

Um Estudo Sobre o Uso de Prompts LLM para a Avaliação da Acessibilidade Web dos Portais das Prefeituras da Região dos Sertões dos Crateús

Lúcio Cauper Freitas de França
luciofranca@alu.ufc.br
Universidade Federal do Ceará -
Campus Crateús

Lisieux Marie M. dos S.
Andrade
lisieuxandrade@ufc.br
Universidade Federal do Ceará -
Campus Crateús

Windson Viana de Carvalho
windson@virtual.ufc.br
Universidade Federal do Ceará -
Instituto UFC Virtual

ABSTRACT

This study is an initial investigation into the use of Large Language Models (LLMs) with prompts written in Portuguese for web accessibility evaluation, using a set of 13 municipal government portals as a case study. ChatGPT 4o was employed to verify 37 success criteria from WCAG 2.2, and its coverage was compared with that of the automated inspection tools IBM, Achecker, and WAVE. Preliminary results indicate that the LLM provides positive contributions as a support tool in the web accessibility verification process.

KEYWORDS

Large Language Models, accessibility, web.

1 INTRODUÇÃO

O acesso à informação é uma garantia de direito previsto em Lei no Brasil, todo cidadão possui o direito ao acesso à informação pública [4] e os procedimentos previstos cabem ser observados pela União, Estados, Distrito Federal e Municípios, contudo, nem sempre o acesso à informação é, de fato, acessível [2, 3, 15]. Estando aqui um aspecto importante a ser destacado: a diferença existente entre acesso e acessibilidade. O acesso refere-se à disponibilidade da informação, enquanto a acessibilidade abrange a diminuição de barreiras para a garantia equitativa da informação por todos, sem distinção, e que no Brasil sua garantia também é prevista em Lei[5].

A disponibilização digital da informação na web possui modelos, normas e diretrizes de acessibilidade que fomentam regras definidas para o atendimento da acessibilidade, a exemplo da *Web Content Accessibility Guidelines* (WCAG) [17], do Modelo de Acessibilidade em Governo Eletrônico (e-MAG) [6] e, recentemente, da Norma Brasileira de Acessibilidade em Conteúdo e Aplicações Web - Requisitos (NBR 17225:2025) [1], estando as duas últimas alicerçadas na primeira. Porém, mesmo com o estabelecimento de tais diretrizes, trabalhos como os de Pardini et al. [12], Sheoran et al. [16], Dorneles and Melo [7] e Macedo et al. [11] apresentam investigações sobre pontos de melhoria na acessibilidade e na usabilidade em portais governamentais, fato que destaca a temática como relevante e importante, o que motiva a realização da presente pesquisa.

Outro aspecto motivador é o avanço de discussões e contribuições acerca do uso de Inteligência Artificial Generativa por meio

das *Large Language Models* (LLMs) no apoio ao processo de desenvolvimento de softwares no contexto web [9]. Nesta perspectiva, verifica-se que investigar o uso de LLMs como assistente no processo de avaliação da acessibilidade web pode contribuir para o estabelecimento de portais mais acessíveis e para a disponibilização do acesso à informação pública, coerente com o direito de todo cidadão, refletindo assim na composição de uma sociedade mais igualitária.

Desta forma, a presente pesquisa teve como estudo de caso verificar a acessibilidade web dos portais das prefeituras que compõem a região dos Sertões dos Crateús, um conjunto de 13 cidades localizadas no oeste do estado do Ceará. Tendo a seguinte pergunta de pesquisa: utilizar LLM, com *prompt* em português, pode apoiar o processo de verificação da acessibilidade de portais web?

Os resultados alcançados neste estudo são base para investigações futuras com outras LLMs; com o uso do mesmo processo metodológico aplicado à verificação da violação de diretrizes de acessibilidade web nacionais (e-MAG e NBR 17225:2025); e na exploração do viés multilíngue de LLMs na avaliação da acessibilidade web, uma vez que estudos anteriores já evidenciam esta característica no processo de geração de códigos [18].

Este artigo está estruturado em cinco seções. Na segunda, é apresentado como o uso de LLM tem contribuído em investigações sobre estudos da acessibilidade web. Na terceira, são apresentadas as definições da metodologia e os materiais utilizados nesta pesquisa. Na quarta seção, são apresentados os resultados dos processos de avaliação realizados, essenciais para a resposta à pergunta de pesquisa lançada. E, por fim, a última seção apresenta as considerações finais da pesquisa.

2 LLM NO CONTEXTO DA ACESSIBILIDADE WEB

Estudos sobre a aplicação de LLMs como apoio ao processo de correções de erros ou problemas, e no desenvolvimento de portais ou recursos para web, têm revelado discussões, contribuições e avanços. Contudo, até o momento, são poucos os estudos que abordam o uso de LLMs no processo de verificação da acessibilidade web.

López-Gil and Pereira [10] utilizaram o ChatGPT (nas versões 3.5 e 4)¹, Anthropic Claude², Google Bard³ e LangChain⁴, como apoio para a análise de três critérios de sucesso da WCAG: 1.1.1

¹<https://openai.com/pt-BR/>

²<https://claude.ai>

³<https://bard.google.com/>

⁴<https://www.langchain.com/>

Conteúdo não textual, 2.4.4 Finalidade do link (em contexto) e 3.1.2 Idioma das partes, por meio de *prompts* escritos em língua inglesa, e evidenciaram contribuições positivas que o uso de LLM pode promover no processo de verificação da acessibilidade.

O estudo de He et al. [8] apresenta uma investigação nesta perspectiva, com o desenvolvimento da ferramenta GenAI1y, que extrai informações de uma página web e, por meio de *prompts*, escritos em inglês, com breves descrições dos critérios analisados da WCAG 2.2 para ativar a memória da LLM, aplica-os ao ChatGPT-4o para a realização da análise. O estudo avaliou o uso da ferramenta em casos de testes, e seu desempenho demonstrou a detecção de mais violações de critérios de sucesso do que as quatro ferramentas de inspeção automática utilizadas no processo de comparação.

Inspirado nestes estudos e em atendimento à questão de pesquisa levantada aqui, o presente trabalho faz uso de *prompts*, em estilo *zero-shot* (sem exemplos ou descrições dos critérios da WCAG), escritos em português, aplicados à LLM ChatGPT-4o para verificar seu desempenho no alcance de critérios de sucesso da WCAG 2.2, quando comparado a três ferramentas automáticas de avaliação no estudo de caso utilizado. O estudo de caso, os materiais e o processo metodológico adotado são apresentados na seção a seguir.

3 MATERIAIS E MÉTODOS

Esta pesquisa possui natureza aplicada e descriptiva quanto aos objetivos, e tem uma abordagem quantitativa, visto o seu propósito de avaliar o comportamento de uma LLM na análise da acessibilidade dos sites das prefeituras que compõem a região, frente ao desempenho de três ferramentas de inspeção automática[13]. A etapa metodológica foi inspirada no trabalho de Albuquerque et al. [2], sendo composta por quatro subetapas sequenciais: seleção da amostra, definição dos instrumentos de uso para inspeção, execução da inspeção e análise e interpretação dos resultados.

Quanto ao processo de seleção da amostra, este correspondeu ao acesso aos 13 portais das prefeituras que compõem o estudo de caso, conforme Tabela 1. A escolha deste estudo de caso foi estabelecida para verificar as contribuições do uso de LLM em sites reais e governamentais brasileiros. O processo de seleção da amostra ocorreu em julho de 2025. A definição do uso da diretriz WCAG

Tabela 1: Portais das prefeituras que pertencem a região dos Sertões dos Crateús.

ID	Cidade	Portais
[1]	Ararendá	https://ararendá.ce.gov.br/
[2]	Catunda	https://catunda.ce.gov.br/
[3]	Crateús	https://crateús.ce.gov.br/
[4]	Hidrolândia	https://hidrolândia.ce.gov.br/
[5]	Independência	https://independência.ce.gov.br/
[6]	Ipaporanga	https://ipaporanga.ce.gov.br/
[7]	Ipueiras	https://ipueiras.ce.gov.br/
[8]	Monsenhor Tabosa	https://monsenhor-tabosa.ce.gov.br/
[9]	Nova Russas	https://novarussas.ce.gov.br/
[10]	Novo Oriente	https://novooriente.ce.gov.br/
[11]	Poranga	https://poranga.ce.gov.br/
[12]	Santa Quitéria	https://santa-quitéria.ce.gov.br/
[13]	Tamboril	https://tamboril.ce.gov.br/

2.2 foi estabelecida para fins de comparação do desempenho da LLM com as ferramentas de inspeção. Os instrumentos de uso para inspeção adotados foram uma LLM e três ferramentas automáticas de inspeção: IBM Equal Access Toolkit (IBM)⁵, Web Accessibility

⁵<https://www.ibm.com/able/toolkit/>

Evaluation Tools (WAVE)⁶ e Accessibility Checker (AChecker)⁷, que utilizam a WCAG 2.2 como referência.

A LLM utilizada foi o ChatGPT-4o. Para o estabelecimento do uso dessa LLM, foram avaliadas também: Copilot⁸, Sabiá 3.1⁹ e Llama 4¹⁰. Contudo, para estas, algumas limitações foram encontradas, tais como: o não acesso a páginas web via *prompt* e o não recebimento de arquivos com extensão .html, o que impossibilitou as ações de análise pretendidas nesta pesquisa. Desta forma, a justificativa para a definição do uso do ChatGPT-4o foi a possibilidade de indexação de arquivos HTML estáticos dos portais (Tabela 1).

Para o uso da LLM, um *prompt* (Figura 1) foi estabelecido como padrão, inspirado no trabalho de He et al. [8], porém escrito em português. A justificativa para o uso da língua portuguesa foi a simulação da perspectiva de usuários cuja língua nativa é o português, que buscam utilizar LLMs como recurso auxiliar no processo de detecção da acessibilidade em portais web; e também para a verificação do desempenho do *prompt* estabelecido, uma vez que estudos anteriores os *prompts* foram escritos em inglês.

Contexto: Você é um Especialista em Acessibilidade (Especialista WCAG) responsável por detectar violações da WCAG 2.2 em sites. Sua experiência é crucial para tornar a web mais acessível a todos. Não omita nenhuma questão. Após analisar detalhadamente todos os elementos, faça uma análise individualizada elemento por elemento. Não é necessário prévias ou exemplos, apresente o resultado final, detalhado e geral.

Necessidade Geral: Analise todo os elementos HTML, CSS, Javascript e outros, fornecidos para a avaliação completa e detalhada do critério [número] da WCAG 2.2. Realize análises estáticas.

Formato da Saída: Sua saída deve ser estruturada em arquivo único com a inspeção detalhada e completa do critério [número] da WCAG 2.2, no seguinte formato Json: Se houver violações, a resposta não deverá ter violações duplicadas (razões iguais para mesmo elemento) e ser no seguinte formato:

```

{
  "critério": "",
  "violação": "Sim",
  "quantidade": quantidade total,
  "elementos_e_razões_violados": [
    {
      "elemento": "",
      "razão": "",
      "recomendação": ""
    },
    {
      "elemento": "",
      "razão": "",
      "recomendação": ""
    }
  ]
}
Se não houver violações, a resposta deverá ser:
{
  "critério": "",
  "violação": "Não",
  "quantidade": 0,
  "elementos_e_razões_violados": []
}

```

Figura 1: Prompt utilizado para a avaliação individual de cada portal.

Testes prévios foram realizados para a definição do *prompt* utilizado. Observou-se que, ao fazer uso de um *prompt* com propósito geral, sem o estabelecimento de um critério de sucesso para a verificação, a LLM não conseguiu realizar o processo de avaliação por completo e, por vezes, apresentou alucinações [14]. Desta forma, foi necessário adotar procedimento semelhante ao de He et al. [8], com um *prompt* dedicado a cada critério de sucesso. Sendo assim, o campo [número] no *prompt* apresentado (Figura 1) foi preenchido com o número de cada critério de sucesso avaliado (total de 37 critérios).

Os critérios de sucesso verificados da WCAG 2.2 foram os mesmos abordados por He et al. [8]. A lista completa dos critérios utilizados está disponível em um repositório online¹¹.

⁶<https://wave.webaim.org/>

⁷<https://www.accessibilitychecker.org/>

⁸<https://copilot.microsoft.com>

⁹<https://www.maritaca.ai/>

¹⁰<https://www.llama.com/>

¹¹Confira o ReadMe em <https://github.com/luciocauper/Resultados-das-analises-realizadas>

Não houve a apresentação prévia de exemplos ou a utilização de recursos para o resgate da memória da LLM para a descrição da WCAG 2.2, diferente do trabalho de He et al. [8]. Portanto, utilizou-se um *prompt zero-shot*, junto a uma instrução “use temperatura = 0” para minimizar variação nas respostas¹², executado manualmente e aplicado sobre toda a página em HTML, e não apenas sobre trechos.

Ainda no processo de testes, algumas limitações da LLM foram evidenciadas, como: a execução de *scripts* para simular comportamentos que só aparecem após o carregamento de um conteúdo ou em resposta a eventos como cliques, foco ou rolagem; a simulação de dispositivos ou navegadores reais; e o carregamento de recursos como folhas de estilo. Tais características semelhantes também foram relatadas por He et al. [8] e estão presentes nas limitações das ferramentas de inspeção automática. Desta maneira, o processo de avaliação solicitado para a execução dos *prompts* à LLM foi uma análise estática.

A execução da inspeção dos portais foi realizada no mês de julho de 2025, sendo cada ferramenta de inspeção automática e cada *prompt* LLM utilizados uma única vez em cada portal. Os dados coletados foram tabulados em planilhas eletrônicas e estão disponíveis para visualização sumarizada por ferramenta¹³, e a análise e interpretação são descritas na próxima seção.

4 RESULTADOS E DISCUSSÕES

Pontua-se que a proposta desta pesquisa não foi realizar uma comparação quantitativa quanto ao tempo de execução da LLM estudada versus as ferramentas de inspeção utilizadas, ou quanto à quantidade de erros encontrados na LLM em comparação com cada ferramenta de inspeção utilizada, até porque cada ferramenta possui uma dinâmica diferente para apresentar e classificar a quantidade de erros ou problemas encontrados. Por vezes, há o agrupamento dos critérios de sucesso não atendidos, a exemplo do retorno da ferramenta WAVE, que associou os critérios 1.1.1, 1.3.1, 2.4.6 e 3.3.2 a um único problema (“Missing form label”).

Desta forma, a análise dos resultados aqui apresentada será voltada para o alcance de quais e quantos critérios foram indicados pela LLM, comparando-a aos que foram indicados pelas ferramentas de inspeção no processo de avaliação do mesmo conjunto de portais. Visto que sabe-se que tais ferramentas não contemplam todos os 86 critérios de sucesso presentes na WCAG, devido às suas limitações e devido também a alguns critérios da WCAG possuírem um nível de subjetividade [8].

Foram executados, para cada portal, 37 *prompts*, um para cada critério de sucesso avaliado, totalizando 481 execuções manuais. Verificou-se, contudo, que alguns critérios de sucesso não apresentaram violações quando analisados pela LLM e pelas ferramentas (Tabela 2). Sendo assim, os resultados obtidos (Tabela 3) apresentam os retornos advindos das execuções dos *prompts* no ChatGPT, que seguiram o modelo apresentado na Seção 3, e da execução das três ferramentas de inspeção automática aos 13 portais (Tabela 1). Os resultados revelam que todos os portais possuem violações ou problemas de não cumprimento de 4 a 8 critérios de sucesso quando

¹²A temperatura corresponde a um parâmetro que, quando próximo de zero, torna o modelo mais determinístico. Embora não possa ser configurada manualmente na interface do ChatGPT, sua inclusão no prompt orienta o modelo a produzir retornos mais estáveis, dessa forma, para este trabalho os valores foram próximos a zero.

¹³<https://github.com/luciocauper/Resultados-das-analises-realizadas>

Tabela 2: Critérios de sucesso não violados.

Critérios não violados	Evidências constatadas
1.4.2 Controle de áudio	Observou-se que os portais utilizados no estudo de caso não possuem áudios.
2.5.8 Tamanho do alvo (mínimo)	Os portais não apresentam falhas nas áreas de itens clicáveis, como botões e links.
3.1.1 Idioma da página	Todas as páginas possuam a definição do idioma no código HTML (<code>lang="pt-BR"</code>).
3.1.2 Idioma das Partes	Em nenhuma página foi verificado o uso de trechos em língua diferente do português.
3.2.2 Em entrada	Não foram identificados nos arquivos HTML elementos interativos que promovessem mudanças de contexto de forma automática.
3.2.5 Alteração Mediante Solicitação	A análise estática permitiu verificar eventos como <code>onchange</code> , <code>onmouseover</code> , <code>oninput</code> , entre outros, que poderiam promover mudanças de contexto. No entanto, não foram evidenciadas violações neste aspecto.

indicados pelas ferramentas de inspeção, e de 23 a 25 pela LLM. Observou-se o não cumprimento de dois critérios pontuados pelas ferramentas de inspeção automática para quase todos os portais, contudo não indicados pela LLM. Foram eles: 1.4.3 Contraste (Mínimo) e 2.4.1 Blocos de desvio. Para o primeiro, esperava-se tal limitação advinda da entrada que foi fornecida para a LLM, visto que esta não possuía acesso ao CSS da página. Já para o critério 2.4.1 a inspeção estática feita pela LLM se baseou apenas no HTML fornecido, a página não foi executada e não houve iteração com ela via teclado. A LLM conseguiu verificar apenas a presença de *landmarks* e *links* no HTML, mas não conseguiu validar o comportamento prático do desvio de blocos, que depende de teste com teclado ou emulador de leitor de tela. Apresentadas estas duas limitações pontuais esperadas, fatores positivos foram evidenciados nas saídas obtidas pela LLM, que identificou violações em 21 critérios não identificados pelas ferramentas de inspeção.

A exemplo do critério 2.5.3, que trata da correspondência entre o rótulo visível e o nome acessível de botões e links. Na Figura 2, é possível constatar que a LLM identificou um botão cujo rótulo visível apresenta o símbolo “x”. Essa configuração não atenderia plenamente ao critério de acessibilidade, visto que há uma divergência semântica entre o conteúdo visual (“x”) e o nome acessível (“Close”). Este critério 2.5.3 e outros foram detectados pela LLM com

```

{
  "critério": "2.5.3 - Rotulagem do None",
  "violação": "Sim",
  "quantidade": 1,
  "elementos_e_razões_violados": [
    {
      "elemento": "<button aria-label=\"Close\" class=\"offcanvas-close\"></button>",
      "razão": "O texto visível \"x\" não está contido no nome acessível \"Close\".",
      "recomendação": "Garanta que o nome acessível (por exemplo, aria-label) contenha o texto visível apresentado ao usuário."
    }
  ]
}
  
```

Figura 2: Evidências de saídas obtidas dos *prompts* executados na LLM.

uma ou mais violações, e não foram identificados pelas ferramentas de inspeção (Tabela 3). Observa-se que isto foi possível devido a uma característica presente nas LLMs e ausente nas ferramentas de inspeção automáticas: a capacidade de reconhecer equivalência semântica, uma evidência também pontuada por López-Gil and Pereira [10] e He et al. [8]. No caso deste critério 2.5.3, a LLM conseguiu identificar que não havia equivalência entre o rótulo visível e o nome acessível do elemento.

Tabela 3: Critérios de sucesso violados.

		Critérios de Sucesso Evidenciados																														
		1.1.1	1.3.1	1.3.2	1.3.3	1.3.4	1.3.5	1.4.1	1.4.3	1.4.4	1.4.5	1.4.6	1.4.8	1.4.9	1.4.10	1.4.11	1.4.12	2.2.1	2.2.2	2.4.1	2.4.2	2.4.4	2.4.5	2.4.6	2.4.8	2.4.9	2.4.10	2.5.3	2.5.5	3.1.4	3.3.2	4.1.2
[1]	AChecker	x						x											x	x	x								x	x		
	WAVE	x	x		x	x	x	x	x				x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x				
	LLM	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x				
[2]	AChecker	x						x											x	x	x							x	x			
	WAVE	x	x					x											x	x	x	x	x	x	x	x	x	x	x			
	IBM	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x				
[3]	AChecker	x	x					x											x	x	x	x	x	x	x	x	x	x	x	x		
	WAVE	x	x					x											x	x	x	x	x	x	x	x	x	x	x	x		
	LLM	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x				
[4]	AChecker	x	x					x											x	x	x	x	x	x	x	x	x	x	x	x		
	WAVE	x	x					x											x	x	x	x	x	x	x	x	x	x	x	x		
	IBM	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x				
[5]	AChecker	x	x					x											x	x	x	x	x	x	x	x	x	x	x	x		
	WAVE	x	x					x											x	x	x	x	x	x	x	x	x	x	x	x		
	LLM	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x				
[6]	AChecker	x	x					x											x	x	x	x	x	x	x	x	x	x	x	x		
	WAVE	x	x					x											x	x	x	x	x	x	x	x	x	x	x	x		
	IBM	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x				
[7]	AChecker	x	x					x											x	x	x	x	x	x	x	x	x	x	x	x		
	WAVE	x	x					x											x	x	x	x	x	x	x	x	x	x	x	x		
	IBM	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x				
[8]	AChecker	x	x					x											x	x	x	x	x	x	x	x	x	x	x	x		
	WAVE	x	x					x											x	x	x	x	x	x	x	x	x	x	x	x		
	IBM	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x				
[9]	AChecker	x	x					x											x	x	x	x	x	x	x	x	x	x	x	x		
	WAVE	x	x					x											x	x	x	x	x	x	x	x	x	x	x	x		
	IBM	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x				
[10]	AChecker	x	x					x											x	x	x	x	x	x	x	x	x	x	x	x		
	WAVE	x	x					x											x	x	x	x	x	x	x	x	x	x	x	x		
	IBM	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x				
[11]	AChecker	x	x					x											x	x	x	x	x	x	x	x	x	x	x	x		
	WAVE	x	x					x											x	x	x	x	x	x	x	x	x	x	x	x		
	IBM	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x				
[12]	AChecker	x	x					x											x	x	x	x	x	x	x	x	x	x	x	x		
	WAVE	x	x					x											x	x	x	x	x	x	x	x	x	x	x	x		
	IBM	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x				
[13]	AChecker	x	x					x											x	x	x	x	x	x	x	x	x	x	x	x		
	WAVE	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x				
	IBM	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x				

Desta forma, verificações como estas que dependem da identificação de significado e contexto, atualmente, são fatores limitadores para as ferramentas de inspeção automática, e que até então só poderiam ser realizadas por meio de testes manuais. Nesse sentido, observou-se nos resultados colhidos que a LLM apresentou-se como um bom recurso de apoio ao processo de verificação da acessibilidade de portais web ao utilizar *prompts* em português.

5 CONSIDERAÇÕES FINAIS

O presente trabalho trata-se de um estudo sobre as investigações do uso das LLMs como apoio ao processo de verificação da acessibilidade web. Para isto, foram utilizados como estudo de caso um conjunto de 13 portais das prefeituras que compõem a região dos Sertões dos Crateús. As páginas foram inspecionadas pelo ChatGPT-4, através de *prompts*, e pelas ferramentas Web Accessibility Evaluation Tools, Accessibility Checker e IBM Equal Access Toolkit.

Os resultados revelaram contribuições interessantes da aplicação de LLM como ferramenta de apoio ao processo de inspeção manual, dada a sua capacidade de interpretação de contextos. Sabe-se, até o momento, que este é o primeiro estudo a utilizar *prompts* escritos em português para esta finalidade. Desta forma, verifica-se o presente estudo como uma contribuição positiva que abre a possibilidade

de exploração futura do viés multilingue de LLMs na avaliação da acessibilidade web.

A fim de investigar melhor seus potenciais, para trabalhos futuros, elenca-se a possibilidade de: variação dos portais explorados, uma vez que o conjunto dos 13 portais das prefeituras pertence à mesma região e alguns apresentam padrões de layout, estilo, fontes e cores, o que pode ser considerado como um fator limitante desta pesquisa à abrangência da avaliação de ocorrências pontuadas pela LLM; investigações sobre o uso e a comparação de desempenho com outras LLMs; a avaliação de limitações e o aperfeiçoamento dos *prompts*, como o impacto da variação de temperatura ou de estilo do *prompt*; a verificação da qualidade se utilizar como entrada toda a página HTML (como feito aqui) ou partes delas, como fez He et al. [8] em sua pesquisa; para além de investigar o uso das LLMs na avaliação da acessibilidade utilizando as diretrizes de acessibilidade nacionais (e-MAG e NBR 17225:2025).

6 AGRADECIMENTOS

Esta pesquisa foi parcialmente apoiada pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), sob a bolsa com número 307831/2025-6 e pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes).

REFERÊNCIAS

- [1] Associação Brasileira de Normas Técnicas ABNT. 2025. ABNT NBR 17225:2025 – Acessibilidade em conteúdo e aplicações web - Requisitos. <https://www.abntcolecao.com.br/mpf/norma.aspx?ID=567818>
- [2] Danyillo Albuquerque, Gustavo Ponciano, João Pedro, Arlan Santos, Fabricio Silva, Caique Oliveira, Felipe Vasconcelos, and Thiago Honorato. 2024. Análise da Conformidade com Acessibilidade Digital: Um Estudo no Contexto dos Websites das Universidades Federais Brasileiras. In *Workshop de Computação Aplicada em Governo Eletrônico (WCGE)*. SBC, 133–144.
- [3] Yang Bai. 2019. The relationship between website accessibility and usability: an examination of us county government online portals. *Electronic Journal of e-Government* 17, 1 (2019), pp47–62.
- [4] Brasil. 2011. Lei n° 12.527, de 18 de novembro de 2011. Diário Oficial da União. http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm
- [5] Brasil. 2015. Lei n° 13.146, de 6 de julho de 2015. Diário Oficial da União. https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13146.htm
- [6] Governo Federal Brasil. 2023. Modelo de Acessibilidade de Governo Eletrônico (eMAG) – Versão 3.1.
- [7] Julielen A Dorneles and Amanda M Melo. 2024. Avaliação Preliminar da Acessibilidade de Sites de Prefeituras dos Municípios em que estão situados os Campi da Unipampa. In *Escola Regional de Engenharia de Software (ERES)*. SBC, 70–78.
- [8] Ziyao He, Syed Fatiul Huq, and Sam Malek. 2025. Enhancing Web Accessibility: Automated Detection of Issues with Generative AI. *Proceedings of the ACM on Software Engineering* 2, FSE (2025), 2264–2287.
- [9] Wenjun Lin, Bamikole Adewale, Min Li, Mahreen Nasir, Ajmery Sultana, Rashid Hussain Khokhar, and Yue Zhang. 2025. Dynamic Web Page Modification for Accessibility Using AI and Large Language Models. In *Computer and Information Science and Engineering: Volume 18*. Springer, 33–46.
- [10] Juan-Miguel López-Gil and Juanan Pereira. 2025. Turning manual web accessibility success criteria into automatic: an LLM-based approach. *Universal Access in the Information Society* 24, 1 (2025), 837–852.
- [11] João Marcelo Alves Macedo, Valdecir Becker, Felipe Melo Feliciano de Sá, Edvaldo Vasconcelos da Rocha Filho, and Daniel de Queiroz Cavalcanti. 2025. Avaliação de sistemas de Governo Aberto e de transparéncia pública nas capitais brasileiras. *Cadernos Gestão Pública e Cidadania* 30 (2025), e90832.
- [12] Rafael Pardini, João Bárbara, Henrique Scheid, Adriano CM Pereira, Wagner Meira Jr, Reinaldo Ferraz, and Beatriz Rocha. 2021. Observatório da acessibilidade da web brasileira. In *Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia)*. SBC, 71–74.
- [13] Cleber Cristiano Prodanov and Ernani Cesar De Freitas. 2013. *Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico-2ª Edição*. Editora Feevale.
- [14] G Pradeep Reddy, YV Pavan Kumar, and K Purna Prakash. 2024. Hallucinations in large language models (LLMs). In *2024 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream)*. IEEE, 1–6.
- [15] Ather Sharif, Sanjana Shivani Chintalapati, Jacob O Wobbrock, and Katharina Reinecke. 2021. Understanding screen-reader users' experiences with online data visualizations. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–16.
- [16] Sudesh Sheoran, S Mohanasundaram, R Kasilingam, and Sanket Vij. 2023. Usability and accessibility of open government data portals of countries worldwide: an application of TOPSIS and entropy weight method. *International Journal of Electronic Government Research (IJEGR)* 19, 1 (2023), 1–25.
- [17] World Wide Web Consortium W3C. 2023. Web Content Accessibility Guidelines (WCAG) 2.2.
- [18] Chaozheng Wang, Zongjie Li, Cuiyun Gao, Wenxuan Wang, Ting Peng, Hailiang Huang, Yueting Deng, Shuai Wang, and Michael R Lyu. 2024. Exploring multi-lingual bias of large code models in code generation. *arXiv preprint arXiv:2404.19368* (2024).