

Acessibilidade em Plataformas de IA Generativa: Uma Análise Comparativa entre Ferramentas Populares

Arthur Diniz de Souza¹, Erica de Castro Silveira², Ingrid Teixeira Monteiro²

¹College of Interdisciplinary Studies – Zayed University – Dubai, U.A.E

²Universidade Federal do Ceará – Quixadá, CE – Brasil

arthurdinizdesouza@gmail.com, ery.cxtr@alu.ufc.br, ingrid@ufc.br

Abstract. **Introduction:** Generative AI platforms, such as ChatGPT, Copilot, Gemini, and Claude, are widely adopted worldwide, yet little is known about their accessibility for people with disabilities. **Objective:** The objective of this study is to analyze the accessibility of web interfaces of leading generative AI platforms, identifying compliance and violations with the Web Content Accessibility Guidelines (WCAG) 2.1 and 2.2. **Methodology:** Automated accessibility evaluations were performed using the “Accessibility Insights for Web” tool, which internally applies the axe-core 4.10 engine to detect accessibility issues in the selected platforms. **Results:** All platforms evaluated presented at least one critical accessibility violation, which may hinder usage by people with disabilities. Based on these findings, a set of best practices is proposed to guide developers in promoting digital inclusion and regulatory compliance.

Keywords Accessibility, Generative AI, Web Interfaces, WCAG, Automated Validation.

Resumo. **Introdução:** Plataformas de IA generativa, como ChatGPT, Copilot, Gemini e Claude, são amplamente utilizadas em todo o mundo, mas pouco se sabe sobre a acessibilidade desses sistemas para pessoas com deficiência. **Objetivo:** O objetivo deste estudo é analisar a acessibilidade de interfaces web das principais plataformas de IA generativa, identificando conformidades e violações em relação às Diretrizes de Acessibilidade para Conteúdo Web (WCAG) 2.1 e 2.2. **Metodologia:** Foram realizadas avaliações automáticas de acessibilidade com a ferramenta “Accessibility Insights for Web”, que utiliza internamente o mecanismo axe-core 4.10 para detectar problemas de acessibilidade nas plataformas selecionadas. **Resultados:** Todas as plataformas avaliadas apresentaram pelo menos uma violação crítica de acessibilidade, que pode dificultar o uso por pessoas com deficiência. A partir desses achados, é proposto um conjunto de boas práticas para orientar desenvolvedores na promoção de maior inclusão digital e conformidade regulatória.

Palavras-Chave Acessibilidade, IA generativa, Interfaces Web, WCAG, Validação automática.

1. Introdução

Interfaces de Inteligência Artificial generativa se tornaram amplamente utilizadas, oferecendo desde assistência em tarefas até suporte criativo. Entretanto, sua conformidade

com as diretrizes de acessibilidade ainda é limitada. Uma análise recente de cinquenta aplicações constatou significativos problemas de acessibilidade, evidenciando barreiras que afetam sobretudo pessoas com deficiência [Acosta-Vargas et al. 2024].

Apesar da crescente dependência desses sistemas por usuários com diferentes perfis e necessidades, poucas avaliações investigam se as próprias plataformas de IA atendem aos requisitos da WCAG (Web Content Accessibility Guidelines)¹. Um estudo de 2024 mostrou que a maioria dos chatbots baseados em IA generativa ainda carece de compatibilidade nativa com leitores de tela, forçando pessoas com deficiência visual a recorrer a soluções externas e agravando as barreiras de uso [Sarangam 2024], evidenciando a importância de uma avaliação aprofundada. Para garantir que essas interfaces sejam utilizáveis por todos, este estudo realiza uma análise comparativa nas páginas iniciais de quatro plataformas populares, avaliando sua aderência às WCAG.

Avaliações anteriores mostraram limites importantes dos verificadores automáticos e divergências entre ferramentas, recomendando cautela na interpretação de resultados [Vigo et al. 2013, Brajnik 2009]. Diferentemente desses estudos gerais, e complementando [Acosta-Vargas et al. 2024, Sarangam 2024], este artigo foca um recorte específico e atual: as páginas públicas de plataformas populares de IA generativa, auditadas com WCAG 2.2 e axe-core 4.10. Essa combinação de domínio (IA generativa) e versões recentes das diretrizes/engine oferece uma comparação transversal atualizada por princípio da WCAG e por gravidade das falhas.

2. Metodologia

Esta avaliação foi realizada em junho de 2025 e usou verificações automatizadas, assegurando replicabilidade no curto prazo. Utilizou-se a extensão Accessibility Insights for Web (v. 2.46) [Microsoft Accessibility Insights Team 2025], com o mecanismo de teste de acessibilidade axe-core 4.10 [Deque Systems Inc. 2025], que executa trinta e cinco testes baseados nas recomendações WCAG 2.2 [World Wide Web Consortium 2023]. Foram analisadas as páginas iniciais públicas (o primeiro ponto de contato com o usuário) dos quatro principais serviços de IA generativa disponíveis no período: ChatGPT, Microsoft Copilot, Google Gemini e Claude. Cada página foi aberta no Google Chrome 125, em ambiente desktop e ampliação padrão (100%), sendo então submetida aos testes, que geraram relatórios com as falhas identificadas pelo axecore.

Foram registradas quatro métricas principais: (i) número absoluto de falhas automáticas por página; (ii) taxa de conformidade, definida como a razão entre verificações aprovadas e total de testes executados; (iii) distribuição das falhas pelos quatro princípios da WCAG 2.2: Perceptível, Operável, Compreensível e Robusto; e (iv) classificação de severidade indicada pelo axe-core (crítica, grave, moderada ou leve).

Essa análise reflete o estado das interfaces na data da avaliação (8 de junho de 2025), portanto alterações posteriores nos sites podem modificar os valores observados.

¹As Web Content Accessibility Guidelines (WCAG) são publicadas pela Iniciativa de Acessibilidade na Web (WAI) do W3C. A versão 2.2 organiza critérios de sucesso segundo os princípios Perceptível, Operável, Compreensível e Robusto (POUR) e níveis A/AA/AAA; na prática, o alvo comum é o nível AA. Ver [World Wide Web Consortium 2023].

O procedimento descrito fornece uma base comparativa consistente e de fácil reprodução para aferir o nível mínimo de acessibilidade inicial das plataformas estudadas.

3. Resultados

Foram geradas de 32 a 26 verificações (*pass* ou *fail*) por página, resultando em um total de 118 verificações nos quatro serviços. O Google Gemini apresentou o melhor desempenho absoluto, com apenas duas falhas; o total de 24 testes aprovados resulta numa taxa de conformidade de 92% (24/26). Na sequência surgem Microsoft Copilot, com sete falhas e 27 aprovações, correspondendo a 79% de conformidade, e Claude, que registou igualmente sete falhas (24 aprovações; 77%). O ChatGPT exibiu o maior número de inconformidades, oito falhas, para 24 testes aprovados, o que configura a menor taxa do conjunto (75%).

Tabela 1. Resultados das verificações automáticas.²

Serviço	Testes aprovados	Falhas	Total analisado	Conformidade (%)
ChatGPT	24	8	32	75
Microsoft Copilot	27	7	34	79
Google Gemini	24	2	26	92
Claude	24	7	31	77

Exemplos por severidade

- **Crítica** (WCAG 2.1.1 — Teclado): controles não alcançáveis por Tab — #upload-file-btn, #system-hint-button.
- **Grave** (WCAG 2.4.3 — Ordem do foco): tabulação salta (focus-guard 28 → 1).
- **Grave** (WCAG 1.4.3 — Contraste): “Chats” = 3,07:1; texto secundário = 3,23:1.
- **Grave** (WCAG 4.1.2 — Nome acessível): botão sem nome programático (*button-name*).
- **Moderada** (WCAG 4.1.2 — ARIA): aria-expanded em div sem papel compatível.

Em termos de gravidade, o axe-core classificou 13% das falhas como críticas, 54% como graves e o restante como moderadas. As inconformidades críticas concentram-se principalmente em controles interativos inacessíveis por teclado (como o botão de *upload* no ChatGPT) e em casos de contraste tipográfico abaixo do limiar intermediário do WCAG (nível AA). Já as falhas graves distribuem-se entre a ausência de rótulos acessíveis em ícones (regra WCAG 4.1.2) e a utilização inadequada de atributos ARIA em elementos neutros, prática observada nos quatro serviços.

Ao analisar a distribuição por princípio WCAG 2.2, verificou-se que Perceptível e Operável respondem, em conjunto, por aproximadamente 63% de todas as falhas. O domínio Perceptível é dominado por problemas de contraste: o token de cinza claro (#8F8F8F) surge na barra lateral do ChatGPT, nos contadores do Copilot e nos rodapés do Gemini, violando a relação mínima de 4,5:1 exigida para texto de corpo. No domínio Operável, predominam problemas de navegação por teclado e de foco visível,

²Dados completos e relatórios por plataforma: <https://drive.google.com/drive/u/4/folders/1pTw706dusCfBnXHpmELRbZLg16FdOJou>.

além de controles acionados por <div> sem função semântica de botão. Os princípios Compreensível e Robusto exibem menor incidência, ainda que existam links genéricos (“Learn more”) sem destino claro e estados ARIA aplicados de forma inadequada.

Em resumo, apesar das diferenças nos valores, as quatro plataformas compartilham dois padrões problemáticos: design cromático que compromete contraste e ausência de rotulagem acessível em ícones e botões de menu. Mesmo o Gemini, que obteve o melhor resultado, possui alguns desses problemas, evidenciando que a adoção de bibliotecas ou tokens comuns repercutem de maneira sistemática na acessibilidade. Esses resultados fornecem um panorama comparável do “nível mínimo” de acessibilidade garantido pelas plataformas, servindo de base para futuras inspeções complementares.

3.1. Recomendações imediatas

Os resultados apontam três intervenções relativamente simples que podem eliminar a maior parte das falhas: (i) substituir o cinza dos textos secundários por uma cor que garanta contraste mínimo de 4,5:1; (ii) assegurar que todos os ícones e botões recebam rótulos acessíveis (aria-label ou texto visível); e (iii) corrigir a sequência de tabulação, atribuindo foco visível e tabindex="0" a todos os controles. A adoção dessas medidas resolveria a maioria das violações críticas e graves sem exigir refatoração extensa.

4. Conclusão

A análise mostra que, mesmo entre plataformas de IA generativa de grande visibilidade e recursos avançados, persistem violações das diretrizes WCAG, que restringem o acesso de pessoas com deficiência. Embora o Google Gemini tenha apresentado o melhor resultado, todos os serviços analisados exibiram pelo menos uma falha crítica, geralmente concentrada em contraste insuficiente e falta de navegação plena pelo teclado. Esses resultados confirmam que soluções de altamente tecnológicas não garantem, por si só, interfaces inclusivas.

As recomendações propostas indicam que a correção desses pontos exige, na maioria dos casos, ajustes limitados ao nível de *front-end*, sem necessidade de reescrever a lógica de aplicação. A adoção de verificações automatizadas, integradas ao fluxo de desenvolvimento, pode prevenir regressões e consolidar a cultura de acessibilidade desde as primeiras iterações do produto.

4.1. Próximos Passos Planejados

Neste pôster, apresentamos os dados de uma avaliação inicial de acessibilidade das interfaces de IA generativa. Como próximos passos desta pesquisa, estão inclusos os seguintes pontos:

- **Testes assistivos manuais:** *keyboard-only* e leitores de tela (NVDA/JAWS/VoiceOver/TalkBack) em fluxos reais (login, chat, upload, imagem, configurações). Métrica: taxa de sucesso sem mouse, eventos anunciados e ordem de foco.
- **Ampliar escopo:** páginas autenticadas, estados dinâmicos (modais, toasts, menus), mobile web e, quando existirem, apps nativas; múltiplos navegadores/SO.
- **Cobrir WCAG 2.2 específicos:** 2.4.11/2.4.12 (foco visível/não encoberto), 2.5.7 (arrastar com alternativa), 2.5.8 (alvo mínimo), 3.3.7 (entrada redundante) e 3.3.8 (autenticação acessível).

4.2. Cuidados Éticos

Este estudo não realizou experimentos com participantes humanos, estando, portanto, isento de submissão a comitê de ética conforme as Resoluções do Conselho Nacional de Saúde. As análises foram conduzidas exclusivamente sobre informações de acesso público disponíveis na Internet, sem envolver dados pessoais ou qualquer intervenção que pudesse representar risco ético.

Referências

- Acosta-Vargas, P. G., Guillén, M. L., e Quezada, A. M. (2024). Web accessibility evaluation of fifty ai applications: Empirical findings and challenges. *International Journal of Human–Computer Interaction*.
- Brajnik, G. (2009). Validity and reliability of web accessibility guidelines. In *Proceedings of the 2009 International Cross-Disciplinary Conference on Web Accessibility (W4A '09)*. ACM.
- Deque Systems Inc. (2025). *axe-core: Accessibility Testing Engine, versão 4.10*. Deque Systems. Disponível em: <https://github.com/dequelabs/axe-core>. Acesso em: 09 jun. 2025.
- Microsoft Accessibility Insights Team (2025). *Accessibility Insights for Web, versão 2.46*. Microsoft Corporation. Disponível em: <https://accessibilityinsights.io>. Acesso em: 09 jun. 2025.
- Sarangam, R. (2024). Screen-reader compatibility in generative chatbots: An exploratory study. In *Proceedings of the 2024 ACM Web Conference*. ACM.
- Vigo, M., Brown, J., e Conway, V. (2013). Benchmarking web accessibility evaluation tools: Measuring the harm of sole reliance on automated tests. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility (W4A '13)*. ACM.
- World Wide Web Consortium (2023). Web content accessibility guidelines (wcag) 2.2. Technical report, W3C. Disponível em: <https://www.w3.org/TR/WCAG22/>. Acesso em: 09 jun. 2025.