toolkit*. BCcampus.

[26] J Cooper. 2019. How to respond to a school website accessibility complaint from OCR. CampusSuite.

[27] Michael Crabb, Michael Heron, Rhianne Jones, Mike Armstrong, Hayley Reid, and Amy Wilson. 2019. Developing accessible services: Understanding current knowledge and areas for future support. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[28] Alireza Darvishy and Hans-Peter Hutter. 2013. Comparison of the effectiveness of different accessibility plugins based on important accessibility criteria. In *Universal Access in Human-Computer Interaction. Applications and Services for Quality of Life: 7th International Conference, UAHCI 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013, Proceedings, Part III 7*. Springer, 305–310.

[29] Alireza Darvishy, Hans-Peter Hutter, Alexander Horvath, and Martin Dorigo. 2010. A flexible software architecture concept for the creation of accessible PDF documents. In *Computers Helping People with Special Needs: 12th International Conference, ICCHP 2010, Vienna, Austria, July 14-16, 2010. Proceedings 12*. Springer, 47–52.

[30] Alireza Darvishy, Hans-Peter Hutter, and Oliver Mannhart. 2011. Web application for analysis, manipulation and generation of accessible PDF documents. In *Universal Access in Human-Computer Interaction. Applications and Services: 6th International Conference, UAHCI 2011, Held as Part of HCI International 2011, Orlando, FL, USA, July 9-14, 2011, Proceedings, Part IV 6*. Springer, 121–128.

[31] Alireza Darvishy, Thomas Leemann, and Hans-Peter Hutter. 2012. Two software plugins for the creation of fully accessible PDF documents based on a flexible software architecture. In *Computers Helping People with Special Needs: 13th International Conference, ICCHP 2012, Linz, Austria, July 11-13, 2012, Proceedings, Part I 13*. Springer, 621–624.

[32] Heather Devine, Andres Gonzalez, and Matthew Hardy. 2011. Making accessible PDF documents. In *Proceedings of the 11th ACM symposium on Document engineering*. 275–276.

[33] Luchin Doblies, David Stolz, Alireza Darvishy, and Hans-Peter Hutter. 2014. PAVE: A web application to identify and correct accessibility problems in PDF documents. In *Computers Helping People with Special Needs: 14th International Conference, ICCHP 2014, Paris, France, July 9-11, 2014, Proceedings, Part I 14*. Springer, 185–192.

[34] Olaf Drümmer. 2012. PDF/UA (ISO 14289-1)–applying WCAG 2.0 principles to the world of PDF documents. In *Computers Helping People with Special Needs: 13th International Conference, ICCHP 2012, Linz, Austria, July 11-13, 2012, Proceedings, Part I 13*. Springer, 587–594.

[35] Olaf Drümmer and Bettina Chang. [n. d.]. PDF/UAin. ([n. d.]).

[36] Kerry A Falloon. 2020. Effectively Evaluating the Accessibility of Electronic Monographs Using VPATs and Other Resources at the College of Staten Island Library-CUNY. *Serials Review* 46, 2 (2020), 98–113.

[37] Nosheen Fayyaz, Shah Khusro, et al. 2023. Enhancing Accessibility for the Blind and Visually Impaired: Presenting Semantic Information in PDF Tables. *Journal of King Saud University-Computer and Information Sciences* (2023), 101617.

[38] Nosheen Fayyaz, Shah Khusro, and Shakir Ullah. 2021. Accessibility of tables in pdf documents. *Information Technology and Libraries* 40, 3 (2021).

[39] Katrina Fernandes, Sarika Paramananthan, Lynn Cockburn, and Julius Nganji. 2023. Readily available but how accessible?: An analysis of the web accessibility of healthcare-related resources. *Journal of Accessibility and Design for All* 13, 2 (2023), 188–215.

[40] Catherine S Fichten, Vittoria Ferraro, Jennison V Asuncion, Caroline Chwojka, Maria Barile, Mai N Nguyen, Ryan Klomp, and Joan Wolforth. 2009. Disabilities and e-learning problems and solutions: An exploratory study. *Journal of Educational Technology & Society* 12, 4 (2009), 241–256.

[41] Maja FILIPOVSKA, Goran AJDINSKI, and Aleksandra KAROVSKA RISTOVSKA. 2023. DYSLEXIA TYPEFACE: DOES IT AFFECT READING FLUENCY? *Prizren Social Science Journal* 7, 1 (2023), 46–52.

[42] Leon Franzen, Zoey Stark, and Aaron P Johnson. 2021. Individuals with dyslexia use a different visual sampling strategy to read text. *Scientific reports* 11, 1 (2021), 6449.

[43] Ayaka Fujita, Mashiho Mukaida, Tadahiro Azetsu, and Noriaki Suetake. 2023. Lightness Modification in RGB Color Space for Protanopia and Deuteranopia. In *IECON 2023-49th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 1–6.

[44] Jessica Galliussi, Luciano Perondi, Giuseppe Chia, Walter Gerbino, and Paolo Bernardis. 2020. Inter-letter spacing, inter-word spacing, and font with dyslexia-friendly features: testing text readability in people with and without dyslexia. *Annals of dyslexia* 70, 1 (2020), 141–152.

[45] Beverly Gilbert, Jonathan Stubblefield, Jake Qualls, Xiuzhen Huang, Arianne Pait, Karen Yanowitz, Allen Hays, Eli Richmond, Leslie Parker, and Tamra Washington. 2023. Dyslexia and AI: The Use of Artificial Intelligence to Identify and Create Font to Improve Reading Ability of Individuals with Dyslexia. In *Society for Information Technology & Teacher Education International Conference*. Association for the Advancement of Computing in Education (AACE), 856–865.

[46] Francisco Gomez-Donoso, Felix Escalona, Bessie Dominguez-Dager, Monica Pina-Navarro, Francisco Morillas-Espejo, and Miguel Cazorla. 2023. A SOFTWARE FOR IMPROVED ACCESSIBILITY OF VISUAL CONTENT IN CLASS. In *INTED2023 Proceedings*. IATED, 504–509.

[47] Hritvik Gupta and Mayank Patel. 2021. Method of text summarization using LSA and sentence based topic modelling with Bert. In *2021 international conference on artificial intelligence and smart systems (ICAIS)*. IEEE, 511–517.

[48] Benaris Hajduk and Ismail Ali. 2023. Study of accessibility needs of people with visual impairments and how they align with WCAG 2.0 guidelines: A Study of WCAG 2.0 Guidelines.

[49] Shawn Lawton Henry. 2012. Developing text customisation functionality requirements of PDF reader and other user agents. In *Computers Helping People with Special Needs: 13th International Conference, ICCHP 2012, Linz, Austria, July 11-13, 2012, Proceedings, Part I 13*. Springer, 602–609.

[50] Hendrik Heuer and Elena Leah Glassman. 2023. Accessible Text Tools: Where They Are Needed & What They Should Look Like. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–7.

[51] Andrew Hewson and Emma Tonkin. 2012. Supporting PDF accessibility evaluation: early results from the FixRep project. *New Trends In Qualitative And Quantitative Methods In Libraries: Selected Papers Presented at the 2nd Qualitative and Quantitative Methods in Libraries* (2012), 417–424.

[52] Chelsea Hosein. [n. d.]. Review article on colour vision defect. ([n. d.]).

[53] Guanhua Hou and Ying Hu. 2023. Designing combinations of pictogram and text size for icons: effects of text size, pictogram size, and familiarity on older adults' visual search performance. *Human Factors* 65, 8 (2023), 1577–1595.

[54] Crescentia Jung, Shubham Mehta, Atharva Kulkarni, Yuhang Zhao, and Yea-Seul Kim. 2021. Communicating visualizations without visuals: Investigation of visualization alternative text for people with visual impairments. *IEEE transactions on visualization and computer graphics* 28, 1 (2021), 1095–1105.

[55] AA Karcioglu and AC Yasa. 2020. Automatic Summary Extraction in Texts Using Genetic Algorithms [Genetik Algoritmalar Kullanilarak Metinlerde Otomatik Ozet Cikarma]. (2020).

[56] Carol Kaufman-Scarborough. 2001. Accessible advertising for visually-disabled persons: the case of color-deficient consumers. *Journal of Consumer Marketing* 18, 4 (2001), 303–318.

[57] Takatomo Kodera. 2023. Accessibility-friendly approach for responsive web design–Perspectives for User experience and User interface. (2023).

[58] Arni Kristjansson and Heida Maria Sigurdardottir. 2023. The role of visual factors in dyslexia. *Journal of Cognition* 6, 1 (2023).

[59] Serhat Kurt. 2019. Moving toward a universally accessible web: Web accessibility and education. *Assistive Technology* (2019).

[60] Anil Kus and Cigdem Inan Aci. 2023. Performance Evaluation of the Extractive Methods in Automatic Text Summarization Using Medical Papers. *Gazi Journal of Engineering Sciences (GJES)/Gazi Mühendislik Bilimleri Dergisi* 9, 4 (2023).

[61] Sanne M Kuster, Marjolijn van Weerdenburg, Marjolein Gompel, and Anna MT Bosman. 2018. Dyslexie font does not benefit reading in children with or without dyslexia. *Annals of dyslexia* 68 (2018), 25–42.

[62] Jonathan Lazar. 2021. Managing digital accessibility at universities during the COVID-19 pandemic. *Universal Access in the Information Society* (2021), 1–17.

[63] Jonathan Lazar. 2022. Managing digital accessibility at universities during the COVID-19 pandemic. *Universal Access in the Information Society* (2022), 1–17.

[64] 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.

[65] Fayaz Ahmad Loan, Aasif Mohammad Khan, Syed Aasif Ahmad Andrabi, Sozia Rashid Sozia, and Umer Yousuf Parray. 2023. Giving life to dead: role of WayBack Machine in recovery of dead URLs. *Data Technologies and Applications* (2023).

[66] Fayaz Ahmad Loan and Ufaira Yaseen Shah. 2020. The decay and persistence of web references. *Digital Library Perspectives* 36, 2 (2020), 157–166.

[67] Elham Madjidi and Christopher Crick. 2023. Enhancing Textual Accessibility for Readers with Dyslexia through Transfer Learning. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–5.

[68] Shrabanti Mandal and Girish Kumar Singh. 2020. LSA based text summarization. *Int J Recent Technol Eng (IRTE)* 9, 2 (2020), 150–156.

[69] Alicia Mason and Sakshi Bhati. 2023. Ready. gov: Who's ready, really? Examining principles of inclusivity and universal design in emergency management and disaster preparedness public information websites. *Journal of Emergency Management and Disaster Communications* (2023).

[70] Rebecca McGuinness, Carl Wilson, Duff Johnson, and Boris Doubrov. 2017. veraPDF: open source PDF/A validation through pragmatic partnership.. In *iPRES*.

[71] S Meenakshi and Anshu Singla. 2023. A novel heuristic method for quantitative assessment of web accessibility for colorblind. *Universal Access in the Information Society* (2023), 1–21.

[72] Frank Mittelbach and Chris Rowley. 2020. LATEX Tagged PDF—A blueprint for a large project. *TUG-boat* 41, 3 (2020), 292–298.

[73] Heather Moorefield-Lang. 2019. Accessibility in online course design. *Library Technology Reports* 55, 4 (2019), 14–16.

[74] Martez E Mott, John Tang, and Edward Cutrell. 2023. Accessibility of Profile Pictures: Alt Text and Beyond to Express Identity Online. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–13.

[75] Kumaresh Nandhini and SR Balasundaram. 2011. Improving readability of dyslexic learners through document summarization. In *2011 IEEE International Conference on Technology for Education*. IEEE, 246–249.

[76] Azadeh Nazemi. 2016. *Non-visual representation of complex documents for use in digital talking books*. Ph. D. Dissertation. Curtin University.

[77] Trude Nergård-Nilssen and Charles Hulme. 2014. Developmental dyslexia in adults: Behavioural manifestations and cognitive correlates. *Dyslexia* 20, 3 (2014), 191–207.

[78] Julius T Nganji. 2015. The Portable Document Format (PDF) accessibility practice of four journal publishers. *Library & Information Science Research* 37, 3 (2015), 254–262.

[79] Julius T Nganji. 2018. An assessment of the accessibility of PDF versions of selected journal articles published in a WCAG 2.0 era (2014–2018). *Learned Publishing* 31, 4 (2018), 391–401.

[80] Aleena Gertrudes Niklaus, Tianyuan Cai, Zoya Bylinskii, and Shaun Wallace. 2023. Digital Reading Rulers: Evaluating Inclusively Designed Rulers for Readers With Dyslexia and Without. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[81] Isa Maria Paiva, Sean Siqueira, and Simone Bacellar Leal Ferreira. 2021. The Windows 10's Color Filter Feature as an Aid for Color Blind People in the Use of Websites. In *Proceedings of the XX Brazilian Symposium on Human Factors in Computing Systems*. 1–11.

[82] Afra Pascual, Mireia Ribera, and Toni Granollers. 2012. Perception of accessibility errors to raise awareness among web 2.0 users. In *Proceedings of the 13th International Conference on Interacción Persona-Ordenador*. 1–2.

[83] Rohan Patel, Pedro Breton, Catherine M Baker, Yasmine N El-Glaly, and Kristen Shinohara. 2020. Why software is not accessible: Technology professionals' perspectives and challenges. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*. 1–9.

[84] Debashish Pradhan, Tripti Rajput, Aravind Jembu Rajkumar, Jonathan Lazar, Rajiv Jain, Vlad I Morariu, and Varun Manjunatha. 2022. Development and Evaluation of a Tool for Assisting Content Creators in Making PDF Files More Accessible. *ACM Transactions on Accessible Computing (TACCESS)* 15, 1 (2022), 1–52.

[85] Aravind Jembu Rajkumar, Jonathan Lazar, J Bern Jordan, Alireza Darvishy, and Hans-Peter Hutter. 2020. PDF accessibility of research papers: What tools are needed for assessment and remediation?. In *HICSS*. 1–10.

[86] Kasarapu Ramani, K Bhavana, A Akshaya, K Sai Harshita, CR Thoran Kumar, and M Srikanth. 2023. An Explorative Study on Extractive Text Summarization through k-means, LSA, and TextRank. In *2023 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*. IEEE, 1–6.

[87] Luz Rello and Ricardo Baeza-Yates. 2013. Good fonts for dyslexia. In *Proceedings of the 15th international ACM SIGACCESS conference on computers and accessibility*. 1–8.

[88] Luz Rello and Miguel Ballesteros. 2015. Detecting readers with dyslexia using machine learning with eye tracking measures. In *Proceedings of the 12th International Web for All Conference*. 1–8.

[89] Luz Rello, Gaurang Kanvinde, and Ricardo Baeza-Yates. 2012. Layout guidelines for web text and a web service to improve accessibility for dyslexics. In *Proceedings of the international cross-disciplinary conference on web accessibility*. 1–9.

[90] Luz Rello, Martin Pielot, and Mari-Carmen Marcos. 2016. Make it big! The effect of font size and line spacing on online readability. In *Proceedings of the 2016 CHI conference on Human Factors in Computing Systems*. 3637–3648.

[91] Mireia Ribera, Ricardo Pozzobon, and Sergio Sayago. 2020. Publishing accessible proceedings: the DSAI 2016 case study. *Universal Access in the Information Society* 19, 3 (2020), 557–569.

[92] Beth Richard. 2023. Key Issues Affecting the Inclusion of Alt Text in Scholarly PDF Publications. *Logos* 34, 1 (2023), 44–60.

[93] Felix M Schmitt-Koopmann, Elaine M Huang, and Alireza Darvishy. 2022. Accessible PDFs: Applying Artificial Intelligence for Automated Remediation of STEM PDFs. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–6.

[94] Sally E Shaywitz. 1998. Dyslexia. *New England Journal of Medicine* 338, 5 (1998), 307–312.

[95] Raju Shrestha. 2021. A Neural Network Model and Framework for an Automatic Evaluation of Image Descriptions based on NCAM Image Accessibility Guidelines. In *Proceedings of the 2021 4th Artificial Intelligence and Cloud Computing Conference*. 68–73.

[96] Hamzah Siddiqui, Saleha Siddiqui, Mukesh Rawat, Anas Maan, Shashaank Dhiman, and Mohd Asad. 2021. Text Summarization using Extractive Techniques. In *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*. IEEE, 28–31.

[97] Jorge Sassaki Resende Silva, André Pimenta Freire, and Paula Christina Figueira Cardoso. 2022. When headers are not there: Design and user evaluation of an automatic topicalisation and labelling tool to aid the exploration of web documents by blind users. In *Proceedings of the 19th International Web for All Conference*. 1–11.

[98] Harlan P Stevens, Carly V Winegar, Arwen F Oakley, and Stephen R Piccolo. 2023. Identifying images in the biology literature that are problematic for people with a color-vision deficiency. *bioRxiv* (2023), 2023–11.

[99] Ron Stewart, Vivek Narendra, and Axel Schmetzke. 2005. Accessibility and usability of online library databases. *Library Hi Tech* 23, 2 (2005), 265–286.

[100] Jaime A Teixeira da Silva and Maryna Nazarovets. 2023. Archiving website-based references in academic papers: Problems caused by reference rot, potential solutions and limitations. *Learned Publishing* 36, 3 (2023), 477–487.

[101] Garreth W Tigwell. 2021. Nuanced perspectives toward disability simulations from digital designers, blind, low vision, and color blind people. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–15.

[102] Utku Uckun, Ali Selman Aydin, Vikas Ashok, and IV Ramakrishnan. 2020. Breaking the accessibility barrier in non-visual interaction with pdf forms. *Proceedings of the ACM on Human-computer Interaction* 4, EICS (2020), 1–16.

[103] Andreas Uebelbacher, Roberto Bianchetti, and Markus Riesch. 2014. PDF accessibility checker (PAC 2): the first tool to test PDF documents for PDF/UA compliance. In *Computers Helping People with Special Needs: 14th International Conference, ICCHP 2014, Paris, France, July 9-11, 2014, Proceedings, Part I 14*. Springer, 197–201.

[104] Beat Vollenwyder, Serge Petralito, Glena H Iten, Florian Brühlmann, Klaus Opwis, and Elisa D Mekler. 2023. How compliance with web accessibility standards shapes the experiences of users with and without disabilities. *International Journal of Human-Computer Studies* 170 (2023), 102956.

[105] Lucy Lu Wang, Isabel Cachola, Jonathan Bragg, Evie Yu-Yen Cheng, Chelsea Haupt, Matt Latzke, Bailey Kuehl, Madeleine van Zuylen, Linda Wagner, and Daniel S Weld. 2021. Improving the accessibility of scientific documents: Current state, user needs, and a system solution to enhance scientific PDF accessibility for blind and low vision users. *arXiv preprint arXiv:2105.00076* (2021).

[106] LUCY LU WANG, ISABEL CACHOLA, BRAGG JONATHAN, EVIE YU-YEN CHENG, HAUPT CHELSEA, MATT LATZKE, KUEHL BAILEY, MADELEINE VAN ZUYLEN, and DANIEL S WELD. [n. d.]. Improving the accessibility of scientific documents. ([n. d.]).

[107] Aleena Watson and Shaun Wallace. 2021. Improving reading outcomes using digital reading rulers for readers with & without dyslexia. *Journal of Vision* 21, 9 (2021), 2650–2650.

[108] Jessica J Wery and Jennifer A Diliberto. 2017. The effect of a specialized dyslexia font, OpenDyslexic, on reading rate and accuracy. *Annals of dyslexia* 67 (2017), 114–127.

[109] Jonathan D Wren. 2004. 404 not found: the stability and persistence of URLs published in MEDLINE. *Bioinformatics* 20, 5 (2004), 668–672.

[110] 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.

[111] Yeliz Yesilada, Giorgio Brajnik, Markel Vigo, and Simon Harper. 2012. Understanding web accessibility and its drivers. In *Proceedings of the international cross-disciplinary conference on web accessibility*. 1–9.

[112] Norman E Youngblood and Michael Brooks. 2018. Website accessibility: US veterans affairs medical centers as a case study. *Business and Professional Communication Quarterly* 81, 4 (2018), 440–461.

[113] Chen-Hsiang Yu, Zachary Shelton, Omar Abou Nassif Mourad, and Mohamed A Oulal. 2021. Readability Enhancement for PDF Documents. *Frontiers in Computer Science* 3 (2021), 628832.

[114] Azizah Nurfauziah Yusril. 2020. E-accessibility analysis in user experience for people with disabilities. *Indonesian Journal of Disability Studies* 7, 1 (2020), 107–109.

[115] Xiangling Zhang, Ahmed Tlili, Fabio Nascimbeni, Daniel Burgos, Ronghuai Huang, Ting-Wen Chang, Mohamed Jemni, and Mohamed Koutheair Khribi. 2020. Accessibility within open educational resources and practices for disabled learners: A systematic literature review. *Smart Learning Environments* 7 (2020), 1–19.

[116] Han Zhuang, Tzu-Yang Huang, and Daniel E Acuna. 2023. A computational analysis of accessibility, readability, and explainability of figures in open access publications. *EPJ Data Science* 12, 1 (2023), 5.

[117] Shaban Zulfiqar, Safa Arooj, Umar Hayat, Suleman Shahid, and Asim Karim. 2020. Automated generation of accessible pdf. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–3.