IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Evaluating the Accessibility of ChatGPT for Individuals with Dyslexia

Shradha Sujeeth

Gems Modern Academy, Dubai, 53663, UAE

ABSTRACT: In recent years, Artificial Intelligence (AI) development has accelerated, with Large Language Models and chatbots becoming more prominent in society. However, their accessibility for individuals with dyslexia remains less explored. This paper aims to evaluate the extent of accessibility of ChatGPT for individuals with dyslexia using 3 frameworks: Web Content Accessibility Guidelines (WCAG), British Dyslexia Association (BDA) Style Guide, and Cognitive Load Theory (CLT) to evaluate the accessibility of ChatGPT for individuals with dyslexia by identifying possible gaps and providing recommendations. ChatGPT's general and dyslexia specific accessibility was determined by evaluating its compliance with the WCAG and BDA guidelines, respectively. Additionally, CLT was used to interpret the implications of readability scores on cognitive load. Flesch Reading Ease score (FRE) was used as the metric for output readability. Results reveal moderate accessibility: ChatGPT conforms to several WCAG guidelines but lacks compliance with the BDA style guide. Therefore, improving the accessibility of AI chatbots like ChatGPT requires evolving existing global accessibility guidelines to provide dyslexia specific recommendations, developing AI-specific guidelines to address the dynamic nature of ChatGPT outputs, and integrating personalization features into AI interfaces to address all population needs and individual need differences within disabilities.

KEYWORDS: System software, Human Machine/Interface, Accessibility, Dyslexia, Cognitive Load.

Introduction

In the past decade, the fast growth of Artificial Intelligence (AI) has led to the increased popularity of Large Language Models (LLMs) and chatbots such as ChatGPT. ChatGPT is an AI language model that was developed by OpenAI and released in 2022, and it gained over 100 million active monthly users within 2 months of its initial launch.[1] It's one of the most globally recognized chatbots. Rapid developments of AI technologies have led to their implementation in various sectors such as healthcare, education, and business.[2,3] Despite their rapid development, their accessibility and inclusivity are less explored and evaluated. Dyslexia is a highly widespread cognitive disability affecting 1 out of 10 people worldwide, with an estimated 780 million people.[4] Thus, with the rapid development of AI technologies and the massive population affected by Dyslexia, it is crucial that we ensure that these technologies are accessible to individuals with Dyslexia and that they are not neglected. This can be done by evaluating the accessibility of AI models and wrappers in terms of Dyslexia requirements. Consequently, the following paper utilizes 3 frameworks to assess ChatGPT's accessibility for dyslexic individuals: Web Content Accessibility Guidelines (WCAG), British Dyslexia Association (BDA) Style Guide, and Cognitive Load Theory (CLT).[5–7]

Background

Accessibility Guidelines:

Several guidelines for accessibility for individuals surfing the web have been published, for example Web Content Accessibility Guidelines (WCAG) is a notable accessibility framework published by the World Wide Web Consortium. The version in use is the WCAG 2.2, published in 2023, version in use and these guidelines aid in addressing a range of disabilities, including physical disabilities such as vision impairment and cognitive disabilities such as Dyslexia. Additionally, other guidelines, such as the British Dyslexia Association Style Guide, have been published. These guidelines are more focused and address accessibility requirements of digital and written material for individuals with Dyslexia.[8]

WCAG Guidelines:

The WCAG guidelines are divided into four principles: perceivable, operable, understandable, and robust.[9] 13 guidelines are spread across the principles.[5] The aim of perceivable is to ensure that content is not invisible to all the senses of an individual.[5] For example, under the perceivable principle, one guideline suggested by WCAG is that text alternatives should be provided for non-text elements such as audio or video. Guidelines under operability ensure that a user is able to navigate the web content comfortably, such as by providing keyboard shortcuts to possible challenging navigation.[5] Understandable ensures that content is simple to understand, and Robust ensures that web content is usable with assistive technologies. Each guideline is accompanied by success criteria that are testable. The success criteria are categorized into three levels: A (bare necessity), AA (moderate conformance), and AAA (highest level conformance). However, achieving an AAA conformance does not ensure absolute accessibility as the WCAG has stated that 'even content that conforms at the highest level (AAA) will not be accessible to individuals with all types, degrees, or combinations of disability, particularly in the cognitive

DOI: 10.36838/IJHSR87.301

language and learning areas.[5] This demonstrates a massive limitation of the guidelines, as it does not ensure accessibility for individuals with cognitive disabilities like Dyslexia.

BDA style guide guidelines:

With the latest version in 2023, the British Dyslexia Association (BDA) has published specific accessibility guidelines (namely the BDA style guide) for digital and written content that ensure accessibility for individuals with dyslexia. In 2018, it was considered the most well-known style guide for individuals with Dyslexia.[8] The guidelines are categorized into several sections, such as readable fonts, heading and structure, color, layout, and writing style, that ensure that the content's overall presentation and text are accessible to individuals with Dyslexia. The style guide recommends that these guidelines be considered in combination with the WCAG guidelines during web development.[7] However, it is important to note that while these guidelines are highly recognized in terms of Dyslexia, they serve as a mere recommendation and are not mandated by law, unlike the WCAG guidelines in many regions.[7,9] Therefore, this serves as a major limitation of the guidelines, as addressing its recommendations is not mandated; developers often overlook the BDA style guide guidelines, thus inhibiting accessibility for individuals with dyslexia.

Cognitive Load Theory:

Cognitive load theory is an instructional theory based on our knowledge of human cognition.[6] The theory divides mental load into 3 aspects: intrinsic load, extraneous load, and germane load. Intrinsic load refers to the complexity of a text; the higher the complexity higher the intrinsic load.[10] Complexity of a text is not only influenced by the nature of the information itself, but also by an individual's prior knowledge or familiarity with the topic.[10] Extraneous load refers to the mental effort exerted as a result of the manner in which information is presented.[11] For example, formatting of text and linguistic complexity can result in increased extraneous load as it requires more working memory, but does not directly contribute to learning.[12] Individuals with dyslexia often struggle with word recognition, which exerts an immense cognitive load that could be utilized for comprehension.[13] Germane load refers to the mental effort contributing to understanding and schema formation.[14] Therefore, reducing intrinsic and extraneous load is highly crucial for individuals with dyslexia to focus on germane load.[15] This can be done by improving the readability of texts.[8] The WCAG and BDA guidelines do make recommendations on content readability, but it merely suggests simple content but lacks specific metrics for readability. Cognitive Load Theory will be used as a framework to evaluate the accessibility of ChatGPT.

Comparison of Guidelines:

Table 1: Comparison of accessibility features between WCAG and BDA guidelines.

Feature	WCAG recommendation (general recommendation)	BDA recommendation (dyslexia specific recommendation)
Font type	No suggestion	Sans serif fonts
Font size	Text must be resizable to 200% (Level AA)	16-19px
Letter spacing	At least 0.12 the font size (Level AA)	35% of the average letter width
Line spacing	At least 1.5 the font size (Level AA)	proportional to inter-word spacing; 1.5 / 150% is preferable.
Word spacing	At least 0.16 the font size (Level AA)	Inter-word spacing should be at least 3.5 times the inter-letter spacing.
Text formatting	Avoid justification (Level AAA)	Avoid justified text, underlining, and italics. Use bold for emphasis. Avoid using capital letters and uppercase letters for continuous text.
Color and contrast ratios	Contrast greater than equal to 4:5:1. (Level AA) Contrast greater than or equal to 7:1 (Level AAA)	Use solid backgrounds. Use dark text on a light background (avoid pure white), avoid red, pink, and green
Writing content	When text requires reading ability over secondary level, include supplemental content or a version that does not require reading ability more advanced than the lower secondary level. Level (AAA)	Write in simple, clear language using everyday words Use active rather than passive voice. Be concise; avoid using long, dense paragraphs

Table 1 demonstrates that while there are a few features in which the guidelines partially or completely align, such as word spacing and text formatting, there are several features in which the guidelines misalign. The comparison reveals that WCAG does not provide guidelines for font type, while BDA recommends sans-serif for web interfaces. Font size, letter spacing, and word spacing are partially aligned as they differ in metrics. Line spacing, which was recommended as at least 1.5 times the font size, was aligned between the guidelines. Text formatting was partially aligned between the guidelines, as they both recommend avoiding justification, but WCAG lacks guidelines on the use of capitalization and styles. In terms of color contrast, both guidelines require high contrast, but they are partially aligned, as BDA provides an additional recommendation to avoid a pure white background and specific colors. The writing content guidelines partially align as they both advise simple writing, but differ in scope, with WCAG focusing on supplemental content.

These differences between the guidelines demonstrate that mere conformance to the guidelines proposed by WCAG does not ensure complete accessibility for individuals with dyslexia; thus, this is crucial to consider as several regions implement WCAG guidelines in their legal policies, laws, or frameworks. For example, the United States' Section 508 of the Rehabilitation Act states that federal government agencies must ensure that their digital tools, such as official websites, meet the level A and AA conformance level of the WCAG guidelines.[18] However, the BDA style guide acts as a recommendation and is not integrated into legal frameworks, thus developers often attempt to address general accessibility standards such as WCAG and may neglect the specific needs of individuals

with dyslexia. However, it is important to note that combining guidelines might prove difficult in certain features, such as letter spacing, as the guidelines vary in metric. For example, in letter spacing, WCAG recommends at least 0.12 times the font size while BDA guidelines recommend 0.35 times the average letter width.

Thus, to evaluate the accessibility of ChatGPT for individuals with dyslexia using a holistic approach, it is crucial that a combination of these guidelines is utilized. The WCAG guidelines provide insight into ChatGPT's general accessibility conformance, and the BDA style guide gives insight into dyslexia accessibility conformance.

The paper poses the following research question: To what extent do AI models like ChatGPT comply with WCAG and BDA style guide accessibility guidelines and ensure minimal cognitive load for individuals with Dyslexia?

Methods

First, a qualitative evaluation of ChatGPT's accessibility for individuals against the WCAG and BDA guidelines was conducted, followed by a quantitative readability analysis of ChatGPT's generated text.

1.Evaluating Accessibility:

The accessibility of ChatGPT's user interface was evaluated using the comparative table of the guidelines from the previous section. The study was conducted on macOS Sequoia 15.3.2 on default settings with no user modification. For example, no modification from light mode to dark mode was made. Additionally, the study was conducted on a Google Chrome browser. To ensure the reliability of the study, default settings of the browser were utilized, such as 100% zoom and no extensions. Using the Google Chrome developer tools, the source code (the HTML and CSS code) of ChatGPT's interface was derived to assess the key features such as font type and size; line spacing, character spacing, and word spacing; color and contrast ratios; text formatting or alignment; and the layout of the interface. Using the source code, ensured that the results of the study reflected ChatGPT's default design accessibility rather than user settings and modifications. Writing content criteria from the previous table was exempted as the guidelines lacked specific metrics.

Font type and size:

First, the elements tab in the developer's tool was navigated. The selection of specific texts contained in the specific HTML elements <p> was selected. After which, the computed tab was navigated, which displayed the CSS elements of the user interface for the text selected. In the computed tab, the 'font-family' property was noted to identify the type of font used, and the 'font-size' property value was noted to identify the font size. The process was repeated for headings and general body text. After which, the data derived was compared against the guidelines' recommendations.

Spacing:

Line spacing, character spacing, and word spacing were examined in the same manner as above. However, the property identified in the computed tab was 'line-height' for line spacing, 'word-spacing' property for word-spacing, and 'letter-spacing' property for character spacing. After which, the data derived was compared against the guidelines' recommendations.

Color and Color ratios:

The color properties for text and background were examined in the same manner. However, the property identified in the computed tab was the 'color' property for the text and 'background-color' for the background in the styles tab. When the background color property specified a CSS variable, the variable value was in the stylesheet. The values were inputted in WebAIM, an online color contrast tool. Color values that were recorded in RGB format were converted to hexadecimal codes, as the WebAIM tool only accepts hexadecimal codes to calculate color contrast ratios. The contrasting ratios generated by the WebAIM tool were compared against the guideline recommendations.

Text formatting:

The justification of the text was identified in the same manner as spacing or font type. However, the 'text-align' property was navigated in the computed tab to determine the alignment. After which, the data derived was compared against the guidelines' recommendations.

2. Readability Analysis:

In order to evaluate the readability of a text generated by ChatGPT for individuals with Dyslexia, the study used the Flesch Reading Ease (FRE) score as a metric. The FRE score quantitatively evaluates the readability of a text based on factors such as sentence length and syllable count per word. The score produced ranges from 0 to 100, and the higher the score, the lower the extraneous and intrinsic load, thus ensuring higher readability. According to studies, the recommended benchmark reading level for content for individuals with dyslexia is a grade 5 level.16 A grade 5 level translates to a score of 90-100 on the Flesch Reading Ease score. 17 Lower FRE scores demonstrate the need for increased effort and energy in decoding sentences, thus increasing extraneous and intrinsic load. Thus, less effort and energy are expended on understanding content and forming schemas, which results in reduced germane load.

To evaluate the readability, 10 prompts were developed that reflected the everyday use of ChatGPT, such as explaining a topic in simple terms, creative writing, or step-by-step instructions. While the prompt-dependent nature of ChatGPT's outputs is acknowledged, individuals with dyslexia must not be expected to 'prompt engineer' to get accessible content. Thus, the following paper evaluates the general readability trends of ChatGPT outputs in everyday contexts irrelevant of prompt phrasing. Each prompt was entered into ChatGPT's free version upon clearing browser history to prevent any discrepancies as a result of prior conversations. Each of the

DOI: 10.36838/IJHSR87.301

ten prompts was then entered into ChatGPT twice to assess consistency, resulting in 20 texts. To maintain consistency, bullet points, headings, and hyperlinks were removed from the text outputs. The text outputs were then entered into www.webfx.com/tools/read-able/ to calculate the Flesch Reading Ease Score. The score for all 20 texts was noted down, and descriptive statistics, including mean, minimum, maximum, and percentage of texts meeting accessibility benchmarks, were calculated. The FRE score was then translated to grade level equivalents using already established frameworks, such as that outlined by Dubay in 'The Principles of Readability'. The results were then tabulated.

After the collection of data, the results were organized into a table. The table includes columns for WCAG recommendation, BDA recommendation, the actual observed results/data of the ChatGPT interface, alignment with WCAG guidelines, and a final column on alignment with BDA guidelines. If a feature met or exceeded guideline requirements, it was determined as 'aligned'. If the feature partially met recommendations or was inconsistent, then it was determined as 'partially aligned'. If the feature did not meet the guideline, it was determined as 'not aligned'.

Results

1. Evaluation of Accessibility using Guidelines:

Table 2: Evaluating accessibility of ChatGPT in reference to guidelines. The table indicates ChatGPT largely complies with WCAG general accessibility requirements, but shows partial or no alignment to several BDA dyslexia specific recommendations, such as spacing.

Feature	WCAG recommendation	BDA recommendation	ChatGPT result	BDA alignment (Alignment with dyslexia needs)	WCAG alignment (General accessibility alignment)
Font type	No suggestion	Sans serif fonts	Sans Serif	Aligned	Aligned
Font size	Text must be resizable	16-19px	Headings 20px Body 16px	Partially aligned: body meets guidelines, but headings exceed range	Aligned
Letter spacing	At least 0.12 the font size (Level AA)	letter-letter / character spacing ideally around 35% of the average letter width	Default spacing, not explicitly mentioned in the source code	Not aligned	Not aligned
Line spacing	At least 1.5 the font size (Level AA)	proportional to inter-word spacing; 1.5 / 150% is preferable.	Heading: 32px Body: 28px	Aligned	Aligned
Word spacing	At least 0.16 the font size (Level AA)	Inter-word spacing should be at least 3.5 times the inter-letter spacing.	Default spacing, not explicitly mentioned in the source code	Not aligned	Not aligned
Text formatting	Text is not justified (Level AAA)	Avoid Underlining and italics. Use bold for emphasis. Avoid using capital letters and uppercase letters for continuous text.	Text is left-aligned. Bold and Italic are used based on the output, which varies from prompt to prompt	Partially aligned: Justification guideline met, but other BDA rules not consistently followed	Aligned
Color and contrast ratios	contrast 4:5:1. (Level AA) Contrast 7:1 (Level AAA)	Use solid backgrounds. Use dark text on a light background (avoid pure white to prevent glare), and avoid red, pink, and green	Contrast 20:4:1	Partially aligned: Pure white is utilized in regions, and white with slight transparency in other regions.	Aligned
Writing content	When text requires reading ability over secondary level, include supplemental content or a version that does not require reading ability more advanced than lower secondary level. Level (AAA)	Write in simple clear language using everyday words Use active rather than passive voice. • Be concise; avoid using long, dense paragraphs	Output dependent	Output dependent	Output dependent

The font type of ChatGPT's UI, which was sans serif, aligned with both WCAG guidelines and the BDA style guide. The font size of ChatGPT was partially aligned with BDA guidelines, as the body text was 16px, which complied with the WCAG font size range of 16-19px, but the heading font size of ChatGPT's UI, which was 20px, exceeded the suggested range. Font size of ChatGPT is resizable and therefore aligned with ChatGPT, which does not implement explicit fixed CSS spacing enhancement and relies on the font's, operating system, and browser's default spacing. However, both WCAG and BDA recommendations require enhanced spacing. Thus, it is not aligned with BDA or WCAG requirements. Both WCAG and BDA recommend a line spacing of at least 1.5 times the font size, and this was met and exceeded by ChatGPT's UI interface. Additionally, in terms of text formatting, ChatGPT is partially aligned with BDA guidelines as it meets the no justification requirement, but it does not meet the recommendation of avoiding the use of bold or italicized text. The color contrast ratio of ChatGPT's interface exceeds the WCAG requirement; therefore, it is aligned. However, the use of pure white color in regions does not meet the BDA style guide recommendation that suggest avoidance of pure white color to avoid glare. However, ChatGPT does utilize a solid background, which is partially aligned with the BDA style guide color contrast recommendation. Writing content compliance is prompt-dependent and variable output-to-output.

3. Readability Analysis:

Table 3: Readability Analysis of ChatGPT outputs using the FRE score. The ChatGPT output snippet shows a single line of response for reference, but the full output text was used for the FRE score calculation. The average FRE score is the mean FRE score between the 2 samples of each prompt, and SEM shows the variation between samples. Grade level corresponds to the FRE score. The results indicate that across the 20 prompts, FRE scores range from 33.3 to 90.4, with a mean score of 69.2 (+- 4.67 SEM), and only one output achieves a FRE of 90 or above.

Prompt	ChatGPT Output Snippet	Average FRE score	Grade Level	Standard error mean
Explain climate change to a teenager.	Climate change means long-term changes in the Earth's temperature, weather patterns, and climate. Right now, it mostly refers to the warming of our planet caused by human activities	74.75	7th grade	6.25

Give step-by-step instructions on how to set up a wireless printer at home.	Step 1: Unbox and Prepare the Printer Unbox the printer and remove all packaging materials.	64.9	8th - 9th grade	3.8
Write a short story about friendship.	Every Saturday morning, beneath the old oak tree in Maple Park, sat two friends—Eli and Sam.	81.55	6th grade	4.35
Describe how to prepare scrambled eggs.	Crack the eggs into a bowl. Add salt, a bit of pepper, and milk or cream if using.	79.35	7th grade	4.65
Explain the importance of recycling.	Recycling is important for several key reasons that impact the environment, economy, and society: 1. Environmental Protection	38.8	Beyond 12th grade, 13-16th grade	5.5
Write a short letter apologizing to a friend for being late.	Dear [Friend's Name], I'm really sorry for being late the other day.	80.5	6th grade	9.9
Summarize the plot of "Three Little Pigs".	"The Three Little Pigs" is a classic folktale about three pig siblings who each build a house to protect themselves from the Big Bad Wolf.	84.95	6th grade	3.25
Provide advice on how to improve sleep quality.	Improving sleep quality often involves changes in lifestyle, environment, and behaviour.	61.45	8th - 9th grade	3.65
Describe what a solar eclipse is.	A solar eclipse occurs when the Moon passes between the Earth and the Sun, blocking all or part of the Sun's light from reaching Earth.	74.4	7th grade	1.0
Explain why it is important to wash your hands regularly.	Washing your hands regularly is important because it helps prevent the spread of germs and infections.	50.9	10th - 12th grade	4.3

The mean Flesch reading Ease Score of the 20 text outputs generated across 10 prompts is 69.2 (+/- 4.67 SEM), which translates to a grade 8-9 level. The Flesch Reading Ease Score of the outputs ranged from 33.3 to 90.4. Using the 5th-grade readability benchmark, which translates to a minimum of 90 Flesch reader score and higher, the score indicates the easier the readability. It was determined that only 1 out of the 20 outputs met the criteria. Therefore, most of the outputs fell below the benchmark.

Discussion

1. Evaluation of Accessibility:

The evaluation of the accessibility of the ChatGPT user interface against the WCAG and BDA guidelines demonstrates overall moderate compliance, as it does not conform to all the requirements of each guideline. ChatGPT displays more compliance with WCAG guidelines than the BDA style guide. However, ChatGPT still does not adhere to certain features of the WCAG guidelines, such as letter spacing and word spacing. ChatGPT does not explicitly declare numerical values for 'letter-spacing' or 'word-spacing' in its source code, relying on the font's, operating system, and browser's default spacing. However, WCAG and BDA have specific spacing requirements for enhanced readability. Since ChatGPT's spacing in its UI may vary across devices, the spacing may not consistently meet the requirement for consistent general or dyslexia specific accessibility. Though the letter spacing and word spacing may not align in metrics between WCAG and BDA guidelines, both guidelines do recommend enhanced spacing for readability. Lack of compliance with these spacings can inhibit efficient readability for individuals, particularly those with Dyslexia.

Although the body text complies with both WCAG and BDA guidelines, the heading font size slightly exceeds the BDA style guide recommendation; thus, this can inhibit accessibility for individuals with dyslexia who require a consistent font size for efficient navigation. The font type and use of line spacing adhere to both the guidelines, thus supporting general accessibility and accessibility for individuals with Dyslexia. The text formatting of content generated by ChatGPT partially aligns with the BDA and WCAG guidelines, as they avoid justification, but the usage of bold and italics in its outputs misaligns with the BDA style guide recommendation, as this can reduce readability for individuals with dyslexia due to the addition of distractions and loss of consistency in the text output.

The color contrast of the text and background of ChatGPT's UI completely aligns with WCAG guidelines, but it only partially aligns with BDA guidelines due to the usage of a pure white background in certain regions of the UI. The usage of pure white can increase glare and result in eye strain, which can serve as a further challenge for individuals with Dyslexia.

Therefore, the compliance of ChatGPT's UI to most of the WCAG guidelines, with the exception of spacing indicate that the interface is accessible to an extent on a general accessibility level and by legal frameworks. However, the lack of compliance with several BDA style guide recommendations demonstrates that ChatGPT's UI is not completely accessible to individuals with dyslexia.

2. Readability Analysis:

The evaluation of the written content or outputs of ChatGPT using the BDA style guide and WCAG guidelines was omitted from the methodology, as the guidelines lack specific metrics to analyze the readability of the text. However, we must analyse the text's readability as linguistic complexity can result in increased cognitive load, particularly extraneous load.[12] Individuals with dyslexia already experience high cognitive load, particularly intrinsic and extraneous load, as they often struggle with word recognition; therefore, factors such as complex linguistics or long sentence structures could add to the extensive present cognitive load.[13] Thus, the Flesch Reading Ease Score formula was used as a metric to analyze the

DOI: 10.36838/IJHSR87.301

readability of the text outputs of ChatGPT. The FRE score evaluates the readability of a text based on its sentence length and syllable counts per word. The higher the FRE score, the lower the extraneous and intrinsic load. When extraneous and intrinsic load is lowered, more energy can be expended on germane load, thus ensuring higher text readability.

Since only 1 out of the 20 text outputs received a Flesch Reading Score higher than 90, which is the benchmark for reading score of text for individuals with dyslexia, it demonstrates that content generated by ChatGPT is not generally accessible to individuals with dyslexia. Additionally, the large range of scores between prompt outputs demonstrates that ChatGPT accessibility is prompt-dependent and not consistent. Additionally, the occasional variance in scores between prompt trials demonstrates that ChatGPT outputs are inconsistent in terms of readability. These inconsistent outputs and readability levels of the texts pose a huge challenge to individuals with dyslexia, as the constant change in reading level can add to the cognitive load, thus making reading/comprehension more difficult. Thus, to an extent, the readability of ChatGPT does not meet Dyslexia standards. However, it is important to note that there are other factors besides sentence length and syllable count that affect the readability of a text, which could either increase or decrease the cognitive load and readability levels. Yet, since the FRE score is widely recognized for the reading level of a text's assessment, it can be used as an estimate of the complexity of a text.

3. Limitations and recommendation:

Although a comprehensive approach was aimed at, the paper has several limitations. The following paper did not conduct a direct dyslexic individual experience study and instead utilized existing guidelines, such as the BDA style guide, to assess the accessibility of the interface. Additionally, the readability analysis used the formula FRE, which used factors such as sentence length and syllable count, but did not take into consideration other factors of linguistic complexity.

The study utilized a MacOS Sequoia 15.3.2 and the browser Google Chrome due to availability. No user modifications were made, and default settings were maintained. Although the developer tools were utilized to extract source code to identify ChatGPT's specific UI choices, as ChatGPT's source code demonstrates what ChatGPT intends to display. However, variation may occur due to rendering differences in certain elements, such as font type in operating systems or browsers. These rendering differences may occur in all web-based platforms; nonetheless, the findings of the study provide a valuable initial assessment of ChatGPT's accessibility.

Due to the lack of complete adherence of ChatGPT's UI to the guidelines, ChatGPT should develop its interface in future versions in adherence to the BDA style guide guidelines to ensure accessibility for individuals with Dyslexia. However, if adhering to all guidelines proves difficult, then ChatGPT must offer personalization, for example, in spacing and text size, so that individuals can tailor the UI as per their needs, thus addressing all population needs.

In terms of guidelines, general guidelines such as the WCAG must evolve more to support more Dyslexia specific needs and add metrics for readability of text displayed. Unlike general websites that are static in nature, AI model and wrapper outputs are highly dynamic in nature and prompt-dependent; these guidelines must evolve and develop specific recommendations to address this inconsistency.

■ Conclusion

As AI technologies such as ChatGPT are increasingly gaining popularity and being integrated into sectors of society and daily life, such as in education, it is important that these technologies are accessible to all populations, including individuals with cognitive disabilities. ChatGPT's UI does not completely comply with the dyslexia specific guidelines, thus suggesting that future chatbot builds should develop solutions to address this gap.

My paper contrasts two solutions: 1. Change the guidelines, 2. Develop better AI Chat builds. Changing the guidelines includes ensuring that existing general accessibility guidelines, such as the WCAG, incorporate dyslexia specific accessibility needs, readability metrics, and possible cognitive load metrics. Additionally, developing distinct AI-specific guidelines that address the dynamic and variable nature of ChatGPT outputs could ensure consistent accessibility for individuals with dyslexia and other cognitive disabilities. Solution two involves developing better AI chat builds by developing its interface in future versions in adherence to the BDA style guide guidelines to ensure accessibility for individuals with dyslexia. However, if addressing the needs of different populations proves difficult, personalization of interfaces can improve accessibility for all individuals with any form of disabilities or individual needs.

While both solutions can offer an immense impact on the overall accessibility of Artificial intelligence chatbots, my analysis suggests that option 2 (developing better AI chatbots) is a more viable solution. Improving AI interfaces and integrating personalization features can be implemented more rapidly than changing global accessibility standards, and this solution also tackles the challenge of individual differences between individuals with disabilities' needs. However, combining the solutions is more ideal as it betters long term AI accessibility while providing immediate impact.

■ Acknowledgments

I would like to thank my mentor, Dr. Jorge A. Avila, for his support and guidance throughout my project.

■ References

1. Andriansyah, Y. The Current Rise of Artificial Intelligence and Religious Studies: Some Reflections Based on ChatGPT. *Millah J. Relig. Stud.* **2023**, ix–xviii. https://doi.org/10.20885/millah.vol22.iss1.editorial.
2. Akpan, I. J.; Kobara, Y. M.; Owolabi, J.; Akpan, A. A.; Offodile, O. F. Conversational and Generative Artificial Intelligence and Human–Chatbot Interaction in Education and Research. *Int. Trans. Oper. Res.* **2025**, *32* (3), 1251–1281. https://doi.org/10.1111/itor.13522.

3. Boscardin, C. K.; Gin, B.; Golde, P. B.; Hauer, K. E. ChatGPT and Generative Artificial Intelligence for Medical Education: Potential Impact and Opportunity. *Acad. Med. J. Assoc. Am. Med. Coll.* **2024**, *99* (1), 22–27. https://doi.org/10.1097/ACM.0000000000005439.
4. *33 Dyslexia Statistics & Facts: How Many People Have Dyslexia?* https://www.crossrivertherapy.com/research/dyslexia-statistics (accessed 2024-08-15).
5. *Web Content Accessibility Guidelines (WCAG) 2.2*. https://www.w3.org/TR/WCAG22/ (accessed 2025-08-09).
6. Sweller, J.; Ayres, P.; Kalyuga, S. *Cognitive Load Theory*; Springer: New York, NY, 2011. https://doi.org/10.1007/978-1-4419-8126-4.
7. Dyslexia Style Guide 2023.
8. Yoliando, F. T. A Comparative Study of Dyslexia Style Guides in Improving Readability for People With Dyslexia; Atlantis Press, 2020; pp 32–37. https://doi.org/10.2991/assehr.k.201202.050.
9. *Web Content Accessibility Guidelines (WCAG) 2.1*. https://www.w3.org/TR/WCAG21/ (accessed 2025-04-04).
10. Sweller, J.; van Merriënboer, J. J. G.; Paas, F. Cognitive Architecture and Instructional Design: 20 Years Later. *Educ. Psychol. Rev.* **2019**, *31* (2), 261–292. https://doi.org/10.1007/s10648-019-09465-5.
11. Lange, C.; Costley, J.; Han, S.-L. The Effects of Extraneous Load on the Relationship Between Self-Regulated Effort and Germane Load Within an E-Learning Environment. *Int. Rev. Res. Open Distrib. Learn.* **2017**, *18* (5). https://doi.org/10.19173/irrodl.v18i5.3028.
12. Strohmaier, A. R.; Ehmke, T.; Härtig, H.; Leiss, D. On the Role of Linguistic Features for Comprehension and Learning from STEM Texts. A Meta-Analysis. *Educ. Res. Rev.* **2023**, *39*, 100533. https://doi.org/10.1016/j.edurev.2023.100533.
13. Tricot, A.; Vandenbroucke, G.; Sweller, J. Using Cognitive Load Theory to Improve Text Comprehension for Students with Dyslexia. In *Handbook of educational psychology and students with special needs*; Educational psychology handbook series; Routledge/Taylor & Francis Group: New York, NY, US, 2020; pp 339–362. https://doi.org/10.4324/9781315100654-17.
14. Debue, N.; van de Leemput, C. What Does Germane Load Mean? An Empirical Contribution to the Cognitive Load Theory. *Front. Psychol.* **2014**, *5*, 1099. https://doi.org/10.3389/fpsyg.2014.01099.
15. López-Resa, P.; Moraleda-Sepúlveda, E. Working Memory Capacity and Text Comprehension Performance in Children with Dyslexia and Dyscalculia: A Pilot Study. *Front. Psychol.* **2023**, *14*. https://doi.org/10.3389/fpsyg.2023.1191304.
16. Matsuura, Y.; Jaeah, C. Dyslexia Articles Unboxed: Analyzing Their Readability Level. *J. Dev. Behav. Pediatr. JDBP* **2024**, *45* (3), e211–e216. https://doi.org/10.1097/DBP.0000000000001274.
17. DuBay, W. H. *The Principles of Readability*; ERIC Clearinghouse, 2004.
18. *U.S. Access Board - Revised 508 Standards and 255 Guidelines*. https://www.access-board.gov/ict/ (accessed 2025-08-10).

Author

Shradha Sujeeth, a high school senior passionate about psychology, is the author of a children's book and a certified adolescent mental health first aider. She earned a perfect 5.0 in AP psychology and regularly engages in academic writing, earning a merit in the John Locke Essay Competition. She plans to study psychology at university.

DOI: 10.36838/IJHSR87.301