INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO
MESTRADO EM ENGENHARIA INFORMÁTICA




Estudo exploratório sobre aplicações de Modelos de Linguagem de Larga Escala em acessibilidade web

LEANDRO BARBOSA DE SOUSA
Junho de 2025

POLITÉCNICO
DO PORTO

ISEP INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO

Estudo exploratório sobre aplicações de Modelos de Linguagem de Larga Escala em acessibilidade *web*

Leandro Barbosa de Sousa

Dissertação para obtenção do Grau de Mestre em
Engenharia Informática
Área de Especialização em Engenharia de Software

Orientadora: Isabel Azevedo (IFP)

Porto, junho 2025

Declaração de Integridade

Declaro ter conduzido este trabalho académico com integridade.

Mais declaro que não plagiei ou apliquei qualquer forma de uso indevido de informações ou falsificação de resultados ao longo do processo que levou à sua elaboração.

Portanto, o trabalho apresentado neste documento é original e de minha autoria, não tendo sido utilizado anteriormente para nenhum outro fim. As exceções estão explicitamente reconhecidas na secção "Considerações éticas" do primeiro capítulo. Esta secção também declara como as ferramentas de IA foram utilizadas e para que finalidade.

Declaro ainda que tenho pleno conhecimento do Código de Conduta Ética do P. PORTO.

ISEP, Porto, 25 de junho de 2025

Leandro Barbosa de Sousa

Resumo

O presente documento investiga a aplicação de *Large Language Models* (LLMs), especificamente do Google Gemini 2.0 Flash, na remediação de problemas de acessibilidade *web* identificados pela ferramenta AXE, com foco nas diretrizes WCAG 2.2.

Com base em resultados promissores evidenciados em estudos anteriores, o potencial dos LLMs é reconhecido como agentes facilitadores na promoção da acessibilidade digital. Este estudo analisa, de forma empírica, a eficácia dos LLMs na resolução de quatro áreas críticas: contraste de cor, atributos ARIA, ausência de texto alternativo/geração de descrições para conteúdo não textual, e correção de hiperligações sem texto discernível, sendo a avaliação conduzida a partir da análise e remediação semiautomática dos 100 websites mais populares do mundo.

Ao reunir dados quantitativos e qualitativos sobre o comportamento do modelo perante diferentes desafios de acessibilidade, este trabalho procura ainda contribuir para o debate sobre a viabilidade do uso de LLMs como ferramentas de apoio à acessibilidade web, explorando tanto o seu potencial como as suas limitações práticas.

Palavras-chave: *Large Language Models*, Acessibilidade *web,* WCAG 2.2

Abstract

This paper investigates the application of Large Language Models (LLMs), specifically Google Gemini 2.0 Flash, in the remediation of web accessibility issues with a focus on WCAG 2.2 guidelines, identified by the AXE tool.

Based on promising results demonstrated in previous studies, the potential of LLMs is recognized as facilitating agents in the promotion of digital accessibility. This study empirically analyses the effectiveness of LLMs in resolving four critical areas: colour contrast, ARIA attributes, lack of alternative text/generation of descriptions for non-text content, and correction of hyperlinks without discernible text, with the evaluation being conducted based on the analysis and semi-automatic remediation of the 100 most popular websites in the world.

By gathering quantitative and qualitative data on the behaviour of the model regarding different accessibility challenges, this work seeks to contribute to the debate on the feasibility of using LLMs as tools to support web accessibility, exploring both their potential and their practical limitations.

Keywords: Large Language Models, Web Accessibility, WCAG 2.2

Agradecimentos

Gostaria de expressar a minha mais sincera gratidão ao Instituto Superior de Engenharia do Porto, instituição que é para mim uma segunda casa e que me albergou e acolheu tão bem durante os últimos 6 anos. Aqui tive oportunidade criar imensas memórias que guardo com grande estima, e conhecer várias pessoas com diferentes personalidades, interesses e individualidades, que não só me ajudaram a crescer profissionalmente como também pessoalmente.

Em especial, gostaria de destacar o meu agradecimento à docente Isabel Azevedo, cuja orientação, dedicação e exigência, foram fundamentais para a elaboração do presente trabalho e sempre motivaram a melhoria constante do mesmo.

Finalmente, um agradecimento especial à minha família e amigos, em especial à Erica de Jesus, pelo apoio incondicional, pela paciência e pela motivação diária, sendo uma das principais razões para a minha ambição e motivação em ser mais e melhor todos os dias.

Índice

Lista de Figuras

Lista de Tabelas

Lista de Extratos de Código

Abreviaturas

ARIA	*Accessible Rich Internet Applications*
CAPTCHA	*Completely Automated Public Turing test to tell Computers and Humans Apart*
CoT	*Chain of Thought*
IA	Inteligência Artificial
IPP	Instituto Politécnico do Porto
LLM	*Large Language Model*
POUR	*Perceivable, Operable, Understandable, Robust*
PRISMA	*Preferred Reporting Items for Systematic reviews and Meta-Analyses*
SVG	*Scalable Vector Graphics*
SEO	*Search Engine Optimization*
SC	*Success Criterion*
SMART	*Specific, Measurable, Achievable, Realistic, and Timely*
SLR	*Systematic Literature Review*
SMR	*Systematic Mapping Review*
SRU	*Screen Reader User*
SVG	*Scalable Vector Graphics*
TIC	Tecnologias da Informação e Comunicação
URL	*Uniform Resource Locator*
WCAG	*Web Content Accessibility Guidelines*

1 Introdução

O presente capítulo visa introduzir o tema do projeto, as suas motivações, e inclui uma exposição geral do contexto do trabalho, objetivos, e considerações éticas.

1.1 Contexto do Trabalho

A evolução da *world wide web* desencadeou uma transformação digital progressiva que levou empresas a migrar produtos, serviços, e operações, para plataformas digitais, com o objetivo de se tornarem mais modernas, melhorarem os seus desempenhos, e aumentarem o seu alcance de mercado. Face a estas mudanças, o acesso a conteúdos e recursos foi também afetado, surgindo novos problemas no que toca à acessibilidade dos mesmos.

Ao aceder a Tecnologias de Informação e Comunicação (TIC), a falta de oportunidades iguais pode levar à exclusão. Acessibilidade *web* é a prática da mitigação destas mesmas exclusões, nomeadamente de pessoas com algum tipo de deficiência, temporária ou permanente, assegurando que nenhuma barreira é imposta a qualquer utilizador que aceda o conteúdo de *websites* [10].

Conteúdo *web* acessível é uma preocupação global que conta com várias iniciativas, diretrizes, e políticas impostas, e requer habilitações especializadas, tempo, e muitos outros recursos. À luz da investigação conduzida, importa salientar a Diretiva Europeia (UE) n.º 2016/2102 [44], que estabelece os requisitos de acessibilidade dos *sites* e aplicações *web* dos organismos do setor público, bem como o Decreto-Lei n.º 83/2018, de 19 de outubro, que procede à sua transposição para o ordenamento jurídico nacional. Ambas estas políticas visam garantir a acessibilidade digital em entidades públicas, respetivamente num âmbito europeu e nacional, e constituem instrumentos normativos de caráter obrigatório. No seu incumprimento, estabelece-se, de acordo com o Decreto-Lei n.º 83/2018, artigo 13, capítulo 4 [45], uma infração legal, qualificada como uma prática discriminatória. Estas diretivas integram ainda a Lei Europeia da Acessibilidade [46], complementando-a ao abranger uma vasta variedade de produtos e serviços do setor privado, e contribui para a harmonização dos requisitos de acessibilidade em toda a União Europeia [5].

Atualmente, o *Web Content Accessibility Guidelines* (WCAG) é um *standard* internacional amplamente adotado, que abrange uma vasta gama de recomendações e práticas, e que auxilia o desenvolvimento de conteúdos *web* mais acessíveis [52]. Neste contexto, os *Large Language Models* (LLMs) podem apresentar-se como um agente relevante no combate à exclusão digital, através da aplicação do seu conhecimento nestas diretrizes para a remediação de violações de acessibilidade.

Sistemas providos de LLM têm vindo a ganhar uma crescente popularidade devido às suas capacidades promissoras no apoio a desenvolvedores em tarefas especializadas [14]. Ao tirar

proveito das capacidades destes modelos, viabiliza-se a remediação automática de erros de acessibilidade, o que poderá reduzir significativamente o tempo necessário para a sua correção, outrora dependente de abordagens manuais.

1.2 Problema

A *WebAIM* é uma organização sem fins lucrativos sediada nos Estados Unidos, dedicada à promoção da acessibilidade *web.* Em 2024, e pelo sexto ano consecutivo, o seu relatório anual sobre a acessibilidade das páginas iniciais dos 1 000 000 *websites* mais visitados do mundo foi publicado [39], disponibilizando métricas como a média de erros detetados por página, grau de complexidade das interfaces e conformidade com as diretrizes WCAG 2.2. Neste, verifica-se que, em média, cada página apresentava 51 erros de acessibilidade, sendo que 96% dos erros detetados pertencem a seis principais categorias, descritas na Tabela 1.

Tabela 1 - Falhas mais comuns do WCAG 2 em páginas iniciais (Fonte: 39)

TIPO DE VIOLAÇÃO WCAG	% DAS PÁGINAS
Texto com baixo contraste	81%
Ausência de texto alternativo em imagens	54.5%
Rótulos de entrada de formulário ausentes (*form input labels*)	48.6%
Links vazios	44.6%
Botões vazios	28.2%
Ausência da linguagem da página	17.1%

Praticamente inalterados à data, estes são os erros mais frequentemente identificados em *websites* nos últimos 5 anos, à exceção da ausência da linguagem da página, que tem vindo a diminuir ao longo do tempo. Na Figura 1, é possível observar a evolução das diversas violações desde 2019.

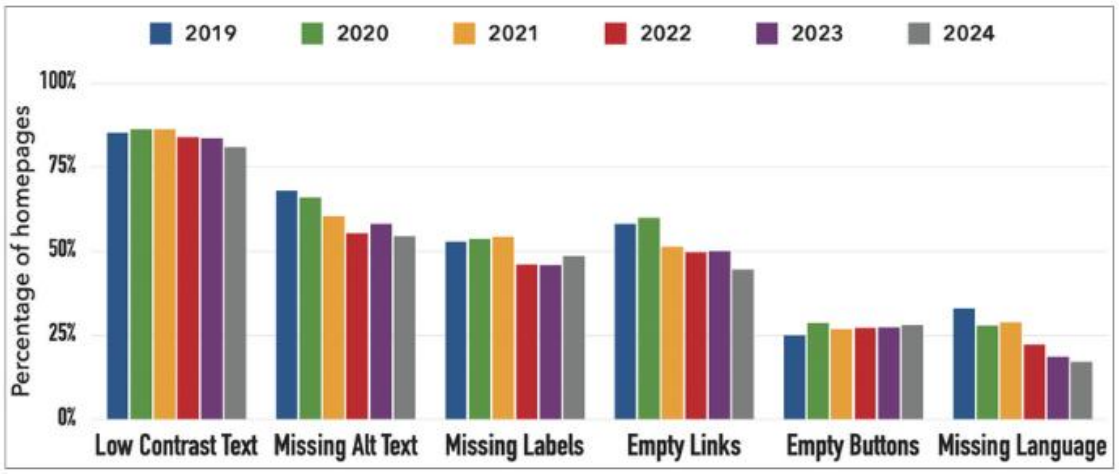


Figura 1 - Percentagem de homepages afetadas pelos principais problemas de acessibilidade de 2019 a 2024 (Fonte: [[39]])

O estudo realizado por Utkarsha Singh [34] direciona a sua investigação ao contexto educacional através da comparação das principais diferenças de acessibilidade em *websites* educacionais de países em desenvolvimento (BRICS) e desenvolvidos (G7). Os resultados obtidos neste estudo apontam para quatro principais categorias de violação: contraste de cor, texto alternativo, links, e elementos sem nomes acessíveis, coincidindo com os mesmos padrões de não conformidade identificados no relatório anual da WebAIM. A recorrência destes problemas em contextos geográficos e funcionais distintos sublinha a necessidade de soluções escaláveis e automáticas, que possam apoiar a remediação eficaz destes obstáculos e promover uma *web* mais inclusiva.

1.3 Objetivos e abordagem

Este estudo tem como finalidade avaliar a competência de *Large Language Models,* no que toca à resolução semiautomática das principais barreiras de acessibilidade, identificadas pela ferramenta AXE, designadamente problemas relacionados com: Contraste de cor, Ausência de papeis, nomes e valores, Conteúdo não-textual e *Links*.

Estes problemas estão relacionados, respetivamente, com os critérios de sucesso 1.4.3 - *Contrast (Minimum) (Level AA)* e 1.4.6 - *Contrast (Enhanced) (Level AAA)*, 4.1.2 - Name, Role, Value (Level A), 1.1.1 - *Non-text Content (Level A)*, e 2.4.4 - *Link Purpose (In Context) (Level A)* e 2.4.9 - *Link Purpose (Link Only) (Level AAA)* da versão 2.2 das diretrizes WCAG, estando os dois últimos critérios relacionados com *links*.

Nas experiências especificas a desenvolver, o modelo Gemini 2.0 Flash da Google [15] será considerado tendo por base critérios especificados na secção 4.1 deste documento. Diversas

instâncias de erro são abordadas pelo LLM com vista a sua remediação, cujo sucesso é ditado pela ausência do erro após a correção e sem o acréscimo de novos problemas. No que toca à resolução de problemas relativos a conteúdo não textual em particular, a simples geração de um texto alternativo, independentemente da sua pertinência, é suficiente para compor um cenário de sucesso. No entanto, isto não constitui uma remediação bem-sucedida pelo que, para além da ausência de erro sem acréscimo de novos, o texto gerado deve seguir os padrões utilizados por especialistas, nomeadamente em termos de brevidade e detalhe apropriado para o tipo de imagem em questão.

Durante o processo, dados empíricos são ainda obtidos e armazenados de forma sistemática, nomeadamente o tipo de erro identificado e o grau de sucesso da solução proposta, os quais fundamentam a análise estatística e discussão dos resultados.

De forma a alcançar os objetivos previamente definidos, foram formuladas quatro questões de investigação:

- RQ 1: Podem os LLMs resolver eficazmente problemas de contraste de cor segundo as diretrizes WCAG 2.2?
    - Motivação

      Apresentando-se como o problema de acessibilidade mais frequentemente identificado em páginas *web,* a eficácia dos LLMs na remediação de violações deste tipo é investigada, assim como a conformidade com o WCAG 2.2, nomeadamente o critério de sucesso 1.4.3 - *Contrast (Minimum)* - e 1.4.6 - *Contrast (Enhanced)* - do mesmo.

- RQ 2: Podem os LLMs resolver eficazmente problemas relacionados com atributos 'name', 'role' e 'value' (ARIA)?
    - Motivação

      O fornecimento de informações sobre nome, papel e valor, em todas as interfaces do utilizador, garante a compatibilidade com tecnologias assistivas, como leitores e ampliadores de ecrã, e software de reconhecimento de voz, frequentemente utilizado por pessoas com debilidades. Com base na ausência destes atributos numa amostra significativa de *websites* e associados a violações do critério de sucesso 4.1.2 - *Name, Role, Value (Level A)* – é investigado neste trabalho se os LLMs podem ser aplicados na resolução destes problemas.

- RQ 3: Podem os LLMs ser capazes de gerar corretamente descrições alternativas ('alt text') para imagens, de acordo com os padrões usados por especialistas?
    - Motivação

      A falta de textos alternativos, designadamente em imagens, compõe um dos maiores obstáculos de acessibilidade identificados por utilizadores de ferramentas assistivas. No âmbito do trabalho, é investigada a eficácia dos LLMs na solução de problemas relacionados com a ausência de textos

alternativos em imagens, assim como a qualidade dos textos gerados face aos padrões utilizados por especialistas, nomeadamente em termos de brevidade e detalhe apropriados para o tipo de imagem em questão.

- RQ 4: Podem os LLMs resolver eficazmente problemas de *links* vazios ou sem texto discernível, garantindo conformidade com boas práticas de acessibilidade?
    - Motivação

      Links vazios ou sem texto discernível apresentam-se, à semelhança das violações mencionadas anteriormente, como um dos problemas de acessibilidade mais frequentes em conteúdo *web*, dificultando a navegação sobretudo em páginas com vastas cargas de informação. No contexto do presente trabalho, é investigada a eficácia dos LLMs na remediação destas violações, destacando os critérios:
        - 2.4.4: *Link Purpose (In Context) (Level A)*,
        - 2.4.9 - *Link Purpose (Link Only) (Level AAA)*.

1.4 Considerações Éticas

Como estudante do Instituto Politécnico do Porto (IPP), todas as considerações e decisões encontram-se alinhadas e ao abrigo do código de conduta do IPP [43], garantindo o respeito pela dignidade individual, legalidade, transparência, e imparcialidade, sob o artigo 2º, número 1, e visando a prossecução do interesse público (número 2). O artigo 6º menciona os deveres dos estudantes, referindo nas alíneas a) a d), aspetos relativos ao respeito e zelo da propriedade física e intelectual do P. Porto, e o respeito pela sua comunidade nas restantes alíneas.

No âmbito do mestrado em engenharia informática, o código de conduta e ética do IEEE [47, 48] também foi considerado pela sua relevância dentro do ramo de estudo. Dito isto, princípios como manter os mais altos padrões de integridade, através do comportamento responsável e conduta ética no exercício de atividades profissionais, mencionado no número 1 do código de ética, como o tratamento respeitoso e justo, indicado no código de conduta foram seguidos.

1.4.1 Utilização de ferramentas de Inteligência Artificial

Durante a elaboração do presente documento, ferramentas generativas foram utilizadas, designadamente o *ChatGPT*, modelo de geração de linguagem em larga escala da *OpenAi*. Aplicado ao para-fraseamento e melhoria da legibilidade e clareza do discurso, os textos gerados foram revistos, corrigidos, e validados, de forma a garantir a precisão e qualidade do conteúdo. Todos os textos encontrados no presente documento são da autoria do autor, não havendo texto gerado por IA (Inteligência Artificial).

Ademais, a própria natureza do trabalho obrigou à utilização de ferramentas identificadas ao longo do documento, pelo que constituem componentes importantes para a elaboração do mesmo.

2 Planeamento

O presente capítulo é composto pelos subcapítulos “Gestão de competências” e “Planeamento e controlo do projeto”.

No primeiro, é realizado um levantamento de competências-chave para o desenvolvimento do trabalho, seguido de uma comparação entre as capacidades atuais do autor e as proficiências requeridas para o sucesso do trabalho. Com isto, pretende-se identificar lacunas e áreas de melhoria por meio de um autodiagnóstico de competências, permitindo a identificação de direções estratégicas para o aprimoramento das mesmas.

O subcapítulo “Planeamento e controlo do projeto” visa apresentar a metodologia a adotar, e identificar as fases do desenvolvimento e organização do trabalho.

2.1 Gestão de competências

Como foi mencionado anteriormente, o presente subcapítulo tem como objetivo o desenvolvimento das competências mais relevantes para a elaboração da dissertação, sendo realizado um diagnóstico e comparação das aptidões atuais do autor. Este processo auxilia a identificação de competências a desenvolver, posteriormente integradas num plano de ação que visa a melhoria das mesmas.

2.1.1 Identificação de competências

No que toca ao desenvolvimento do trabalho, várias competências revelam-se cruciais para o sucesso do mesmo, destacando-se:

- **Análise e resolução de problemas**

  A análise e resolução de problemas é a base para lidar com cada etapa da elaboração de uma dissertação, desde a formulação do problema à apresentação de conclusões. Esta competência é relevante pois permite a identificação precisa de questões importantes, a estrutura e desenvolvimento de uma argumentação sólida, e a interpretação e análise de dados de maneira crítica.

- **Gestão de tempo**

  A gestão de tempo é uma competência essencial dado que ela potencializa o cumprimento de prazos, facilita o planeamento, organiza o processo de pesquisa, e reduz o stress, favorecendo o equilíbrio entre a vida pessoal e a vida académica.

- **Comunicação escrita e discurso técnico**

A comunicação escrita é fundamental para transformar uma boa pesquisa numa dissertação que seja acessível, compreensível e persuasiva, aumentando a qualidade e a credibilidade do trabalho final.

Na Tabela 2, encontram-se analisadas as competências acima mencionadas face à proficiência, impacto, assim como algumas estratégias para o seu desenvolvimento.

Tabela 2 - Análise de Competências

Competência	Proficiência requerida (0-5)	Proficiência atual (0-5)	Impacto para o sucesso do projeto	Estratégias de desenvolvimento de competências
Análise e resolução de competências	5	4	Alto	Desenvolver a capacidade de identificação de problemas. Melhorar a capacidade de análise critica.
Gestão de tempo	5	3	Alto	Criar um cronograma realista para atividades. Estabelecer um sistema de revisão e ajuste do planeamento.
Comunicação escrita	5	3	Alto	Melhorar a clareza e a objetividade escrita. Expandir o vocabulário e repertório de expressões escritas.

2.1.2 Definição de objetivos

A definição de objetivos é essencial pois visa dar clareza sobre o que se pretende alcançar e estabelecer um propósito claro para o diagnóstico, definindo o nível desejado para a competência, assim como os resultados que são pretendidos alcançar com o desenvolvimento da mesma.

Para garantir uma definição clara de metas, a metodologia *Specific, Measurable, Achievable, Realistic, and Timely* (SMART) foi adotada, que garante que todos os objetivos definidos são

específicos, mensuráveis, alcançáveis, e realistas, características estas que contribuem para o cumprimento dos mesmos.

Assim, os seguintes objetivos foram definidos para cada uma das competências identificadas:

- **Análise e resolução de problemas**
    - Desenvolver a capacidade de identificar e definir problemas de maneira clara e precisa, praticando com situações reais ou simuladas, pelo menos uma vez por semana durante os próximos três meses;
    - Melhorar a capacidade de análise crítica, através da prática da avaliação de causas e efeitos para problemas específicos, duas vezes por mês, durante os próximos três meses.
- **Gestão de tempo**
    - Estabelecer um cronograma detalhado para os próximos três meses do projeto, definindo metas intermediárias semanais e prazos para cada etapa;
    - Implementar um sistema de revisão diário e semanal do planeamento para ajustar metas e prioridades, ao longo de três meses.
- **Comunicação escrita**
    - Escrever textos claros e objetivos, livres do uso de frases longas e palavras desnecessárias na escrita de documentos nos próximos três meses;
    - Ampliar o vocabulário e enriquecer a comunicação escrita, através da aprendizagem e aplicação de pelo menos cinco novas palavras ou expressões por semana, durante três meses.

2.1.3 Plano de ação

Face às competências previamente identificadas, o Anexo 4 - Plano de ação foi desenvolvido. Este dispõe um plano de ação que tem como objetivo o desenvolvimento pessoal do autor no que toca às aptidões referidas, por meio da definição de objetivos, ações, e indicadores de desempenho.

2.2 Planeamento e controlo do projeto

Neste subcapítulo é feita, como indica o título, a apresentação do planeamento e controlo do projeto. Aqui é introduzida a estratégia "*Design and Creation*" e metodologia *Agile,* utilizadas para guiar o desenvolvimento do projeto, os artefactos gerados fruto do uso destas, incluindo a calendarização e entregáveis esperados.

2.2.1 Estratégia e metodologia

De forma a desenvolver um projeto focado no cumprimento dos objetivos gerais definidos de forma eficiente, escalável, e alinhados com as necessidades dos *stakeholders,* uma estratégia e metodologia foram definidas.

A primeira está assente na abordagem "*Design and Creation*", que enfatiza o desenvolvimento iterativo e a resolução de problemas de forma criativa, colocando igual importância em ambas as fases de design, que envolve a idealização e prototipagem, e criação, onde os conceitos são materializados através do desenvolvimento iterativo e melhoria contínua.

A metodologia *Agile* foi escolhida para implementar a estratégia "*Design and Creation*", e é uma abordagem baseada nos valores e princípios descritos no manifesto *Agile* [31], denominados como *Agile Values* e *Agile Principles,* criada para tornar o desenvolvimento de *software* mais flexível e adaptável. Esta é particularmente adequada para este projeto dado que promove a flexibilidade, colaboração e melhoria contínua, permitindo à equipa responder à mudança de requisitos de forma eficaz, incorporando várias metodologias especificas, tendo estas o propósito de entregas frequentes e atempadas.

Dito isto, a metodologia *Agile* consiste principalmente nas seguintes características:

- Divisão do desenvolvimento em iterações curtas, chamados "*sprints*", que podem durar entre 1 e 4 semanas.
- Comunicação e feedback constante efetuado no final de cada sprint.
- Flexibilidade e agilidade no ajuste de requisitos.

Ainda nesta matéria, o Anexo 1 – Procedimentos de monitorização e controlo foi elaborado, onde são discutidas e comparadas esta e outras metodologias candidatas, identificadas face à estrutura, flexibilidade, foco no cliente, gestão de riscos, e ciclo de desenvolvimento.

2.2.2 Calendarização e definição de entregáveis

Face ao plano do projeto, é feita uma divisão do trabalho por etapas, estando estas identificadas na Figura 2 - Timeline do Projeto, e contribuindo desta forma para o cumprimento dos objetivos definidos. Durante cada uma destas fases, é feito o acompanhamento semanal do progresso de desenvolvimento com vista a gestão e ajuste regular das tarefas a desenvolver.

Figura 2 - Timeline do Projeto

Seguidamente, encontra-se na Figura 3, o diagrama *Work Breakdown Structure* (WBS) correspondente ao trabalho, composto por 5 fases, e que inclui o entregáveis definidos, assim como as tarefas para os obter. Em complemento, o Anexo 2 - Diagrama de Gantt do planeamento do projeto e o Anexo 3 - Dicionário EAP / WBS foram desenvolvidos de forma a fornecer mais contexto sobre cada passo do desenvolvimento.

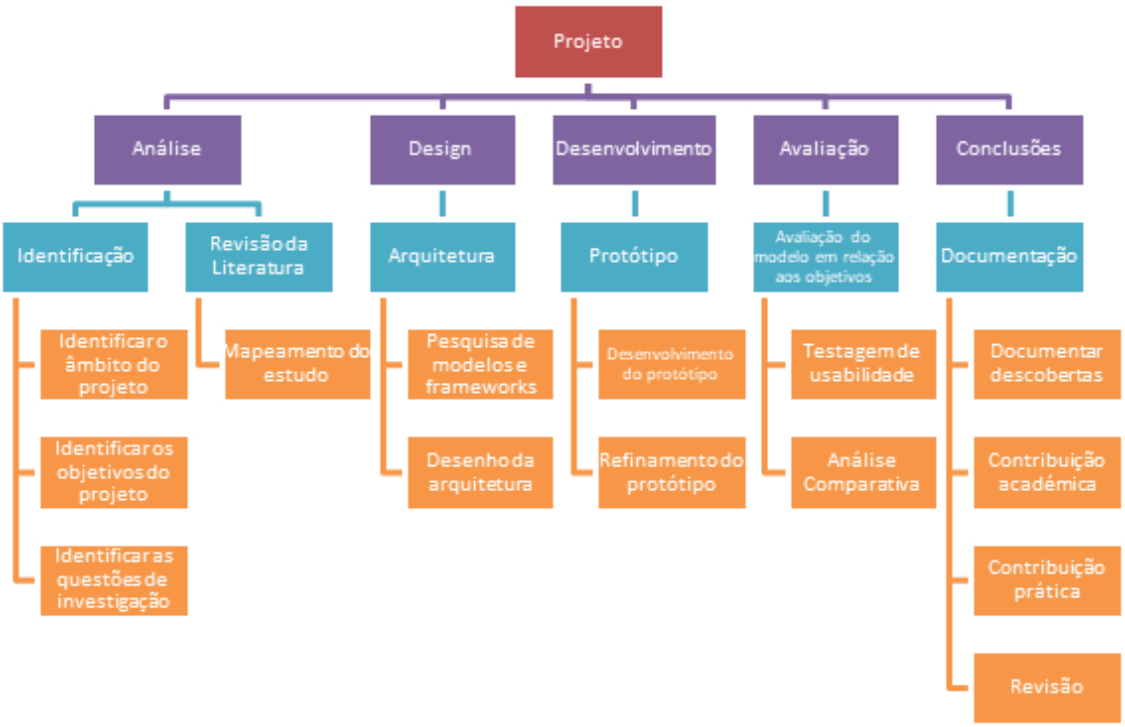


Figura 3 - Diagrama WBS

2.2.3 Gestão de riscos

Neste subcapítulo é feito um levantamento das ameaças em cada uma das etapas do projeto, sendo posteriormente feita a análise e resposta.

2.2.3.1 Identificação

Face ao percurso de desenvolvimento do projeto, as seguintes ameaças foram identificadas:

- **Complexidade tecnológica;**

A complexidade tecnológica surge devido a fatores como desconhecimento sobre a tecnologia, problemas de integração, falta de experiência, ou rápidas mudanças tecnológicas. Estes podem necessitar de tempo acrescido para o estudo e resolução de problemas, o que, por sua vez, pode atrasar entregas.

- **Gestão de tempo**

Uma pobre gestão de tempo pode levar a falhas nas entregas, entregáveis apressados e insatisfação do cliente devido à diminuição da qualidade geral do projeto, apresentando-se como um risco significativo se não for gerido apropriadamente.

- **Orçamento**

Devido ao orçamento restritivo, algumas escolhas associadas ao LLM podem ver-se fora de alcance, dado que muitos dos modelos presentes no mercado atual englobam custos, ainda mais se o *fine-tuning* for realizado. Esta limitação levanta várias barreiras no desenvolvimento do presente trabalho pelo que este aspeto é abordado.

2.2.3.2 Análise

De forma a fazer uma boa avaliação dos riscos identificados, nesta secção será conduzida uma análise dos mesmos segundo uma matriz de risco [17], presente na Tabela 3, sendo averiguado e utilizado o impacto e a probabilidade de cada risco para determinar a sua criticidade.

Tabela 3 - Criticidades dos riscos identificados

DESCRIÇÃO	PROBABILIDADE (P)	IMPACTO (I)	CRITICIDADE (P * I)
Complexidade tecnológica	3	5	15
Gestão de tempo	2	4	8
Orçamento	4	4	16

2.2.3.3 Respostas

Para concluir a componente de gestão de riscos, nesta secção é feito o levantamento das ações que dão resposta aos riscos identificados, por forma a evitar ou minimizar o impacto de cada ameaça.

Complexidade tecnológica

- Simplificar a complexidade – Decomposição do problema em componentes mais pequenos e simples. Através desta divisão, é possível trabalhar sobre um sistema modular onde cada parte pode ser desenvolvida e testada individualmente, evitando abordar o problema todo de uma vez.
- Identificar alternativas – Mesmo simplificando a problemática ou desenvolvendo capacidades sobre a tecnologia, esta poderá não ser integrada e aplicada. Deste modo, devem ser identificadas alternativas para o desenvolvimento do projeto.

Gestão de tempo

- Criação de uma *timeline* realista – através da estimativa realista e minimamente pessimista, é possível traçar um programa fiável e que tenha em conta possíveis atrasos.

- Monitorização constante do progresso – através de ferramentas de gestão de projetos, como o Jira e diagramas de Gantt, é possível acompanhar o desenvolvimento diário do projeto, permitindo a ação imediata face a atrasos.
- Identificação e endereçamento de dependências de tarefas – Ao identificar e endereçar dependências antecipadamente é possível minimizar atrasos provenientes de fatores externos.

Orçamento

- Escolha de modelos com planos gratuitos – vários LLMs encontram-se disponíveis para uso e integração em soluções de forma gratuita graças a planos sem custos ou pelo seu cariz *open-source*.

  No caso de modelos *open-source*, estes têm necessidade de vastos recursos computacionais que dificilmente são encontrados fora de um ambiente dedicado, pelo que, apesar de se apresentarem como gratuitos, os recursos necessários para os utilizar acarretam custos substanciais.

  No caso de modelos com planos gratuitos, estes normalmente restringem a velocidade de processamento, como é exemplo o plano gratuito do Gemini 2.0 Flash, porém não colocam entraves ao seu uso, viabilizando um projeto sem ou com poucos custos.

3 Revisão da Literatura

De forma a assegurar a qualidade dos estudos incluídos no trabalho, uma revisão da literatura foi conduzida. Das várias técnicas de revisão propostas, como o *snowballing* [40, 41] e *Systematic Literature Review* (SLR) [16], o *Systematic Mapping Review* (SMR) [27] foi escolhido. Devido ao seu carater e escopo exploratório, o mesmo traça com maior abrangência o panorama geral da área de investigação, enfatizando a amplitude da pesquisa ao invés da sua profundidade, uma vez que prioriza a categorização e mapeamento dos estudos.

Numa breve introdução, o SMR baseia o seu processo na construção de questões de investigação alinhadas com os objetivos do projeto, das *queries* posteriormente usadas e conjugadas com critérios de elegibilidade previamente definidos em fontes eletrónicas, com vista a reunir os estudos que apresentassem relevância e adequação ao propósito da revisão.

Por fim, as diretrizes *Preferred Reporting Items for Systematic Reviews and Meta-Analyses* (PRISMA) [21] foram adotadas para promoção da transparência, replicabilidade e credibilidade da revisão.

3.1 Fontes

A Tabela 4 apresenta as fontes de onde foram extraídos os estudos utilizados na revisão da literatura.

Tabela 4 - Fontes utilizadas

Fonte	*Uniform Resource Locator* (URL)
ACM *Digital Library*	https://dl.acm.org/
ScienceDirect	https://www.sciencedirect.com/
ResearchGate	https://www.researchgate.net/

A fonte primária utilizada foi a libraria digital *ACM Digital Library,* sendo a grande maioria dos estudos identificados provenientes desta. De forma a garantir uma análise mais abrangente e credível, alternativas foram consultadas com o objetivo de reduzir o enviesamento e considerar diferentes perspetivas e abordagens sobre o tema.

3.2 Termos de pesquisa

Os termos de pesquisa, usados na *query* de pesquisa durante a extração de dados são:

- *Large-Language Models*
- LLM

- Acessibilidade Web
- *Web Content Accessibility Guidelines* (WCAG)
- Texto alternativo
- Contraste de cor
- "*Name, value, role*" / Atributos ARIA
- *Links*

Com estes, a seguinte *query* de pesquisa foi criada:

- *AllField:("Large-Language Models" OR "LLM" OR "Large Language Models") AND AllField:("web accessibility" OR "WCAG" OR "web content accessibility guidelines") AND AllField:("Code Quality Issues" OR "alt text" OR "alternative text" OR "color contrast" OR "link" OR "name, value, role" OR "ARIA attributes")*

3.2.1 Critérios de Elegibilidade

Por forma a evitar tendências e subjetividades, procedeu-se à definição de critérios de seleção. Assim, critérios de inclusão (I) e exclusão (E), foram aplicados durante o mapeamento da literatura, identificados na Tabela 5:

Tabela 5 - Critérios de Elegibilidade de inclusão (I) e exclusão (E) para os estudos da literatura

Nr.		Critério
1	I	Estudos disponíveis na íntegra
2	I	Estudos que envolvem a aplicação de LLMs aplicados à acessibilidade *web*
3	E	Procedimentos, trabalhos em progresso, pequenos resumos ou posters, não descartando as suas referências

3.2.2 Extração de dados

A Ilustração 1 apresenta o fluxograma PRISMA, representativo do fluxo de informações em diferentes etapas de uma revisão da literatura, e retrata o processo de extração e inclusão de estudos.

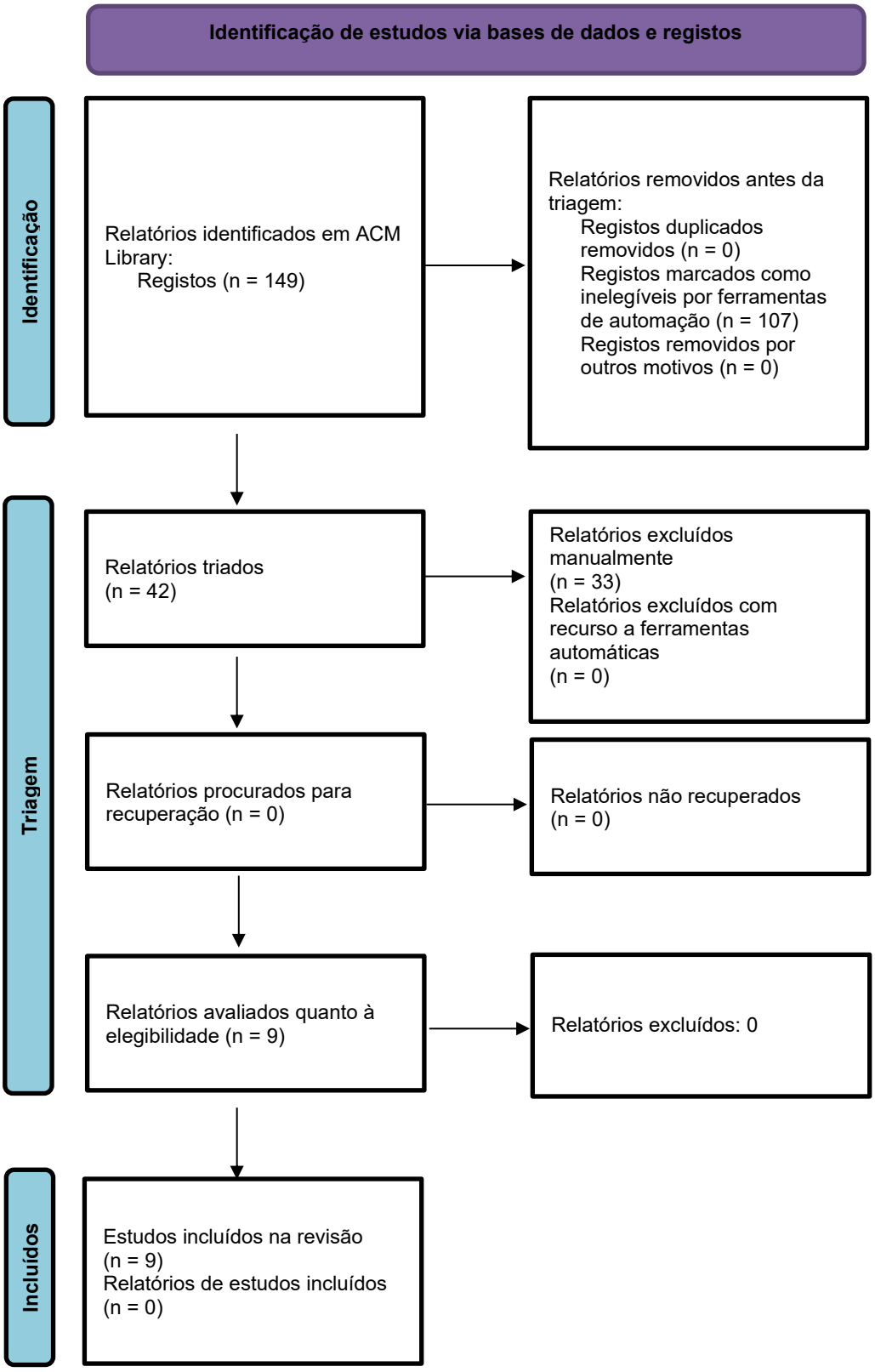


Ilustração 1 - Mapeamento do estudo

- **Identificação**: Aqui estão agregados todos os artigos obtidos das bibliotecas digitais, resultado da aplicação da *query* de pesquisa e que cumprem parcialmente os critérios de elegibilidade.
- **Triagem**: Composta por duas partes. Na primeira fase, é feita uma filtragem dos artigos através da sua exclusão conforme o seu alinhamento com os critérios de elegibilidade definidos na Tabela 5. Na fase seguinte, os restantes artigos são analisados de forma mais cuidada, avaliando se estes contêm conteúdo relevante para responder às questões de investigação, assim como capacidade de sustentar estas mesmas respostas através de argumentos válidos.
- **Inclusão**: Os estudos aqui agregados são incluídos na revisão e identificados na Tabela 6.

Tabela 6 - Estudos identificados

Nome	Referência
Accessibility Analysis of Educational Websites Using WCAG 2.0	[34]
Caption Anything: Interactive Image Description with Diverse Multimodal Controls	[33]
CodeA11y: Making AI Coding Assistants Useful for Accessible Web Development	[26]
Does ChatGPT Generate Accessible Code? Investigating Accessibility Challenges in LLM-Generated Source Code	[36]
Enhancing Accessibility in Software Engineering Projects with Large Language Models (LLMs)	[37]
FigurA11y: AI Assistance for Writing Scientific Alt Text	[23]
Fostering websites accessibility: A case study on the use of the Large Language Models ChatGPT for automatic remediation	[2]
From Provenance to Aberrations: Image Creator and Screen Reader User Perspectives on Alt Text for AI-Generated Images	[19]
ImageExplorer Deployment: Understanding Text-Based and Touch-Based Image Exploration in the Wild	[3]

3.3 Discussão

Realizada a análise dos estudos, esta secção tem o intuito de responder às questões de investigação identificadas, fundamentadas com os dados obtidos nos documentos incluídos na revisão.

3.3.1.1 RQ 1: Podem os LLMs resolver eficazmente problemas de contraste de cor segundo as diretrizes WCAG 2.2?

No âmbito da resolução de problemas de contraste de cor insuficiente, as avaliações conduzidas apontam para resultados promissores através da geração de código HTML que vai ao encontro das *guidelines* WCAG 2.2. Com base nos estudos analisados [36, 2, 37], o LLM *ChatGPT*, alvo de análise nos mesmos, demonstrou ser uma ferramenta eficaz na remediação de problemas de acessibilidade previamente identificados, capaz de resolver corretamente uma panóplia de erros, como a ausência de texto alternativo, contraste insuficiente, e texto de hiperligações não descritivo.

Na Figura 4 encontra-se um levantamento dos dados relativos à remediação de diversas violações de acessibilidade encontradas em código gerado pelo próprio LLM. Destes dados, destaca-se o item número 8 e 9, que dizem respeito à vertente de contraste.

No.	Principle	Success Criteria	Level	Guideline	Violation Type	# Violation	# Fixed	# Fixed%
1	Perceivable	1.1.1	A	Non-text Content	img element missing alt attribute	56	56	100%
2	Perceivable	1.1.1	A	Non-text Content	Image used as anchor is missing valid alt text	2	2	100%
3	Perceivable	1.3.1	A	Info and Relationships	Input element, type of "text", has no text in label	62	46	74%
4	Perceivable	1.3.1	A	Info and Relationships	Input element, type of "text", missing an associated label	69	56	81%
5	Perceivable	1.3.1	A	Info and Relationships	Input element, type of "checkbox", has no text in label	4	0	0%
6	Perceivable	1.3.1	A	Info and Relationships	Input element, type of "file", has no text in label	5	2	40%
7	Perceivable	1.3.1	A	Info and Relationships	Label text is empty for select element	6	6	100%
8	Perceivable	1.3.2	A	Meaningful Sequence	The contrast between the color of text and its background for the element is not sufficient to meet WCAG2.0 Label AA	38	32	84%
9	Perceivable	1.4.3	AA	Contrast (Minimum)	Very low contrast	232	139	60%
10	Perceivable	1.4.4	AA	Resize Text	I (italic) element used	277	273	99%
11	Operable	2.4.2	A	Page Titled	title element is empty	1	1	100%
12	Operable	2.4.4	A	Link Purpose (In Context)	Anchor contains no text	54	35	65%
13	Operable	2.4.6	AA	Headings and Labels	Header nesting	3	1	33%
14	Understandable	3.1.1	A	Language of Page	Language missing or invalid	5	5	100%
15	Understandable	3.1.1	A	Language of Page	Document language not identified	4	4	100%
16	Understandable	3.1.1	A	Language of Page	Document has invalid language code	5	5	100%
17	Understandable	3.3.2	A	Labels or Instructions	Empty Button	7	4	57%
18	Understandable	3.3.2	A	Labels or Instructions	Empty form label	11	11	100%
19	Understandable	3.3.2	A	Labels or Instructions	Missing form label	14	14	100%
20	Understandable	3.3.2	A	Labels or Instructions	select element missing an associated label	6	6	100%
21	Understandable	3.3.2	A	Labels or Instructions	Label text is empty	6	6	100%
22	Robust	4.1.1	A	Parsing	id attribute is not unique	1	1	100%

Figura 4 - Número de violações da diretriz WCAG 2.2 detetadas por ferramentas de verificação de acessibilidade em código gerado pelo ChatGPT e número de violações corrigidas pelo mesmo (Fonte: [36]).

Na Figura 5 estão presentes o número de violações de acessibilidade encontradas em projetos de código aberto extraídos do GitHub, e o número destas que o *ChatGPT* foi capaz de resolver. Ainda nesta figura, cada violação foi associada ao seu respetivo princípio e diretriz WCAG, obtendo-se as seguintes taxas de sucesso:

- *Perceivable* – 70%
- *Operable* – 64.4%
- *Understandable* – 70%
- *Robust* – 89.2%

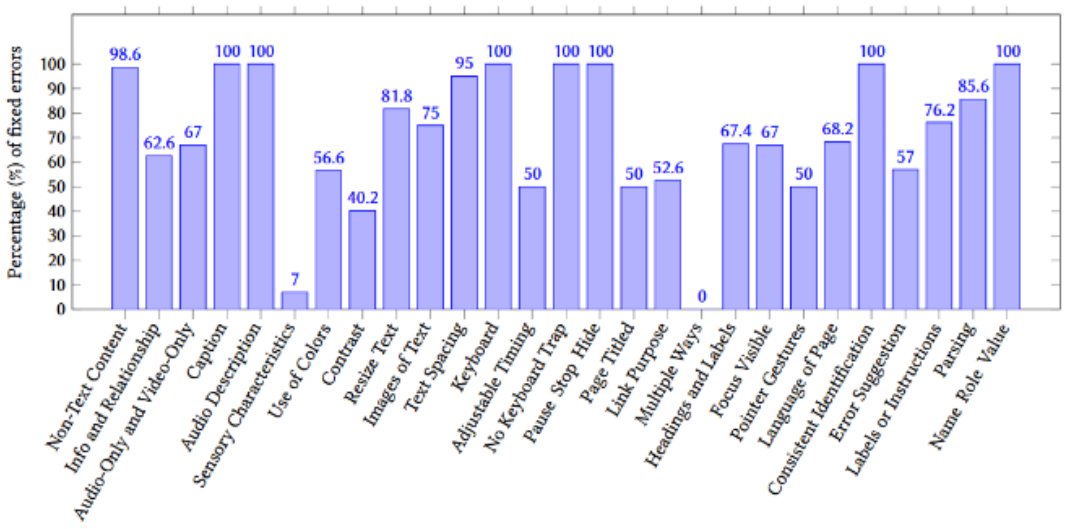


Figura 5 - Taxa de correção do ChatGPT para erros de acessibilidade em código-fonte obtido de projetos *open-source* (Fonte: [33]).

Apesar dos resultados promissores, avaliações subjetivas, como por exemplo a eficácia do contraste de cor, pode ser visto como um desafio para esta ferramenta, o que torna métodos de teste automáticos inviáveis, exigindo uma testagem sobretudo manual.

No que toca à diretriz WCAG em si, a taxa de sucesso para a correção de problemas relacionados com contraste é de 40.2%. Em relação à remediação de inconformidades presentes no código gerado pelo próprio LLM, a taxa de sucesso ronda os 60 a 84%. No geral, e considerando as suas limitações, dados apontam para uma taxa de sucesso moderada na correção de problemas de acessibilidade identificados, revelando ser uma área que requer desenvolvimento adicional.

No estudo conduzido por P. Mowar [26], a aplicação de LLMs na construção de conteúdos *web* acessíveis como uma ferramenta para desenvolvedores, contribuiu significativamente para a melhoria da acessibilidade, designadamente na garantia do contraste de cor para diferentes estados de botões, por meio de sugestões de código.

3.3.1.2 RQ 2: Podem os LLMs resolver eficazmente problemas relacionados com atributos 'name', 'role' e 'value' (ARIA)?

O fornecimento de informações sobre nome, papel e valor em todas as interfaces do utilizador garante a compatibilidade com tecnologias assistivas, como leitores e ampliadores de ecrã, e software de reconhecimento de voz, frequentemente utilizado por pessoas com debilidades. Os *Accessible Rich Internet Applications* (ARIA) são um conjunto de atributos que melhoram a acessibilidade de conteúdos e aplicações *web* através da definição destes mesmos papéis,

estados e propriedades dos elementos das interfaces, especialmente úteis em componentes dinâmicos, como modais e menus suspensos, cujo desenvolvimento não se restringe apenas a HTML padrão e à aplicação de estilos.

Com base nos valores apresentados na Figura 4, conclui-se que o critério '*Name, Role, Value'* apresenta uma alta taxa de remediação por parte do LLM, tendo sido capaz de resolver na íntegra todas as violações detetadas nesta componente.

3.3.1.3 RQ 3 – Podem os LLMs ser capazes de gerar corretamente descrições alternativas ('alt text') para imagens de acordo com os padrões usados por especialistas?

LLMs e modelos multimodais têm sido utilizados para a geração de melhores descrições de imagens [33, 3]. Graças às suas capacidades de geração de texto variada e de alta qualidade, estas ferramentas tornam-se bastante úteis e eficazes nestas tarefas.

O trabalho realizado por Deyao Zhu [33] aborda a geração de descrições visuais detalhadas através do uso dois LLMs, o ChatGPT e BLIP-2, proficientes em áreas distintas. O projeto denominado *ChatCaptioner,* conta com um mecanismo automático de questões baseado na capacidade *zero-shot* do ChatGPT para efetuar uma série de perguntas informativas sobre imagens, às quais o modelo BLIP-2, um *visual-language model* proficiente na interpretação de imagens, responde. No final, todo o contexto é reunido e processado de forma a produzir uma descrição detalhada da imagem.

Com base nos dados obtidos neste estudo, aproximadamente 80% das legendas geradas foram consideradas corretas por parte de especialistas na área. Dentro da restante percentagem, resultados apontam responsabilidades ao BLIP-2, tendo sido apenas capaz de responder corretamente a 67% das perguntas colocadas. Perante uma questão cuja resposta levanta incertezas, este revela ainda tendência para alucinar, deste modo corrompendo a qualidade dos resultados. De forma a prevenir esta situação, o fator de incerteza foi introduzido, que previne que, quando deparado com perguntas sem resposta ou situações de dúvida, o BLIP-2 responda honestamente que desconhece ou que não sabe. Com a implementação deste novo fator, verificou-se uma descida de cerca de 15% das respostas erradas, porém, uma pequena porção de questões permanecem incorretas.

Enquanto os estudos anteriores abordam imagens simples, o trabalho realizado por Nikhil Singh [23] abrange imagens complexas como gráficos, frequentemente presentes em estudos científicos. Com base nos resultados, a assistência interativa desenvolvida neste projeto contribuiu para a composição de textos alternativos melhores, mais longos e detalhados.

Numa outra perspetiva, imagens geradas por IA têm vindo a proliferar como uma nova forma de *media.* Maitraye Das [19] aborda no seu estudo textos alternativos para este novo formato, inerentemente inacessível devido à falta de geração de texto alternativo. Com a ajuda de diversos criadores de imagens geradas por IA e utilizadores de leitores de ecrã, a análise

conduzida destaca as perspetivas dos criadores sobre o texto alternativo, expõe as necessidades dos *Screen Reader Users* (SRUs) no que toca a este, e discute os desafios e oportunidades de usar os *prompts* usados para a geração da imagem na construção de textos alternativos.

No geral, a utilização de LLMs aplicados à geração de textos alternativos revela-se uma ferramenta útil e que contribui para uma maior eficácia, qualidade e produtividade.

3.3.1.4 RQ 4 –Podem os LLMs resolver eficazmente problemas de links vazios ou sem texto discernível, garantindo conformidade com boas práticas de acessibilidade?

A quarta violação de acessibilidade mais frequente em conteúdo *web* encontra-se no conteúdo de hiperligações (*links*), e impacta diretamente a navigabilidade das páginas para utilizadores de tecnologias assistivas, tornando confusa a sua experiência.

No estudo conduzido por Wajdi Aljedaani [36], esta questão é abordada. Com base nos dados apresentados na Figura 4 e 5, o ChatGPT apresenta uma taxa de sucesso modesta na resolução de problemas ligados a *links* sob a diretiva *Link Purpose* (52.6% - 65%). Uma possível explicação para estes valores surge do cariz subjetivo da tarefa pelo que se apresenta como um desafio à performance dos modelos de linguagem.

Um exemplo prático várias vezes encontrado em conteúdo *web* dinâmico, é a utilização de *placeholders.* Estes criam um *link* sem um destino concreto, utilizado muitas vezes por desenvolvedores para acionar um *script JavaScript* que afeta o comportamento da página. Na perspetiva do LLM, a ausência de texto descritivo pode ser detetada, porém esta situação não é facilmente remediada, pois não é possível determinar se a rotulagem do *link* é necessária ou se o contexto visual é suficiente. Neste aspeto, a solução poderá passar pela sugestão de alterações para uma melhor acessibilidade, como o uso de *role="button"* ou *href="ID-da-secção"*. Não obstante, caso o LLM não for capaz de determinar o propósito real do *link,* poderão ser propostas alterações erróneas e sem sentido.

4 Análise e Design

O presente capítulo demarca o processo de análise e desenho da aplicação desenvolvida. Estando dividido em duas partes, é, numa primeira fase, abordado o processo de seleção do modelo a utilizar. Posteriormente é abordada a componente arquitetural segundo o modelo 4+1 [25] que a apresenta sob diferentes perspetivas e níveis de granularidade.

4.1 Escolha do *Large-Language Model*

As métricas de qualidade desempenham um papel essencial na avaliação, comparação e melhoria dos modelos linguísticos. No contexto do desenvolvimento de LLMs, estas métricas asseguram a eficácia e a fiabilidade, sendo cruciais para conduzir escolhas informadas e fundamentadas.

Com base no objetivo central deste projeto, a seleção do LLM reveste-se de particular importância. Para garantir uma escolha consciente e adequada, diversos modelos foram analisados e comparados com base em diferentes métricas de desempenho e fatores.

Tabela 7 - Métricas-chave de avaliação de qualidade

Métrica	*Benchmark*
Conhecimento e Raciocínio (Reasoning & Knowledge)	MMLU (Massive Multitask Language Understanding)
Conhecimento e Raciocínio Científico (Scientific Reasoning & Knowledge)	GPQA (General Purpose Question Answering)
Raciocínio Quantitativo (Quantitative Reasoning)	MATH
Codificação (Coding)	HumanEval
Comunicação (Communication)	LMSys Chatbot Arena ELO Score
Matemáticas (Maths)	MGSM (Mathematics Grade School Math)

A Tabela 7 apresenta as principais métricas de qualidade utilizadas para avaliar modelos em diferentes áreas:

- **MMLU** - Conhecimento geral e raciocínio entre diferentes matérias [12]
- **GPQA** - Raciocínio científico profundo e compreensão [24]
- **MATH** - Raciocínio quantitativo e lógico [30]
- **HumanEval** - Proficiência em escrita e compreensão de código [7]
- **LMSys Chatbot Arena ELO Score** - Comunicação eficaz e envolvente nas conversas [8]
- **MGSM** - Habilitações matemáticas fundamentais para propósitos educacionais [13]

4.1.1 Análise de modelos

A proficiência na escrita e compreensão de código, que identifica os modelos mais capazes de interpretar e gerar conteúdo programático correto, é a métrica mais relevante a ter em consideração na escolha do modelo. Face ao objetivo de remediação de problemas de acessibilidade identificados através do uso de LLMs, é importante que o modelo tenha proficiência suficiente para interpretar o código fornecido, e que seja capaz de o corrigir ou de gerar alternativas de forma correta. Assim, é com base no *HumanEval* que a escolha do modelo irá concentrar-se principalmente.

Durante a análise feita nesta secção, 15 modelos foram comparados, considerados à data da escrita do presente documento, os melhores no mercado no *benchmark HumanEval*.

Ademais, o presente trabalho foi concebido com o intuito de ser tecnicamente robusto e economicamente viável em diferentes contextos orçamentais, pelo que a relação custo-benefício é considerada. Assim, todas as opções apresentadas ao longo do documento têm em consideração critérios de eficiência de custos, procurando eliminar ou reduzir despesas desnecessárias sem comprometer a qualidade e funcionalidade da solução proposta. Neste âmbito, modelos de acesso livre foram privilegiados, que minimizam a dependência de recursos dispendiosos, garantindo a sustentabilidade económica do trabalho e o acesso por parte de um publico mais abrangente, nomeadamente instituições e utilizadores com recursos limitados.

Por fim, a latência, ou seja, o tempo de resposta entre a solicitação e a entrega da resposta, é um fator crítico na avaliação do desempenho de LLMs em contextos aplicacionais. Modelos com menor latência, usados como agentes conversacionais, oferecem uma experiência mais imersiva e natural, ao passo que latências elevadas introduzem atrasos percetíveis entre as ações do utilizador e as respostas do modelo. Além disso, a latência está intrinsecamente relacionada com a capacidade de o modelo lidar com múltiplos pedidos simultaneamente sem degradação de performance, essencial para cenários de larga-escala com alto tráfego. Assim, a minimização da latência torna-se fundamental não só para otimizar a satisfação do utilizador, mas também garantir a escalabilidade e robustez de aplicações interativas em tempo real.

4.1.1.1 *HumanEval*

Na Figura 6 estão presentes os resultados das avaliações conduzidas pela Artificial Analisys [1], onde foram efetuadas testes qualitativos de forma independente em todos os *endpoints* de cada LLM, sendo o *Artificial Analysis Quality Index* a média dos resultados destas mesmas avaliações segundo uma estratégia *zero-shot prompting* [18]. Ao contrário do que acontece com outras técnicas, onde são fornecidos exemplos ou que se revolvem na refinação contínua de *prompts* com base nos *outputs* obtidos, esta abordagem classifica os diferentes modelos com base sua própria habilidade de compreensão de tarefas e quantidade de dados utilizados no seu treino, sendo uma estratégica de avaliação flexível e que elimina a necessidade de *task-specific fine-tuning*.

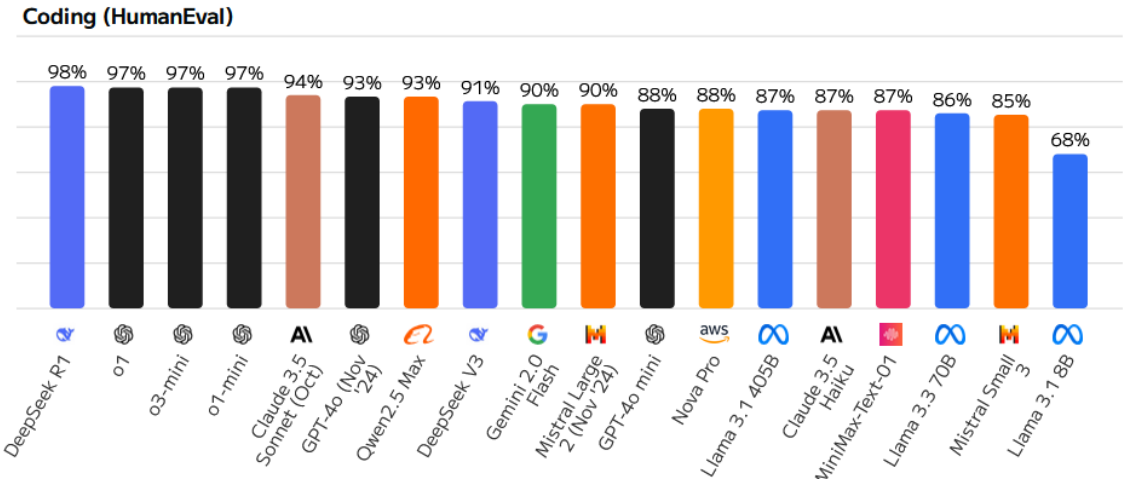


Figura 6 - *Benchmark HumanEval* dos diferentes modelos (Fonte: *ArtificialAnalisys.ai* [1])

Tabela 8 - Modelos IA mais bem qualificados (Fonte: ArtificialAnalisys.ai [1])

Modelo de IA	Criador(a)	HumanEval
DeepSeek R1	Deepseek	98%
o1	OpenAi	97%
o3-mini	OpenAi	97%
o1-mini	OpenAi	97%
Claude 3.5 Sonnet	Anthropic	94%
GPT-4o	OpenAi	93%
Qwen2.5 Max	Qwen (Alibaba)	93%
DeepSeek V3	DeepSeek	91%
Gemini 2.0 Flash	Google	90%
Mistral Large 2	Mistral	90%
GPT-4o mini	OpenAI	88%
Nova Pro	AWS	88%
Llama 3.1 405B	Meta	87%
Claude 3.5 Haiku	Anthropic	87%
MiniMax-Text-01	MiniMax	87%

Os resultados apresentados na Tabela 8 são, por si só, insuficientes para justificar a adequação de qualquer modelo. Além de um bom desempenho, é necessário considerar a relação custo-benefício, que privilegia os modelos que combinam performance e acessibilidade, bem como a latência, que impacta diretamente o tempo-útil de resposta.

Nesse sentido, nas subsecções seguintes é conduzida uma análise comparativa focada nestes aspetos-chave.

4.1.1.2 Relação Custo-benefício

A Figura 7 ilustra a comparação dos principais modelos de IA identificados, organizando-os de acordo com o seu desempenho e custo por milhão de *tokens* processados. Com este paralelo, é possível obter dados relevantes sobre como os diferentes modelos posicionam-se face à sua relação custo-benefício.

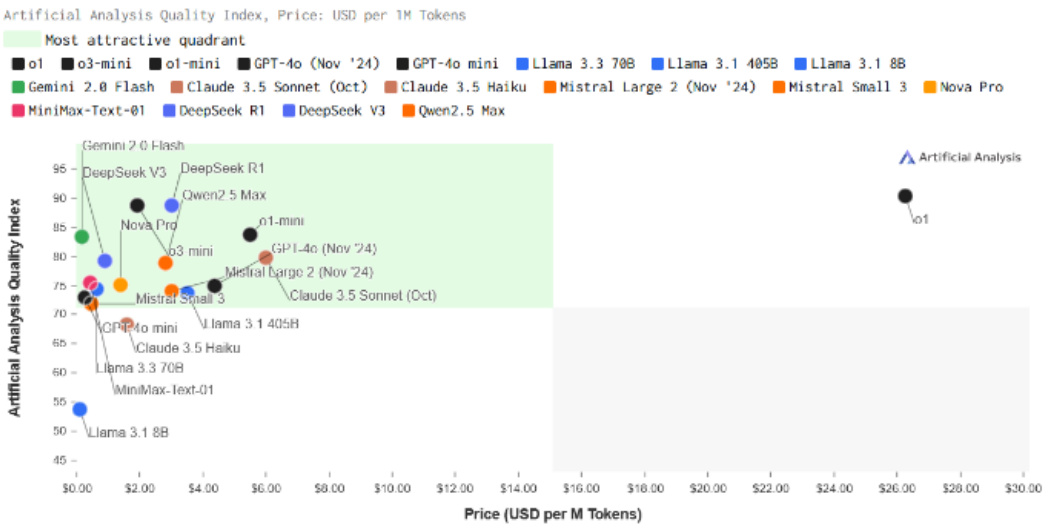


Figura 7 - Relação custo-benefício dos modelos IA por 1M de tokens (Fonte: *ArtificialAnalisys.ai* [1])

Fazendo uma interpretação dos dados, no primeiro quadrante estão presentes os modelos com o melhor desempenho face à sua acessibilidade. Destes destaca-se o *Gemini 2.0 Flash* da *Google*, o *o3-mini* da *OpenAi* e o *DeepSeek R1*, como principais líderes.

4.1.1.3 Latência

A latência, ou seja, os segundos demorados até o primeiro *token* ser recebido após o pedido à API ser enviado, é também uma métrica relevante a analisar, pois influencia diretamente o

tempo de resposta, composto pela combinação da latência de ida (*request*) e a latência de volta (*response*) [20].

Os resultados presentes na Figura 8 oferecem uma perspetiva que contrasta com os dados da Figura 7. Em termos de custo-benefício, os modelos *o3-mini* e *DeepSeek R1* lideram, no entanto, estes modelos apresentam os piores resultados no que diz respeito à latência.

Dos três melhores modelos custo-benefício, o Gemini 2.0 Flash mantém-se destacado registando uma latência de 320 milissegundos - o terceiro mais valor mais baixo - a 30 milissegundos do modelo mais rápido.

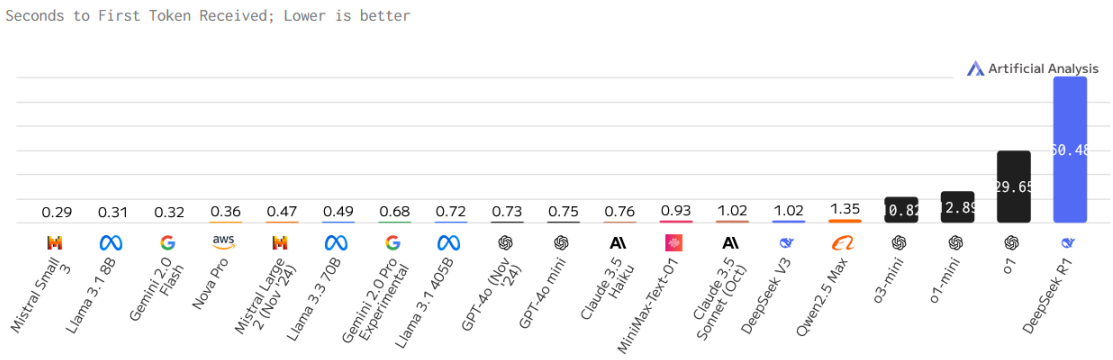


Figura 8 - Latência em modelos IA (Fonte: ArtificialAnalisys.ai [1])

4.1.1.4 Conclusão

A análise efetuada nas subsecções anteriores proporciona uma visão abrangente sobre o atual panorama dos LLM no que diz respeito à sua proficiência na compreensão e geração de código, relação custo-benefício, e latência.

Tendo em consideração o enquadramento do projeto, destaca-se a relevância do *benchmark HumanEval*, que avalia a capacidade de um modelo compreender e gerar código de forma eficaz. Embora forneça uma orientação rudimentar sobre os modelos mais promissores para integração na solução, por si só, não é suficiente para fundamentar uma escolha definitiva. Por este motivo, uma análise mais pormenorizada em diferentes áreas, designadamente a relação custo-benefício e o tempo de resposta, é conduzida em 4.1.1.2 e 4.1.1.3, respetivamente.

No que toca a custo-benefício, a Figura 7 coloca frente a frente a performance e o preço dos diferentes LLM, distribuindo-os de acordo. Nesta avaliação, conclui-se que, apesar de certos modelos apresentarem grandes desempenhos, estes vêm acompanhados com maiores custos - exemplificado pelo modelo o1 da OpenAi - surgindo alternativas mais acessíveis, como o *Gemini 2.0 Flash*, o modelo *o3-mini,* e o *DeepSeek R1*, que lideram esta componente.

Estes três últimos modelos destacam-se pela sua performance e acessibilidade e podem ser considerados boas alternativa para integração na solução, com os modelos *o3-mini* e *DeepSeek R1* a apresentarem desempenhos e custos superiores quando comparados com o *Gemini 2.0 Flash*. Devido a esta igualdade, os dados apresentados na Figura 8 servem de critério de desempate. Durante a pesquisa efetuada, conclui-se que o tempo de resposta é bastante diferente entre os modelos devido à latência. Nos modelos com melhor performance, a latência é superior a dez segundos e, no caso do *DeepSeek R1,* acima de cinquenta segundos, que se traduz num tempo de resposta entre um e dois minutos.

Em conclusão, com os dados obtidos nesta análise, a escolha do *Google Gemini 2.0 Flash* é justificada, sendo adotado para integração na solução como um componente integral e crucial para o desenvolvimento do projeto.

4.2 Design

Esta secção foca-se exclusivamente na arquitetura da solução, apresentada de forma abrangente, em diferentes pontos de vista e níveis de granularidade, sustentada pelo modelo 4+1.

4.2.1 Lógica de design da solução

No que toca à aplicação implementada, esta é composta por duas componentes principais: Analisador e API.

No primeiro, é utilizado um *handler* que despoleta os casos de uso e que toma partido dos módulos *axe-webdriverjs* e *selenium-webdriver* para a análise das páginas *web*. Aqui é feito o levantamento dos problemas de acessibilidade, sendo posteriormente enviados para o *backend,* juntamente com o código inconforme e componentes relevantes.

A *API RESTful* baseia-se numa arquitetura em camadas e atua como intermediário na comunicação entre o analisador e o modelo IA. No exercício das suas funções, este realiza ainda qualquer tratamento necessário dos dados recebidos para fazer as comunicações entre as duas partes.

4.2.2 Vista lógica

A arquitetura lógica sustenta os requisitos funcionais, tirando partido de princípios de abstração, encapsulamento e herança, e decompõe o sistema em abstrações-chave para facilitar a análise funcional, a identificação de mecanismos comuns, e elementos de *design* entre as várias partes do sistema.

Neste contexto, a abordagem *Rational/Booch* [49] foi utilizada para representar a estrutura lógica da solução, através de diagramas de classes/componentes para ilustrar cada parte do sistema e as suas relações lógicas com detalhe. A Figura 9 segue esta mesma notação e traça com pormenor as relações dos diversos componentes do sistema.

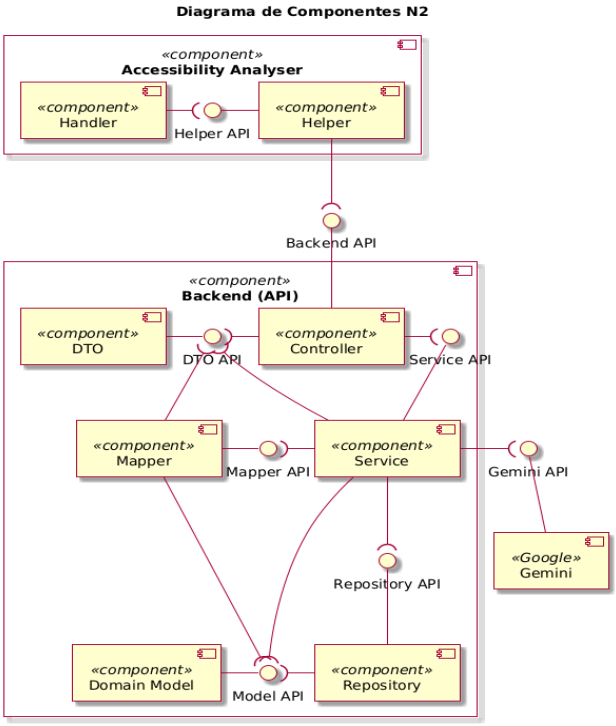


Figura 9 - Nível 2: Vista de Lógica – Diagrama de Componentes

Já mencionado anteriormente, o sistema desenvolvido é composto por dois componentes principais: o Analisador de Acessibilidade e o *Backend*. No desenho destes, os princípios SOLID [28] foram considerados com o intuito de criar um código robusto, elegante e sustentável.

A API, responsável pela receção dos pedidos enviados pelo analisador, serve de intermediário entre este e o modelo IA. A arquitetura escolhida para o desenvolvimento desta componente tem em mente os princípios da programação orientada a objetos [42], e é baseada em camadas que fomentam um desenvolvimento e testagem fácil.

No que diz respeito ao domínio desenvolvido, este foi concebido de forma a facilitar a sistematização das conclusões obtidas durante o trabalho, e está intrinsecamente associado com o processo de remediação. Neste contexto, são armazenados dados acerca da página analisada, violações detetadas, e soluções propostas, facilitando a posterior consulta dos dados e o sustento das fundamentações apresentadas.

Na Figura 10 está presente o modelo adotado, constituído por dois componentes principais e o terceiro, resultado da sua relação.

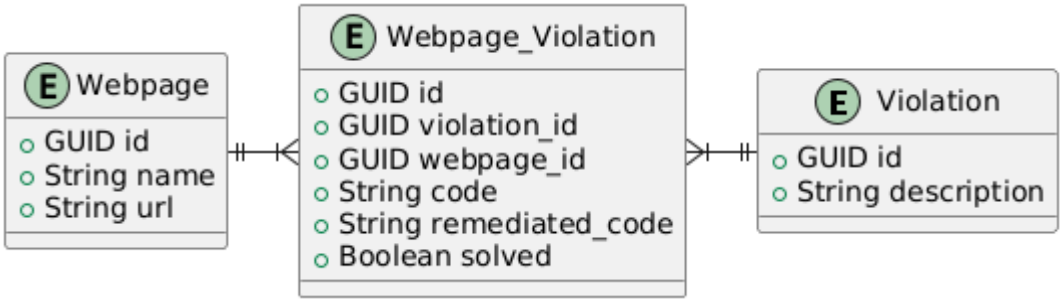


Figura 10 - Modelo de domínio

Devido à complexidade e modo de uso da ferramenta AXE, o Analisador é desenvolvido segundo uma arquitetura mais simplista quando comparado com a API. Por este motivo, a arquitetura adotada é uma variação da arquitetura *Layered* [50], composto por duas componentes principais: *Handler* e *Helper*. Na primeira, *Handler*, a análise da página e o levantamento dos problemas de acessibilidade presentes são geridos, sendo obtidos os dados necessários para a efetuar a remediação do problema. Quanto ao Helper, o seu único propósito é efetuar a comunicação com a API.

4.2.3 Vista de processos

A vertente processual da arquitetura tem em consideração os aspetos não-funcionais dos requisitos e lida com os elementos dinâmicos do sistema, explicitando os seus processos, comunicações, e outros aspetos relativos ao tempo de execução.

Na Figura 11 e 12, encontra-se apresentada a *user story* “Remediar Conteúdo”, genérica, sob diferentes níveis de granularidade, retratando as interações da aplicação no momento da remediação de problemas identificados de acessibilidade. Em complemento, os diagramas de sequência referentes aos casos de uso específicos identificados encontram-se disponibilizados sob a forma de anexos e posteriormente abordados.

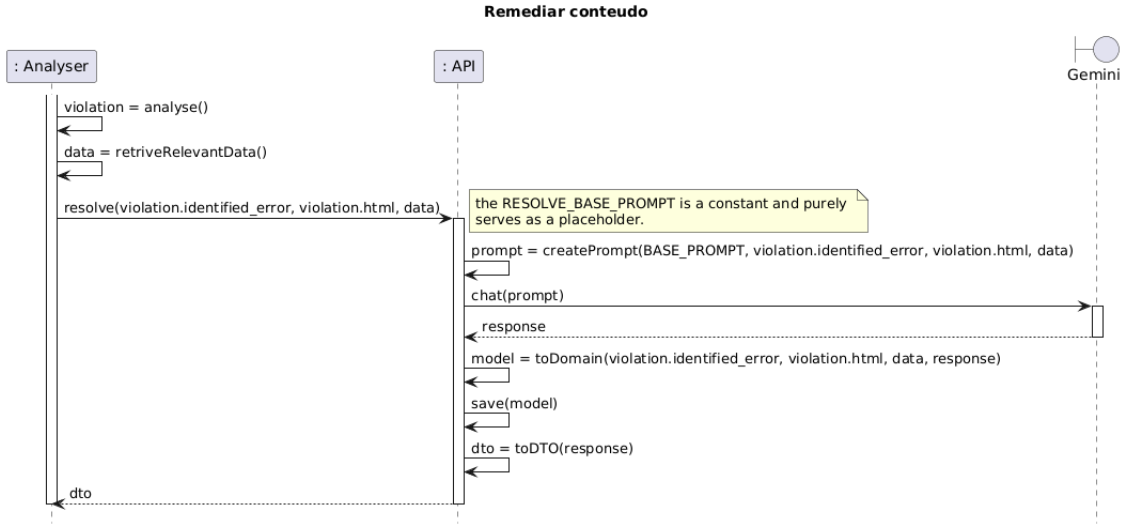


Figura 11 - Nível 1: Vista de Processos - Remediar Conteúdo

Através da passagem do erro identificado pelas ferramentas de análise, bem como o trecho de código onde este foi detetado, o sistema procede ao tratamento destas informações e gera o *prompt* que será posteriormente passado ao Google Gemini.

Na Figura 12 está presente um nível mais detalhado da vista de processos, que descreve o fluxo de comunicações entre as partes envolventes no mesmo. É ainda feita referência ao *prompt* utilizado para efetuar a *query* ao Gemini para obter a remediação do problema identificado. O sistema incorpora neste, o erro e o código HTML onde foi identificado, antes de ser utilizado na comunicação com o Gemini. Após o seu processamento, o conteúdo remediado é retornado.

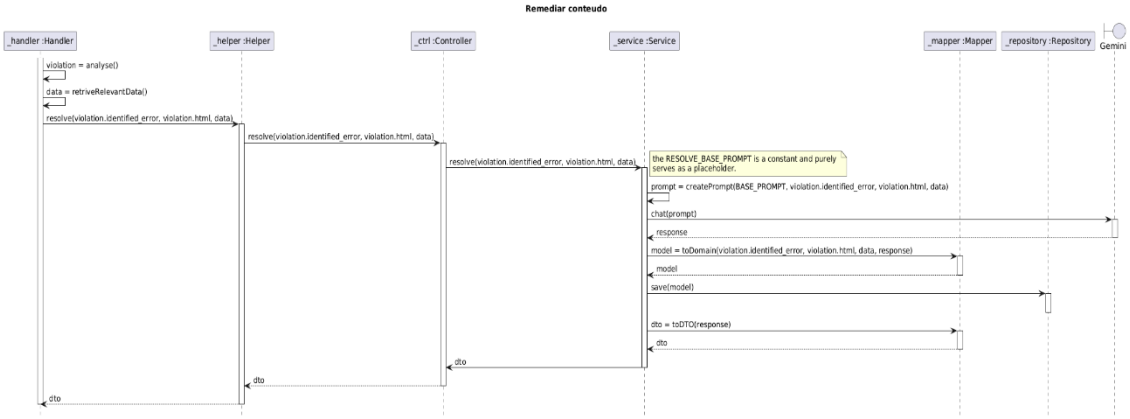


Figura 12 - Nível 2: Vista de Processos - Remediar Conteúdo

4.2.4 Vista de implementação/desenvolvimento

A Figura 13 apresenta a vista de implementação do sistema segundo uma variação da notação de *Booch* com o objetivo de limitar os itens significativamente relevantes.

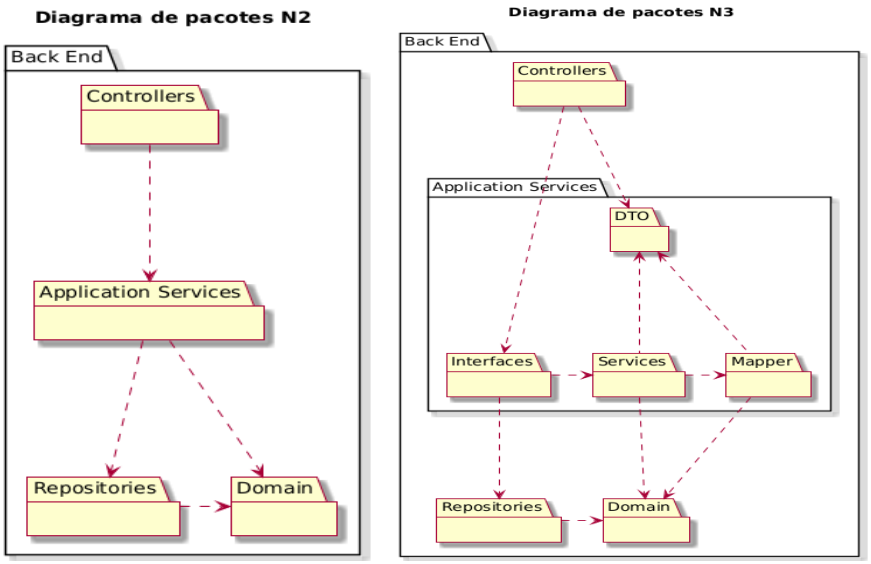


Figura 13 - Vista de Implementação níveis 2 e 3

O pacote *"Controllers"* remete à camada mais superficial da aplicação e onde se encontram os *endpoints* para os quais o analisador irá enviar os pedidos HTTP/HTTPS. No pacote "*Application Services*", diretamente abaixo, encontra-se a camada responsável pela funcionalidade e serviços, incluindo interfaces e serviços responsáveis pela coordenação dos pedidos, o desencadeamento de processos, e a gestão geral do fluxo da aplicação. Por fim, nas camadas mais baixas encontram-se os componentes relativos ao domínio e aos repositórios.

4.2.5 Vista de física/ implantação

A arquitetura física toma em consideração requisitos não funcionais do sistema, como a disponibilidade, confiabilidade, performance, e escalabilidade, e identifica os diferentes elementos, redes, e objetos, e respetivas configurações físicas.

Na solução desenvolvida, todos os componentes do sistema encontram-se hospedados na máquina local à exceção do Google Gemini, que está alojado nos servidores da Google. A Figura 14 ilustra os diferentes componentes físicos da solução, estando a comunicação entre as diferentes partes assegurada pelo protocolo HTTP/HTTPS.

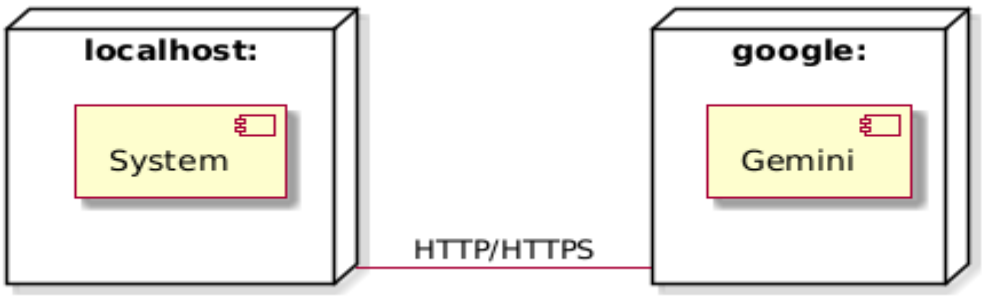


Figura 14 - Nível 1: Vista Física

4.2.6 Cenários

A remediação de problemas identificados de acessibilidade por parte de LLMs é o principal foco de estudo do projeto. Tendo em conta as violações mais frequentemente encontradas em conteúdo *web*, na Figura 15 estão presentes os processos identificados, remetendo cada uma a uma violação diferente.

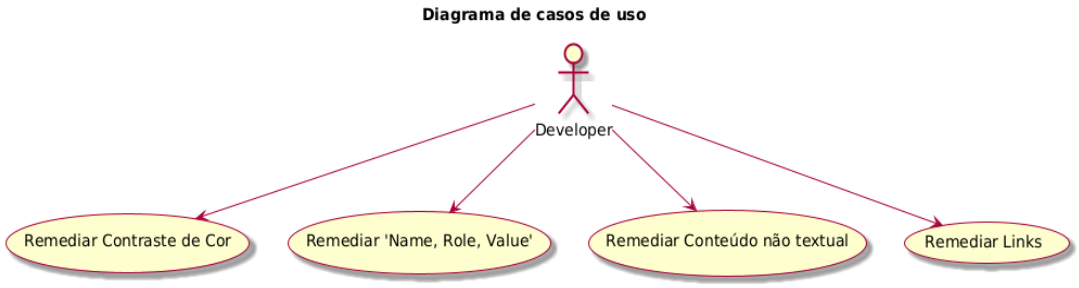


Figura 15 – Funcionalidades desenvolvidas

4.2.6.1 Remediar Contraste de Cor

Dados presentes em estudos [39, 32, 4] apontam o contraste insuficiente de cores como o problema de acessibilidade mais comum e que afeta a legibilidade do conteúdo *web,* estando presente em cerca de 80% das páginas analisadas.

Neste caso de uso, objetiva-se a correção desta componente de forma a que o rácio de contraste seja de, no mínimo:

- 4.5:1 para texto normal e 3:1 para textos grandes como títulos e cabeçalhos, de forma a obter conformidade com o *sucess criterion* (SC) 1.4.3 – *Contrast (Minimum)* (Nível AA) do WCAG 2.2.

- 7:1 para texto normal e 4.5:1 para textos grandes como títulos e cabeçalhos, de forma a obter conformidade com o SC 1.4.6 - *Contrast (Enhanced)* (Nível AAA) do WCAG 2.2.

4.2.6.2 Remediar "*Name, Role, Value*"

O presente critério reforça a necessidade de os componentes da interface gráfica terem nomes, papeis, e valores, de forma a possibilitar a interação com tecnologias assistivas. Uma vasta percentagem das páginas analisadas (68% [4]) apresenta inconformidades nesta vertente que se reflete numa dificuldade frequente na interação com estas páginas.

Neste caso de uso objetiva-se a correção de controlos HTML *standard*, garantindo uma compreensão e interação mais fácil para utilizadores de tecnologias assistivas.

4.2.6.3 Remediar Conteúdo não textual

Conteúdo não textual, como imagens e vídeos, devem apresentar um formato alternativo para que possa ser interpretado por utilizadores de tecnologias assistivas. Com base nos estudos analisados [39, 4], cerca de 30% das páginas *web* não apresenta conformidade neste critério e onde a grande maioria das situações identificadas remete para a falta de textos alternativos em imagens.

Com base nas capacidades dos LLMs, este caso de uso visa a remediação destes problemas através da análise do conteúdo não textual e da geração de texto alternativo de acordo com os padrões utilizados por especialistas, designadamente a nível de:

- **Precisão** – o texto alternativo deve ser representativo da imagem em questão, tanto em conteúdo como em função.
- **Sucintez** – o conteúdo e função, se existirem, devem ser apresentados de forma sucinta, sem sacrifício da precisão.
- **Redundância** – o texto alternativo não deve ser redundante ou fornecer a mesma informação repetida.
- **Uso de frases como "Imagem de" ou "Gráfico de"** – A utilização destas torna-se redundante dado que leitores de ecrã já fazem esse anúncio.

4.2.6.4 Remediar *Links*

Ligações, ou *links,* vazios ou sem texto percetível, presente em 52% das páginas *web* analisadas [4], surgem maioritariamente devido ao uso de imagens como conteúdo único dos mesmos e sem nenhum nome acessível disponibilizado. Cerca de 13,2% dos *links* encontrados continham texto ambíguo, como é exemplo de "Clique aqui", "Mais", ou "Continuar", totalizando uma média de 6 casos por página. *Links* que direcionam o utilizador para diferentes partes da página, denominados de *skip links* [38], foram ainda detetados em 13,3% das páginas analisadas, nos quais um em cada dez encontrava-se não funcional, ora escondidos de forma a estarem inacessíveis ou o seu conteúdo alvo não se encontrava presente [39].

Neste caso de uso, estes pontos são adereçados de forma a serem corrigidos e identificadas quaisquer limitações ou dificuldades encontradas.

5 Implementação da Solução

Este capítulo tem como foco a descrição dos detalhes e resultados da implementação da solução analisada e desenhada previamente. Inicialmente é introduzido o conceito de *Prompt Engineering,* utilizado durante a comunicação com o LLM, seguido da descrição da implementação.

5.1 Prompt engineering

Efetuar o *fine-tuning* do modelo [6, 11] consiste no retreino do LLM sob um *dataset* especializado, e é recomendado para obter os melhores resultados e adaptar o modelo a responder apropriadamente no contexto específico. Apesar de benéfico, este processo é muitas vezes inexequível pois requer custos acrescidos, tempo, e a grandes de quantidades de dados para efetuar o treino. Modelos que passam pelo *fine-tune* são ainda bastante suscetíveis a problemas, podendo ter dificuldades em adaptar-se à nova informação, *overfit,* e até perder conhecimento obtido do seu treino original. Contudo, na inviabilidade do *fine-tuning,* devido à falta de dados de qualidade ou pela pura desnecessidade, uma alternativa se levanta, o *prompt engineering*.

Este conceito foi adotado durante o desenvolvimento da solução, e resume-se na formulação de *queries*/*inputs* passados a um modelo de IA, através da tradução dos requisitos vagos e complexos do utilizador, em instruções claras, diretas e concisas, que contribui para uma compreensão do problema e produção de respostas apropriadas [22].

Modelos de IA dependem fortemente dos dados fornecidos durante o seu treino, e os resultados gerados têm por base padrões aprendidos durante esta fase de aprendizagem. A engenharia de *prompts* torna-se então importante pois permite produzir *inputs* que estejam enquadrados com estes mesmos padrões, com o intuito de extrair respostas mais relevantes e precisas.

Para o desenvolvimento dos *prompts* apresentados neste documento, diferentes técnicas foram consideradas, designadamente *zero-shot, one-shot, few-shot* e *chain-of-thought* (CoT). Apesar de não haver nenhum método considerado como mais eficaz, as abordagens *zero* e *few-shot* foram adotadas dado que tiram proveito dos conhecimentos pré-existentes e compreensão do modelo, ao passo que promovem simplicidade e facilidade de implementação. Por outro lado, métodos com base em CoT, como o *Chain-of-Thought prompting, Tree-of-Thought e Self-Consistency,* frequentemente utilizados em tarefas mais complexas e que requerem maior raciocínio, foram desconsiderados devido à introdução de uma complexidade desnecessária ao quebrar em partes menores uma tarefa simples.

Assim, sendo todos os *prompts* desenvolvidos e apresentados neste documento seguem uma abordagem *zero-shot* ou *few-shot*, para inferir respostas apropriadas através da atribuição de

uma tarefa a um modelo IA com o fornecimento opcional de exemplos. Ainda, de acordo com *Ekin* [29], existem pelo menos 5 fatores relevantes a ter em consideração na construção de um *prompt* para obter respostas de alta qualidade: Intenção do utilizador, Compreensão do modelo, Especificidade do domínio, Clareza e especificidade, e Limitações.

A *framework* 3 Cs [9] é baseada nestes princípios, estando estruturada em três áreas principais:

- *Clarity* (Clareza) – fornecimento de detalhes específicos e intenções claras para guiar o modelo;
- *Context* (Contexto) – ajuda o modelo a entender o nível de complexidade da tarefa;
- *Contraints* (Limitações) – limitação do formato, tamanho, estilo, entre outros aspetos, do *output*.

Na Figura 16 encontra-se melhor descrita esta metodologia assim como um exemplo para a construção de um *prompt* segundo uma abordagem *few-shot.*

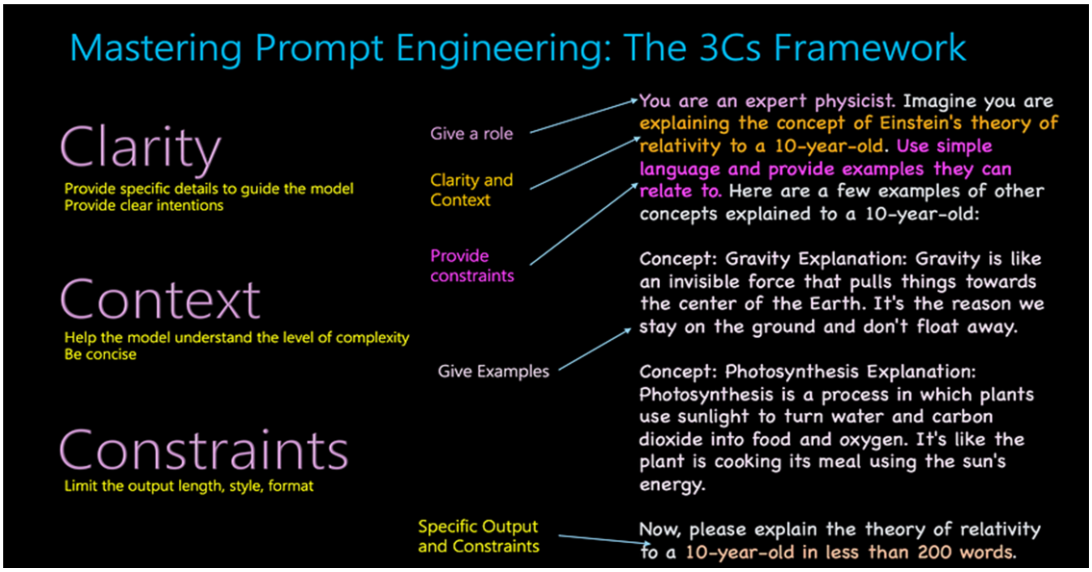


Figura 16 - Framework 3Cs e processo de construção (Fonte: [9])

5.2 Procedimento

O presente subcapítulo tem como propósito introduzir conceitos e procedimentos gerais que serão posteriormente detalhados na descrição da implementação. Pretende-se descrever aqui o fluxo genérico de cada caso de uso, e identificar processos que são partilhados entre os mesmos.

O uso do *webdriver Selenium* na solução, usado para manipular os diferentes elementos da página HTML, implica a sua configuração inicial. No Extrato de Código 1 - Configuração inicial do webdriver Selenium*,* encontra-se exemplificado este processo de configuração do *driver.*

```
var driver = new Builder().forBrowser(Browser.CHROME).build();
```

Extrato de Código 1 - Configuração inicial do webdriver Selenium

Uma vez configurado, este *driver* é usado em conjunto com o *AxeBuilder* para configurar a análise segundo regras e *tags*. No modo de funcionamento do AXE, os problemas de acessibilidade encontram-se categorizados segundo diferentes regras, sendo que cada regra possui uma *tag* que indica qual a versão e nível do WCAG a que esta pertence. Na configuração apresentada no Extrato de Código 2 - Configuração base do *AxeBuilder*, as *tags* definidas ajudam a orientar a pesquisa violações de acessibilidade para o nível A e AA do padrão de acessibilidade WCAG 2.2.

```
AxeBuilder(driver).withTags(['wcag22aa'])
```

Extrato de Código 2 - Configuração base do *AxeBuilder*

5.3 Descrição da Implementação

As secções posteriores descrevem com detalhe o processo de desenvolvimento associado a cada caso de uso identificado.

5.3.1 Remediar Contraste de Cor

No presente caso de uso pretende-se remediar o código inconforme de forma a cumprir os seguintes critérios de sucesso:

- 1.4.3 - *Contrast (Minimum) (Level AA)* da versão 2.2 do WCAG;
- 1.4.6 - *Contrast (Enhanced) (Level AAA)* da versão 2.2 do WCAG.

Para identificar apenas problemas de acessibilidade relacionados com o contraste de cor, a análise foi limitada através da definição da regra "*color-contrast*", que assegura a conformidade de contraste entre a cor de fundo e a da letra segundo os critérios do WCAG 2.2.

Após a análise da página, os erros são identificados e abordados de forma individual através de um processo cíclico, onde são adquiridas todas as informações necessárias para o tratamento do contraste insuficiente de cores, como:

- O elemento e a cor da fonte onde o erro foi identificado;
- O elemento e a cor de fundo onde a cor da fonte contrasta.

Neste segundo ponto, a obtenção destes elementos revela-se um processo mais complexo, uma vez que a cor da fonte pode contrastar tanto com a cor de fundo do próprio elemento, como de um dos seus ascendentes na árvore de nós HTML.

```
let parentElement = await driver.executeScript(async function(ele) {
                  let bgColor = window.getComputedStyle(ele).backgroundColor;
                      // If background is fully transparent, check the parent
                  while (ele.parentElement && (bgColor === 'rgba(0, 0, 0, 0)'
                        || bgColor === 'transparent')) {
                    ele = ele.parentElement;
                    bgColor = window.getComputedStyle(ele).backgroundColor;
                  }

                  return {
                    "element" : ele,
                    "backgroundColor" : bgColor};
                }, element);
```

Extrato de Código 3 - Processo de procura do elemento responsável pela cor de fundo

O Extrato de Código 3 ilustra o processo de procura do elemento responsável pela cor de fundo. Inicialmente, é passado como parâmetro o elemento no qual o erro foi detetado e avaliada a presença de cor de fundo. Na presença, o mesmo é retornado sob o formato JSON para posterior integração no *payload* do pedido remetido à API. Na ausência da cor de fundo, o processo é recursivamente repetido para o elemento ascendente (nó pai) até ser identificado o elemento que tenha esta propriedade definida.

Uma vez na API, um *prompt* pré-definido é utilizado, que incorpora todas as informações recebidas no pedido para a comunicação com o LLM. O conteúdo deste *prompt* pode ser consultado no Extrato de Código 4.

```
Your objective is to solve accessibility issues from HTML code.
To do that, I need you to be very precise, specific and, most importantly, as short as
possible.
This said, the content I want you to analyse is the following element:
 %2$s
In it the content provided, the following accessibility error related to color contrast was
found: %1$s
The font color applied to the element is the following:
 %3$s
The element css selector is: %4$s
The parent element is the following:
 %5$s
The background color of this element is %6$s
Provide me with a proper solution either using inline styling or css rules and without
explanations.
```

Extrato de Código 4 - *Prompt* padrão para a remediação de contraste de cor.

O diagrama de sequência referente ao presente processo é apresentado em anexo [Anexo 5 – Remediar Contraste de Cor (Nível 2)] e ilustra com maior detalhe o fluxo do mesmo.

5.3.2 Remediar "Name, Role, Value"

No presente caso de uso pretende-se remediar o código inconforme de forma a cumprir os seguintes critérios de sucesso:

- 4.1.2 - *Name, Role, Value (Level A)* da versão 2.2 do WCAG.

A remediação de problemas relacionados com "*Name*", "Role", e "Value" implica a modificação do elemento em questão, de forma a incluir os atributos ARIA necessários para garantir a conformidade com o padrão WCAG 2.2. Para isso, o *AxeBuilder* foi parametrizado com as seguintes regras:

- *aria-allowed-attr* – que garante que papel (*role*) do elemento suporta os atributos ARIA definidos.
- *aria-hidden-body* – que garante que inexistência de elementos no documento com o atributo aria-hidden="true".
- *aria-input-field-name* – que garante que todos os campos de *input* têm um nome acessível.
- *aria-required-attr* – que assegura que todos os elementos com papéis ARIA possuem os atributos ARIA obrigatórios.
- *aria-roles* – que garante que todos os elementos com papéis ARIA usam um valor válido.
- *aria-toggle-field-name* – que garante que todos os elementos ARIA com comportamentos *toggle* têm um nome acessível.
- *aria-valid-attr-value* – que garante que todos os atributos ARIA tem um valor válido.
- *label* – que assegura que todos os elementos de formulário têm um rótulo.

Os erros identificados são enviados diretamente à API após a análise, sem necessidade de proceder à obtenção de dados extra, uma vez que, para este caso de uso, as únicas informações necessárias para a remediação destas violações são o erro e o código onde este se encontra.

Os dados enviados pelo analisador são então tratados pela API para compor o *prompt* encarregue de transmitir a tarefa ao LLM. No Extrato de Código 5, encontra-se presente este mesmo *prompt*, que serve como *placeholder* para a API manipular e fazer a integração das informações recebidas.

```
Your objective is to solve accessibility issues from HTML code.
To do that, I need you to be very precise, specific and, most importantly, as short as
possible.
This said, the content I want you to analyse is the following: %2$s

In it the content provided, the following accessibility error was found:
 %1$s
Provide me with a better alternative, in a way that could be simply copied and pasted
without any explanation, maintaining the content and resolving the error according with the
Web Content Accessibility Guidelines (WCAG)
Be sure that all the elements in the generated HTML code are properly closed and in plain
text format.
```

Extrato de Código 5 - Prompt padrão para a remediação de violações relacionadas com *Name, Role e Value*

O diagrama de sequência referente ao presente processo é apresentado em anexo [Anexo 6 – Remediar *Name, Role* e *Value* (Nível 2)] e contém maior detalhe sobre o fluxo do mesmo.

5.3.3 Remediar conteúdo não textual

No presente caso de uso pretende-se remediar o código inconforme de forma a cumprir os seguintes critérios de sucesso:

- 1.1.1 - *Non-text Content (Level A)* da versão 2.2 do WCAG.

A remediação de conteúdo não textual, principalmente relacionada com a falta de textos alternativos em imagens e gráficos, requer que o LLM analise e interprete estes elementos para ser capaz de resolver estes problemas. Para detetar violações neste âmbito, o AxeBuilder foi configurado com as seguintes regras:

- *area-alt* – que garante que elementos '*area*' de mapas de imagem têm texto alternativo;
- *image-alt* – que garante que elementos '*image*' têm texto alternativo ou *role=”none”;*
- *input-image-alt* – que garante que campos de *input* do tipo '*image*' têm texto alternativo.

Quanto ao levantamento do conteúdo não textual para transmissão posterior ao LLM, este é enviado também no *payload* do pedido à API, onde é feita a sua codificação e transformação em *array* de *bytes*, formato este exigido para a comunicação de imagens ao LLM. De notar que, na resposta proveniente do modelo de IA, não só é possível obter uma *tag* HTML conforme com os padrões WCAG, mas também um texto alternativo gerado que toma em consideração o nível de detalhe e estrutura indicada como correta por especialistas. Para sustentar estes resultados, o *prompt* presente no Extrato de Código 6 foi elaborado.

```
Your objective is to solve accessibility issues from HTML code, more specifically non-
textual content.
To do that, I need you to be very precise, specific and, most importantly, as short as
possible.
In the following code, there were some issues regarding its accessibility: %1$s;
The error found was the following: %2$s;
Your objective is to fix the code provided by generating alternative text for the image,
providing only the fixed code.
For this alternative text, start with a short summary, about 1-2 sentences, keeping focus on
the key elements without additional commentary, then add more 1 to 4 more sentences
conveying any missing details or relationships that might be relevant to understand the
figure, avoiding adding repeating content already provided.
```

Extrato de Código 6 - Prompt padrão para a remediação de violações relacionadas com conteúdo não textual

O diagrama de sequência referente ao presente processo é apresentado em anexo [Anexo 7 – Remediar conteúdo não textual (Nível 2)] e contém maior detalhe sobre o fluxo do mesmo.

5.3.4 Remediar *links*

Na remediação de *links,* a passagem da *tag* HTML, juntamente com o erro detetado, é insuficiente, pelo que a interpretação do propósito e conteúdo de cada *link*, do ponto de vista do LLM, é uma tarefa complexa.

Em certos casos, a falta de texto discernível requer um contexto mais detalhado da página em questão, pelo que o LLM não o consegue inferir através do URL isoladamente. Possíveis soluções passam pelo resumo de ambas as páginas, atual e destino, deixando o modelo fazer a correlação das mesmas, porém esta abordagem implicaria um maior esforço computacional e riscos de segurança, dificultando a remediação automática.

Com a implementação de *links* vazios, desenvolvedores muitas vezes manipulam o comportamento normal da página através de *scripts.* Se utilizados indevidamente, estes podem causar problemas de acessibilidade pois podem ser mal interpretados por leitores e outros dispositivos assistivos, provocando confusão aos seus utilizadores, nomeadamente se ocorrer um recarregamento da página ou se o foco for manipulado de forma errada. Neste aspeto, os

LLMs dificilmente serão capazes de interpretar o comportamento que a interação com o *link* irá despoletar.

No presente caso de uso pretende-se remediar o código inconforme de forma a cumprir os seguintes critérios de sucesso:

- 2.4.4 - *Link Purpose (In Context) (Level A)* da versão 2.2 do WCAG;
- 2.4.9 - *Link Purpose (Link Only) (Level AAA)* da versão 2.2 do WCAG.

Para detetar violações neste âmbito, o *AxeBuilder* foi configurado com as seguintes regras:

- *link-name*– que garante que todos os *links* têm um texto discernível*;*

Levantados os erros, a aplicação segue o seu fluxo normal, sendo enviadas as violações identificadas, assim como o código HTML inconforme, à API, onde são integrados no *prompt* apresentado no Extrato de Código 7.

```
Your objective is to solve accessibility issues from HTML code, most specifically links.
To do that, I need you to be very precise, specific and, most importantly, as short as
possible. Here's some examples:

Example 1 - Basic Empty link text:
    Input: <a href="https://www.google.com/"></a>
    Output: <a href="https://www.google.com/" aria-label="Go to Google's
homepage">Google</a>

Example 2 - Link with an icon only:
    Input: <a href="/settings"><i class="fa fa-cog"></i></a>
    Output: <a href="/settings" aria-label="Settings"><i class="fa fa-cog" aria-
hidden="true"></i></a>

Example 3 - Link with only an image
    Input: <a href="/profile"><img src="profile.png"></a>
    Output: <a href="/profile"><img src="profile.png" alt="Go to your profile"></a>

Example 4 - JavaScript-only placeholder
    Input: <a href="#" onclick="openHelp();"></a>
    Output: <button type="button" onclick="openHelp()">Open Help</button>

This said, the content I want you to analyse is the following: %2$s
In it the content provided, the following accessibility error was found:
 %1$s

Provide me with a better alternative, in a way that could be simply copied and pasted
without any explanation, maintaining the content and resolving the error according with the
Web Content Accessibility Guidelines (WCAG)
Be sure that all the elements in the generated HTML code are properly closed and in plain
text format.
```

Extrato de Código 7 - *Prompt* para a remediação violações de acessibilidade relacionadas com links

Dados os diferentes casos, este *prompt* segue uma abordagem *few*-shot, onde são fornecidos exemplos de situações genéricas e de como é expectável o LLM agir perante elas. Após o processamento do LLM, os dados obtidos durante o processo de remediação são salvaguardados para posterior análise.

O diagrama de sequência referente ao presente processo é apresentado em anexo [Anexo 8 – Remediar *links* (Nível 2)] e contém maior detalhe sobre o fluxo do mesmo.

6 Resultados

6.1 Protocolo

No processo de extração de resultados, os 100 *websites* mais populares extraídos do *ranking* Tranco [35] foram analisados, *ranking* este escolhido pela sua robustez metodológica, resultado da agregação de várias listas de popularidade provenientes de diversas fontes, que contribui para a mitigação de enviesamentos e manipulações artificiais.

Da amostra inicial, foram excluídas páginas que retornaram erros (por exemplo, 404), que dispunham menos de 10 elementos HTML ou que incluíam mais de 200 links para o mesmo domínio – um critério estabelecido com base em práticas de otimização para motores de busca (*Search Engine Optimization*, SEO).

No total, foram levantadas e analisadas cerca de 600 violações de acessibilidade, divididas entre as quatro áreas de acessibilidade identificadas e distribuídas segundo os valores percentuais apresentados na Figura 17.

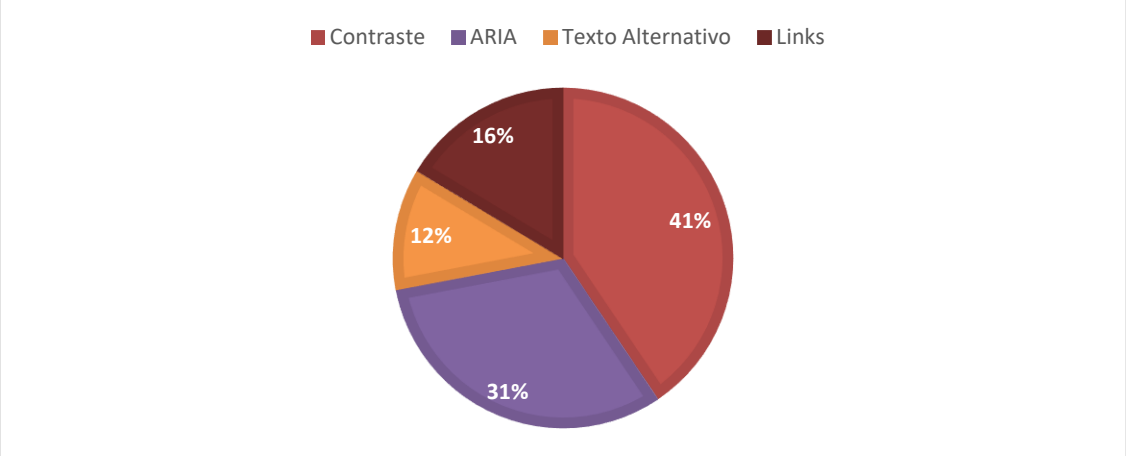


Figura 17 - Distribuição das violações de acessibilidade

6.2 Contraste de Cor

O propósito deste ponto é avaliar a eficácia do Gemini na remediação de problemas relacionados com a insuficiência de contraste em páginas *web.* Neste âmbito, os contrastes de cor gerados pelo modelo devem respeitar os rácios definidos pelas diretivas WCAG 2.2, isto é, um rácio mínimo de 4.5:1 para textos de grandes dimensões (como títulos e cabeçalhos) e de 7:1 para textos normais, com o objetivo de obter, respetivamente, conformidade com os

critérios de sucesso SC 1.4.3 – *Contrast (Minimum)*, de nível AA, e SC 1.4.6 - *Contrast (Enhanced)*, de nível AAA.

Para a avaliação do contraste de cor, recorreu-se à API *Contrast Checker* da WebAIM. Após cada proposta de remediação gerada pelo LLM, o par de cores (texto e fundo) foi submetido à API, que devolveu o respetivo rácio de contraste e a sua conformidade com os diferentes níveis de acessibilidade definidos pelas WCAG (A, AA e AAA).

A análise dos resultados demonstrou que o Google Gemini atingiu uma taxa de sucesso aproximada de 65% na remediação de casos de contraste insuficiente. Nas instâncias onde a correção não foi bem-sucedida, observou-se que, embora as soluções propostas cumprissem os requisitos mínimos do nível AA, não satisfaziam os critérios os critérios mais exigentes do nível AAA.

Para colmatar as falhas, o *prompt* base foi reformulado com o objetivo de explicitar de forma mais clara o critério de sucesso. Especificamente, foi instruído ao modelo que fizesse uma validação prévia dos rácios de contraste das suas propostas, garantindo o cumprimento mínimo de 7:1, conforme requerido pelo SC 1.4.6.

Esta reformulação do *prompt* resultou numa melhoria significativa do desempenho do modelo, com um aumento aproximado de 25 pontos percentuais na taxa de sucesso. Assim, a taxa de remediação bem-sucedida passou a situar-se nos 89.79%.

6.3 “Name, Role, Value”

No âmbito da remediação dos atributos ARIA - *name, role* e *value* -, o analisador foi inicialmente parametrizado com objetivo de identificar violações segundo oito regras distintas. Destas, apenas quatro revelaram inconformidades: *aria-allowed-attr, aria-roles*, *aria-valid-attr-value e label.* A Figura 18 apresenta a distribuição das violações detetadas relativamente a estes atributos.

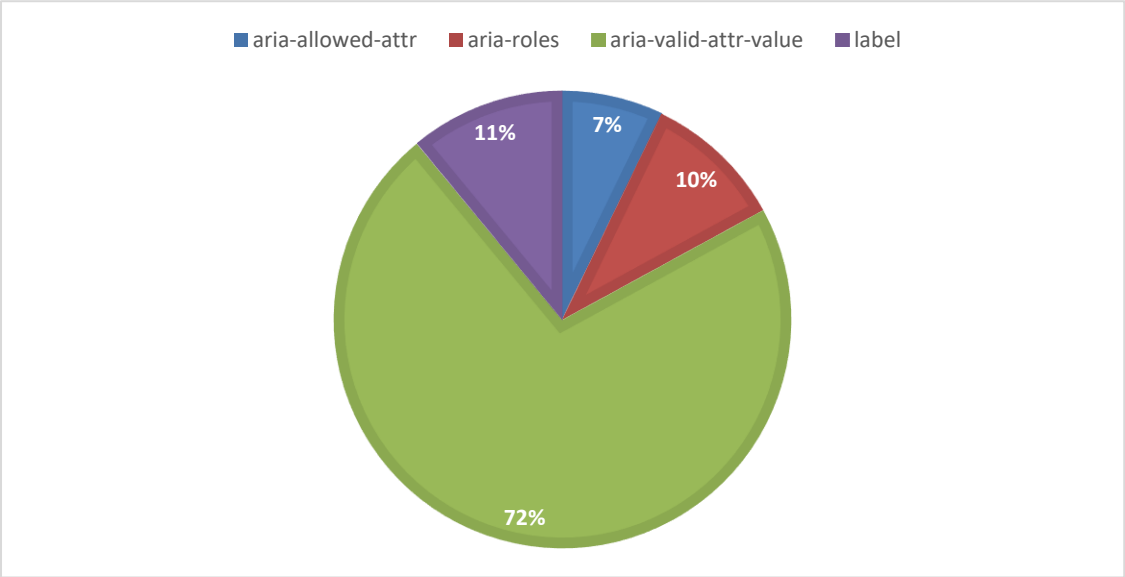


Figura 18 - Distribuição dos erros *name, role* e *value*

Os resultados apresentados indicam que a maioria das não conformidades relacionam-se com a atribuição de valores inválidos a atributos ARIA. No Extrato de Código 8 encontra-se um exemplo de inconformidade identificado, onde o atributo *aria-disabled* encontra-se sem qualquer valor, inválido para esta propriedade.

```
<button id="tabs-pill-bar-oc44a7_tab1-tab" data-cmp-hook-tabs="tab" class="pill-bar__item "
                href="#tabs-pill-bar-oc44a7_tab1" role="tab" data-mount="ocr-tabs" aria-
                label="Para utilização pessoal"
                aria-controls="tabs-pill-bar-oc44a7_tab1" aria-selected="false"
                aria-disabled="" tabindex="-1"
                data-bi-an="Tab_2_For personal" data-bi-ecn="For personal" data-bi-ehn="Get
        the latest on Microsoft Outlook"
                data-bi-ct="Pill Nav" data-bi-hn="Get the latest on Microsoft Outlook"
                data-bi-cn="For personal"
                data-bi-compnm="Carousel - Filtered"
                data-bi-view="News and tips | Get the latest on Microsoft Outlook"></button>
```

Extrato de Código 8 – Inconformidade *aria-valid-attr-value* (*aria-disabled*)

Neste caso em específico, o elemento contém ainda o atributo *tabindex* com o valor -1, o que sugere que o botão poderá estar inacessível ao foco, mas não necessariamente desabilitado. Desta forma, a remediação dependerá da intenção do desenvolvedor na medida em que, caso o botão tenha com propósito a interação com o utilizador, o atributo *aria-disabled* poderá ser declarado como *false* ou até mesmo removido. Caso contrário, a atribuição do valor *true* é apropriada.

A ausência de rótulos (*labels*) e a definição incorreta ou omissão de papéis semânticos (*roles*) em elementos de interface representam, respetivamente, 11% e 10% das inconformidades identificadas nesta categoria.

Casos como os apresentados nos Extrato de Código 9 e 10 comprometem a experiência de utilizadores de tecnologias assistivas de diversas formas. No primeiro exemplo, a utilização de um atributo *role* com valor vazio ou inválido impede o seu reconhecimento, prejudicando a navegação assistida. No segundo, a ausência de uma *label* explícita ou de atributos como *aria-label* ou *aria-labelledby* dificulta a identificação do propósito do campo de pesquisa por parte de leitores de ecrã.

```
<a href="https://www.microsoft.com/pt-pt/store/b/xbox?icid=MSCOM_QL_Xbox" class=" "
            data-automation-test-id="LinkListItemCTA1-link-list-uide334" aria-label="Botão
        de comprar consolas e jogos Xbox"
            target="_self" data-target=" " role=" " data-bi-mto="" data-bi-cn="Comprar
        consolas e jogos Xbox"
            data-bi-ecn="Comprar consolas e jogos Xbox" data-bi-bhvr="0" data-bi-ct="Link"
        data-bi-pa="Body"
            data-bi-compnm="Link List" data-target-uri="https://www.microsoft.com/pt-
        pt/store/b/xbox?icid=MSCOM_QL_Xbox"></a>
```

Extrato de Código 9 – Inconformidade *aria-role* (*role*)

```
<input type="search" class="lp-form-field__content" name="new" required=""
autocomplete="off" maxlength="60"
            autocapitalize="none">
```

Extrato de Código 10 – Inconformidade *label*

Por fim, a definição inválida de atributos representa cerca de 7% das violações observadas nesta categoria. Elementos como o apresentado no Extrato de Código 11, onde o atributo role é aplicado de forma incorreta a um botão, comprometem a interoperabilidade com navegadores e tecnologias assistivas, resultando em comportamentos instáveis e imprevisíveis. Além de violar a conformidade com os padrões do HTML5, estes erros introduzem inconsistências semânticas que dificultam a navegação e compreensão por parte de utilizadores com deficiência.

```
<button
        class="TUXButton TUXButton--default TUXButton--medium TUXButton--secondary css-
19eo4dx-StyledTUXNavButton ej7on1p0"
        aria-disabled="false" type="button" data-e2e="nav-more-menu" aria-label="Mais"
        role="listitem"
        aria-pressed="false"></button>
```

Extrato de Código 11 – Inconformidade *aria-allowed-attr*

Com base nos dados analisados, o Gemini 2.0 demonstrou ser capaz de resolver aproximadamente 92% das violações identificadas. Entre estas, destaca-se a correção de valores inválidos em atributos ARIA, que representou a maior parte das ocorrências.

A análise conduzida evidenciou, no entanto, algumas limitações do modelo na resolução de componentes personalizados, particularmente aqueles associados a *frameworks* ou bibliotecas especificas, como a *Adobe Experience Manager* (AEM), detetada em determinadas páginas. Esta dificuldade pode estar relacionada com o aumento da complexidade da estrutura DOM, resultado da utilização destes componentes, e que dificulta a análise e correção de problemas de acessibilidade.

O estudo conduzido por Achraf Othman [2] aborda este tema, concluindo que a complexidade do *design* do *website* impacta significativamente a capacidade do ChatGPT identificar e remediar problemas de acessibilidade de forma eficaz. Deste modo, as propostas de remediação fornecidas pelo LLM devem ser encaradas apenas como recomendações sujeitas a erros, sendo crucial que os desenvolvedores realizem uma verificação manual do código gerado de forma a garantir a sua precisão e conformidade.

Por fim, a Figura 19 sintetiza as regras analisadas, o número de violações identificadas por cada uma, e o respetivo número de correções efetuadas.

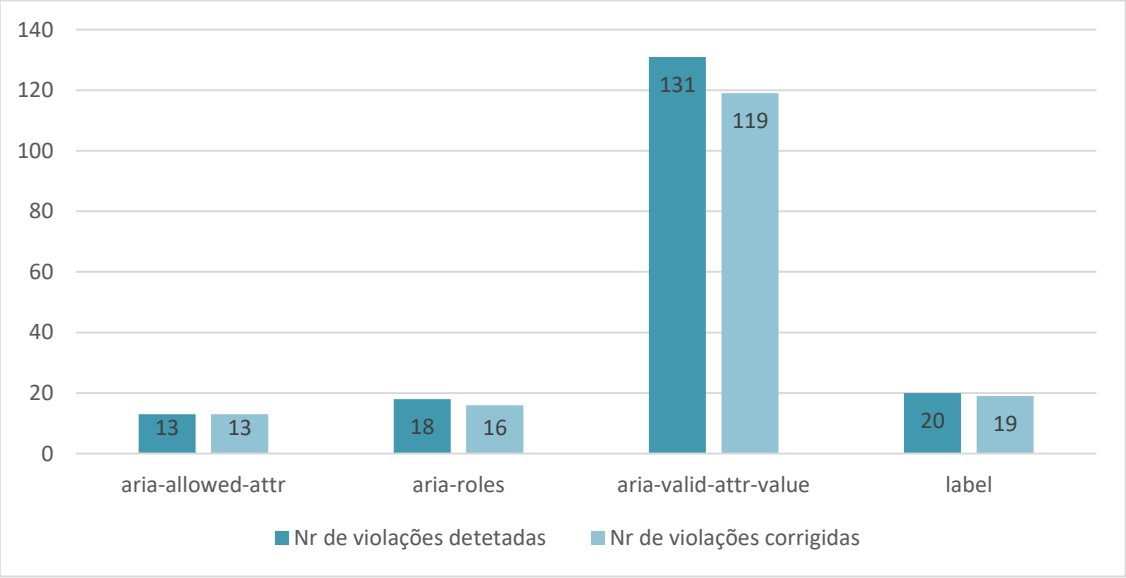


Figura 19 – Número de violações detetadas e corrigidas

6.4 Conteúdo não textual

Entre as categorias analisadas, a de conteúdo não textual registou o menor número de ocorrências, representando apenas 12% do total de violações identificadas.

A remediação deste tipo de conteúdo revela-se um desafio pois, para além de garantir que os elementos remediados passam na análise automática, é igualmente necessário assegurar a conformidade com as boas práticas recomendadas por especialistas e a adequação ao tipo de imagem e ao respetivo contexto [51].

A título de exemplo, a seguinte informação é apresentada para os diferentes tipos de conteúdo não textual:

- **Conteúdo cuja finalidade principal é criar uma experiência sensorial.**
  Por vezes, o conteúdo visa proporcionar uma experiência sensorial que não pode ser traduzida integralmente por meio de palavras, como acontece em performances musicais e obras de arte. Nestes casos, um rótulo textual descritivo mínimo deve ser fornecido e que identifique propriamente o conteúdo, auxiliado, sempre que possível, por uma descrição mais detalhada.

- **Conteúdo não textual que seja um controlo ou que aceite *inputs* de utilizador.**
  Como imagens utilizadas como botões de submissão, mapas de imagem ou animações complexas, um nome tem de ser disponibilizado de modo a que o propósito do conteúdo não textual possa ser compreendido pelo utilizador.

- **Conteúdo não textual que não tem a finalidade de ser visto, compreendido, ou puramente decorativo.**
  Atribuir texto alternativo a estes elementos pode distrair ou confundir utilizadores de ferramentas assistivas, enquanto omiti-los por completo pode gerar incerteza. Nestas instâncias, é necessário marcar estes conteúdos de forma a que sejam ignorados por tecnologias assistivas, garantindo uma experiência acessível e sem ruído informativo.

- **Exercícios não textuais que provem que o utilizador é humano (CAPTCHA).**
  Em determinados *websites,* os CAPTCHA são utilizados para prevenir o acesso automatizado por robôs de *spam* e outros tipos de *software,* recorrendo a tarefas visuais ou auditivas que excedem as capacidades atuais destes agentes. Nestas instâncias, o texto alternativo deve descrever o propósito do CAPTCHA, sendo este complementado por alternativas em diferentes formatos – como versões auditivas - que vão ao encontro das necessidades dos utilizadores com deficiência.

Os resultados obtidos apontam para um desempenho moderado do modelo, justificado por uma taxa de remediação na ordem dos 73%. A análise conduzida revela ainda que as principais

violações identificadas incidiram sobre dois domínios: imagens e elementos gráficos no formato *Scalable Vector Graphics* (SVG).

No contexto de imagens, o modelo conseguiu, de forma geral, resolver satisfatoriamente as inconformidades detetadas. Nas ocorrências de remediação malsucedida, estas deveram-se maioritariamente à inacessibilidade dos recursos referenciados no código, o que inviabilizou a sua correção. Entre os casos de sucesso, destacam-se as seguintes situações:

- **Imagem com atributo "*src*" vazio** – o modelo foi capaz de identificar estas imagens como conteúdo meramente decorativo, atribuindo-lhes o atributo *aria-hidden* de forma a ficarem invisíveis para tecnologias assistivas.
- **Imagens do tipo CAPTCHA** – inicialmente, a correção deste tipo de conteúdo não obedecia aos padrões definidos pelas WCAG. Após ajustar o *prompt* inicial de forma a incluir instruções especificas para este tipo de conteúdo, o modelo passou a gerar soluções conformes.
- **Conteúdos publicitários e imagens com texto em línguas diversas** – o Google Gemini demonstrou capacidade para identificar o carácter promocional das imagens em questão e traduzir os textos presentes nas mesmas, assegurando a acessibilidade da informação.

Em relação aos SVG, verificou-se que a grande maioria das violações identificadas persistiu após a sua correção. Este acontecimento deve-se à tendência do modelo para gerar soluções que utilizam as *tags* <*title*> e <*desc*> para melhoria da acessibilidade. Embora esta abordagem seja válida em vários contextos, o analisador considera a ausência do atributo *aria-label* como uma inconformidade.

Para mitigar este problema, o *prompt* utilizado foi ajustado no sentido de alinhar-se com as recomendações WCAG, para que, em componentes interativos, seja utilizado o atributo *aria-label*, e que, em componentes semânticos, informativos, ou no geral mais complexos, como gráficos, mapas ou infográficos, onde por norma é necessário um maior detalhe, sejam utilizadas as *tags title* e *desc.* Esta alteração resultou numa melhoria significativa, com um aumento de 13%, tendo sido alcançada uma taxa de remediação total de 86%.

Apesar da melhoria, destacam-se casos pontuais em que o modelo não foi capaz de identificar, e por consequência descrever, com eficácia as imagens. Nestes casos, os textos alternativos gerados apresentam um carácter especulativo, que não refletem com precisão o conteúdo visual das imagens.

6.5 Links

No que respeita à remediação de hiperligações sem texto discernível, os resultados obtidos indicam uma taxa de sucesso bastante elevada, próxima da integralidade, porém com alguns reveses.

Na maioria dos casos, o modelo foi capaz de inferir o conteúdo descritivo de cada ligação apenas com base no seu URL, alcançando uma taxa de remediação de 99%. O Extrato de Código 12, ilustra um exemplo de um *link* cujo URL é antidescritivo. Neste tipo de *links,* o modelo consegue identificar com facilidade o seu propósito e gerar um texto alternativo relevante e informativo para o elemento HTML correspondente.

```
<a class="flex justify-end h-auto" href="/features/video-analytics" target="_blank">
```

Extrato de Código 12 – Link auto-descritivo

Entre as limitações observadas, destaca-se a dificuldade na resolução de *links* cujo endereço se baseia em cadeias extensas de caracteres sem significado semântico reconhecível, exemplificado no Extrato de Código 13. Nestes casos, o modelo não consegue precisar o propósito da ligação, comprometendo a geração de um texto alternativo adequado.

```
<a class="mailru-morda-desktop--channel-info__link-1X"
href="https://dzen.ru/gadget_chef?country_code=ru&amp;lang=ru&amp;encoded_pulse_user
_info=G9CJIsxgHUk_MFjI1EDazxG9Est8uZALIs1hwrnrF1Tl1YdIO2UH8utNqHm9YD-
T9KVbXJb31__LT6BU3dxsAXIlKdQ-
XVAZQiCbVLMt0vw%3A1746653543562&amp;from_site=mail&amp;parent_rid=1413965380.1408.17
46653543390.93484&amp;from_parent_id=8352407695173479411&amp;from_parent_type=short_
video&amp;from_page=other_page" rel="dofollow" target="_blank" data-testid="header-
view-link" tabindex="-1">
```

Extrato de Código 13 – Link sem URL descritivo

Embora esta situação seja pouco frequente, a sua correção requer intervenção humana, uma vez que o modelo não pode aceder diretamente aos conteúdos dos URLs por questões de segurança. Deste modo, a colaboração com os programadores torna-se essencial nestes cenários para garantir a conformidade com os requisitos de acessibilidade.

Apesar desta limitação, o modelo conseguiu resolver o problema principal, tendo fornecido uma alternativa HTML conforme com as diretivas WCAG, embora com textos descritivos genéricos como “Anúncio” ou “Título do artigo”.

6.6 Resumo dos resultados

Ao longo deste trabalho, o Gemini 2.0 Flash foi avaliado em diferentes cenários de acessibilidade *web,* através do tratamento de aproximadamente 600 violações identificadas entre os 100 *websites* mais populares do mundo, e categorizadas segundo quatro áreas principais, alinhadas com os critérios estabelecidos nas WCAG 2.2.

No total, o modelo alcançou uma taxa média de remediação de 92%. Das observações realizadas, os melhores desempenhos foram identificados na correção de *links* sem texto discernível e na remediação de atributos ARIA (*name, role, value*). Por outro lado, verificam-se dificuldades na geração de texto alternativo para conteúdos não textuais, especialmente em imagens mais complexas ou contextuais.

Entre as limitações identificadas, o modelo revelou ainda dificuldade em avaliar corretamente contrastes de cor que estejam de acordo com os critérios WCAG 2.2. Embora o critério de sucesso 1.4.3 – *Contrast (Minimum)* (nível AA) tenha sido geralmente cumprido, o mesmo não se verifica para o critério 1.4.6 – *Contrast (Enhanced)* (nível AAA), que exige um rácio mínimo de contraste superior a 7:1.

Adicionalmente, observou-se uma correlação entre a complexidade estrutural das páginas e a eficácia do modelo. Em páginas com grande densidade de elementos HTML ou com componentes personalizados, taxas de sucesso inferiores foram registadas, o que sugere que o desempenho do modelo tende a decrescer em contextos mais intricados.

7 Conclusões

O presente capítulo tem como objetivo a documentação e detalhe de todas as descobertas e conclusões obtidas durante este trabalho. Inicialmente é feita a apresentação das ameaças à validade do estudo, a revisão das questões de investigação motivadoras do presente trabalho, e posterior resposta com base nos dados empíricos obtidos. Por fim, são ainda discutidas as contribuições do estudo, assim como o trabalho futuro.

7.1 Ameaças à validade

No presente estudo são reconhecidas várias ameaças internas e externas que possivelmente podem corromper a validade dos resultados e teses apresentadas.

7.1.1 Ameaças internas

Primeiramente, a utilização de ferramentas automáticas para a deteção de violações de acessibilidade representa uma potencial ameaça, devido à variabilidade nos resultados que estas podem apresentar. A título de exemplo, a ferramenta AXE utilizada neste estudo, reportou um número de ocorrências diferente em instâncias distintas da mesma página. Face a esta questão, a análise foi conduzida de forma recursiva, com várias análises efetuadas por página, de forma a minimizar o risco de omissão de violações. Importa referir ainda que a ferramenta AXE tem sido incorporada em outros estudos [26, 2, 34] devida à sua eficácia na deteção de ocorrências, contribuindo para a sua credibilidade.

Em segundo lugar, a competência do autor na elaboração dos *prompts* pode impactar a qualidade dos resultados, afetando diretamente a remediação gerada pelo modelo de linguagem. Esta ameaça foi atenuada através da utilização de *prompts* de natureza genérica, que permitem uma avaliação neutra da capacidade do modelo, isenta de vieses associados a técnicas avançadas de *prompt engineering*. Assim, cada critério foi analisado com base na capacidade do modelo de compreender a tarefa atribuída, do seu conhecimento na área e da sua proficiência na geração de código HTML acessível e semanticamente correto.

7.1.2 Ameaças externas

A análise feita incidiu em apenas 100 *websites,* o que representa apenas uma pequena amostra, pelo que os resultados obtidos não devem ser interpretados como normativos. Ainda, o presente trabalho focou-se única e exclusivamente na capacidade de remediação do Gemini 2.0 Flash, sendo que os resultados obtidos poderão não ser aplicáveis para os restantes LLMs.

7.2 Revisão das questões de investigação

Este estudo procurou avaliar a capacidade dos LLM, particularmente do Gemini 2.0 Flash, para remediar de forma automática diferentes tipos de violações de acessibilidade em *websites*. A investigação centrou-se em quatro grandes áreas, refletidas nas seguintes questões:

- RQ 1 -Podem os LLMs resolver eficazmente problemas de contraste de cor segundo as diretrizes WCAG 2.2?
- RQ 2 -Podem os LLMs resolver eficazmente problemas relacionados com atributos "*name*", "*role"* e "*value"* (ARIA)?
- RQ 3 -Podem os LLMs ser capazes de gerar corretamente descrições alternativas ("*alt texto"*) para imagens, de acordo com os padrões usados por especialistas?
- RQ 4 - Podem os LLMs resolver eficazmente problemas de links vazios ou sem texto discernível, garantindo conformidade com boas práticas de acessibilidade?

7.2.1 Contraste de cor

Baseado na análise conduzida, os resultados indicam que o modelo é eficaz na maioria dos casos que requerem conformidade com o critério de sucesso 1.4.3 - *Contrast (Minimum)* (nível AA), assegurando um rácio mínimo de contraste de 4.5:1. No entanto, no cumprimento do critério 1.4.6 - *Contrast (Enhanced)* (nível AAA), que estabelece um contraste mínimo de 7:1, o desempenho do modelo apresentou uma redução significativa, demonstrando limitações pontuais na identificação precisa de valores de contraste em certos cenários.

7.2.2 Atributos ARIA (*name, role, value*)

Neste domínio, o modelo apresentou uma taxa de remediação bastante satisfatória, na ordem dos 92%. Contudo, dificuldades persistem na remediação de componentes personalizados, desenvolvidos em *frameworks* específicas, o que limita a generalização da remediação em ambientes complexos ou fortemente customizados.

7.2.3 Texto alternativo para imagens

A remediação de conteúdo não textual revelou ser uma das áreas mais desafiantes. Embora o modelo tenha apresentado capacidade para resolver cenários simples, como imagens decorativas, CAPTCHAs, ou conteúdos publicitários, a taxa de remediação fixou-se nos 73%. Os principais entraves identificados verificaram-se na geração de texto alternativo para imagens contextuais ou SVGs, nos quais o modelo recorreu a descrições especulativas ou desprovidas de utilidade prática. No entanto, e para melhor alinhamento com as diretivas WCAG 2.2, a adaptação do *prompt* permitiu melhorias significativas, aumentando a taxa de remediação em 13%, porém não eliminou totalmente o problema.

7.2.4 Links vazios ou sem texto discernível

Com uma taxa de remediação de 99%, esta área revelou ser onde o modelo demonstrou maior eficácia. Em quase todas as instâncias, o modelo foi capaz de inferir corretamente o propósito da ligação com base no URL associado. No entanto, em situações pontuais, como *links* compostos por cadeias aleatórias de caracteres, a remediação necessitou de intervenção humana, uma vez que o modelo não tem contexto suficiente nem acesso direto ao conteúdo dos URLs, o que inviabiliza a remediação automática destas instâncias.

7.3 Contribuições do estudo

Este trabalho oferece contributos relevantes tanto para o avanço do conhecimento na área da acessibilidade digital assistida por modelos de linguagem de larga escala, como para a prática profissional ligada ao desenvolvimento *web* e à conformidade com os padrões de acessibilidade.

7.3.1 Contribuições académicas

- **Avaliação sistemática de LLMs em acessibilidade *web*:** A investigação conduzida representa uma abordagem sistemática à utilização de modelos de linguagem de grande escala, particularmente o Google Gemini 2.0 Flash, para a remediação semiautomática de violações das diretrizes WCAG, com base em dados reais de *websites.*
- **Análise segmentada por critérios WCAG:** Através da análise estruturada em quatro categorias (contraste de cor, atributos ARIA, conteúdo não textual e *links*), o presente estudo permite uma compreensão granular do desempenho dos LLMs em diferentes dimensões de acessibilidade *web*.

7.3.2 Contribuições práticas

- **Validação do uso de LLMs como ferramentas de apoio à remediação:** Os resultados empíricos obtidos validam o potencial da integração dos LLMs como agentes de correção preliminar ou assistida em ferramentas de análise de acessibilidade, ou até mesmo como agentes ativos no desenvolvimento de conteúdo *web* mais acessível desde as fases iniciais de construção.
- **Relevância para equipas de desenvolvimento e testes:** A análise realizada fornece informações relevantes sobre as áreas em que os LLMs podem ser mais confiáveis e onde a intervenção humana permanece essencial, servindo como argumento para a tomada de decisões estratégicas de planeamento de auditorias e correções em larga escala.
- **Escalabilidade:** Dada a natureza genérica dos *prompts* utilizados ao longo do trabalho, os métodos utilizados são passíveis de ser replicados ou integrados em ambientes de

desenvolvimento contínuo (CI/CD), acelerando a identificação e mitigação de problemas de acessibilidade.

7.4 Considerações finais

Os resultados obtidos neste estudo demonstram o potencial promissor dos LLMs como ferramentas de apoio na remediação de problemas de acessibilidade digital. O Gemini 2.0 Flash, alvo do estudo, apresentou desempenhos elevados em áreas como ARIA e *links,* e resultados positivos na correção de instâncias onde o contraste de cor era insuficiente e ausência de textos alternativos em imagens.

Apesar disso, os modelos de linguagem de larga escala atuais ainda não substituem a análise especializada, pelo que persistem limitações importantes, sobretudo quando a remediação exige interpretação semântica profunda, contexto visual complexo ou conhecimento externo à estrutura do HTML.

O presente estudo, limitado aos 100 *websites* mais populares do mundo e a quatro categorias de acessibilidade WCAG, representa apenas uma fração do universo da acessibilidade digital, dando oportunidade a trabalhos futuros à exploração de componentes como:

- a aplicação de LLMs em contextos multimodais (texto + imagem),
- a integração com ferramentas de design e desenvolvimento *web,*
- e a avaliação da eficácia da remediação com base na experiência de utilizadores reais com deficiência.

Em suma, apesar das suas limitações, os LLMs oferecem atualmente um ponto de partida sólido para a melhoria da acessibilidade *web* de forma automatizada, reduzindo o esforço necessário e promovendo práticas mais inclusivas na construção de interfaces digitais.

8 Referências

[1] AI Model & API Providers analysis, Artificial Analysis. https://artificialanalysis.ai/ (accessed Feb. 09, 2025).

[2] A. Othman, A. Dhouib, and A. N. Al Jabor, "Fostering websites accessibility: A case study on the use of the Large Language Models ChatGPT for automatic remediation," Proc. 16th Int. Conf. PErvasive Technol. Relat. Assist. Environ. (PETRA '23), New York, NY, USA, 2023, pp. 707–713, doi: 10.1145/3594806.3596542.

[3] A. Xu, M. Cai, D. Hou, R.-C. Chang, and A. Guo, "ImageExplorer Deployment: Understanding Text-Based and Touch-Based Image Exploration in the Wild," Proc. 21st Int. Web All Conf. (W4A '24), New York, NY, USA, 2024, pp. 59–69, doi: 10.1145/3677846.3677861.

[4] B. Martins and C. Duarte, “A large-scale web accessibility analysis considering technologyadoption,” Universal Access in the Information Society, Jul. 2023, doi: 10.1007/s10209 - 023-01010-0.

[5] C. Fontinha, “Diretiva Europeia da Acessibilidade Web - acessibilidade.gov.pt,” acessibilidade.gov.pt, Sep. 21, 2023. [Online]. Available: https://www.acessibilidade.gov.pt/blogue/categoria-noticias/diretiva-europeia-da-acessibilidade-web/

[6] C. Pornprasit and C. Tantithamthavorn, “Fine-tuning and prompt engineering for large languagemodels-based code review automation,” Information and Software Technology, vol. 175, p.107523, Jul. 2024, doi: 10.1016/j.infsof.2024.107523.

[7] Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H. P., Kaplan, J., … & Zoph, B. (2021). Evaluating Large Language Models Trained on Code. arXiv preprint arXiv:2107.03374.

[8] Chiang, W.-L., Zheng, L., Sheng, Y., Angelopoulos, A. N., Li, T., Li, D., Zhang, H., Zhu, B., Jordan,M., Gonzalez, J. E., & Stoica, I. (2024). Chatbot Arena: An Open Platform for Evaluating LLMs byHuman Preference. arXiv preprint arXiv:2403.04132.

[9] CryptoGPTo, “Mastering Generative AI Interactions: A Guide to In-Context Learning and Fine-Tuning,” https://rpradeepmenon.medium.com/mastering-generative-ai-interactions-a-guide-to-in-context-learning-and-fine-tuning-9ee620c76246.

[10] European Commission, Communication from the Commission to the Council, the European Parliament and the European Economic and Social Committee and the Committee of the Regions: eAccessibility, [SEC (2005)1095], Sep. 13, 2005. [Accessed: Nov. 19, 2021].

[11] F. Trad and A. Chehab, “Prompt Engineering or Fine-Tuning? A Case Study on Phishing Detectionwith Large Language Models,” Machine Learning and Knowledge Extraction, vol. 6, no. 1, pp.367–384, Feb. 2024, doi: 10.3390/make6010018.

[12] Hendrycks, D., Burns, C., Basart, S., Zou, H., Song, C., & Dietterich, T. (2021). Measuring MassiveMultitask Language Understanding. arXiv preprint arXiv:2104.08604

[13] Hudson, D. M., Tatar, D., Alayrac, J. B., Dziri, A., Cote, E., Li, M., … & Branson, S. (2021). MGSM: A Large-Scale Dataset for Multistep Reasoning in Grade School Math. arXiv preprintarXiv:2104.01663.

[14] J. M. López-Gil and J. Pereira, “Turning manual web accessibility success criteria into automatic: an LLM-based approach,” Universal Access in the Information Society, pp. 1–16, 2024.

[15] J. Whitehead *et al.*, “Conversational Interactions with Procedural Generators using Large Language Models,” in *Proceedings of the 20th International Conference on the Foundations of Digital Games*, ACM, Apr. 2025, pp. 1–7. doi: 10.1145/3723498.3723788.

[16] Kitchenham, B., Charters, S.: Guidelines for performing systematic literature reviews in software engineering. Engineering 2, 1051 (2007)

[17] Leitner, Jesse. (2017). GODDARD TECHNICAL HANDBOOK GSFC-HDBK-8005 Goddard Space Flight Center GODDARD SPACE FLIGHT CENTER GUIDELINE FOR PERFORMING RISK ASSESSMENTS. 10.13140/RG.2.2.10879.74406.

[18] Li, Yinheng. (2023). A Practical Survey on Zero-shot Prompt Design for In-context Learning.10.26615/978-954-452-092-2_069

[19] M. Das, A. J. Fiannaca, M. R. Morris, S. K. Kane, and C. L. Bennett, "From Provenance to Aberrations: Image Creator and Screen Reader User Perspectives on Alt Text for AI-Generated Images," Proc. 2024 CHI Conf. Human Factors Comput. Syst. (CHI '24), New York, NY, USA, 2024, Art. 900, pp. 1–21, doi: 10.1145/3613904.3642325.

[20] M. Elfleet and M. Chollet, “Investigating the impact of multimodal feedback on User-Perceived latency and immersion with LLM-Powered embodied conversational agents in virtual reality,” *IVA ’24: Proceedings of the 24th ACM International Conference on Intelligent Virtual Agents*, pp. 1–9, Sep. 2024, doi: 10.1145/3652988.3673965.

[21] M. J. Page et al., “The PRISMA 2020 statement : An updated guideline for reporting systematic reviews,” PLOS Medicine, vol. 18, no. 3, p. e1003583, 2021. [Online]. Available: https://doi.org/10.1371/journal.pmed.1003583

[22] Morales, Sergio & Clarisó, Robert & Cabot, Jordi. (2025). Impromptu: a framework for model-driven prompt engineering. Software and Systems Modeling. 1-19. 10.1007/s10270-024-01235-4.

[23] N. Singh, L. L. Wang, and J. Bragg, "FigurA11y: AI Assistance for Writing Scientific Alt Text," Proc. 29th Int. Conf. Intell. User Interfaces (IUI '24), New York, NY, USA, 2024, pp. 886–906, doi: 10.1145/3640543.3645212.

[24] Nayak, A., Nishida, T., Chen, H., & Bansal, M. (2022). General Purpose Question Answering withOpen Book Knowledge. arXiv preprint arXiv:2204.07154.

[25] P. Kruchten and Rational Software Corp., “Architectural Blueprints—The ‘4+1’ view model ofsoftware architecture,” journal-article, Nov. 1995. [Online]. Available:https://www.cs.ubc.ca/~gregor/teaching/papers/4+1view-architecture.pdf

[26] P. Mowar, Y.-H. Peng, J. Wu, A. Steinfeld, and J. P. Bigham, "CodeA11y: Making AI Coding Assistants Useful for Accessible Web Development," Proc. 2025 CHI Conf. Human Factors Comput. Syst. (CHI '25), New York, NY, USA, 2025, Art. 45, pp. 1–15, doi: 10.1145/3706598.3713335.

[27] Petersen, K., Feldt, R., Mujtaba, S., Mattsson, M.: Systematic mapping studies in software engineering. EASE 8, 68–77 (2008)

[28] R. C. Martin, “Design Principles and Design Patterns,” 2000. [Online]. Available: https://objectmentor.com/resources/articles/Principles_and_Patterns.pdf

[29] S. Ekin, “Prompt Engineering For ChatGPT: A Quick Guide To Techniques, Tips, And Best Practices.” May 2023. doi: 10.36227/techrxiv.22683919.

[30] Saxton, D., Thrush, G., Bruckner, K., Clark, J., & McCann, B. (2021). Mathematics Dataset: A NewDataset for Pretraining and Evaluating Language Models on Mathematical Problem Solving.arXiv preprint arXiv:2103.03874.

[31] Sy Desirée, “Adapting usability investigations for agile user-centered design,” *Journal of Usability Studies Archive*, May 2007, doi: 10.5555/2835547.2835549.

[32] Tait, A., Davis, S., Niyi-Awosusi, O., Wilhelm, G., Paige, K.: Accessibility. (2021). Accessed in 18 of January of 2022. https://almanac.httparchive.org/en/2021/accessibility

[33] Teng Wang, Jinrui Zhang, Junjie Fei, Yixiao Ge, Hao Zheng, Yunlong Tang, Zhe Li, Mingqi Gao, Shanshan Zhao, Ying Shan, et al. 2023. Caption anything: Interactive image description with diverse multimodal controls. arXiv preprint arXiv:2305.02677 (2023).

[34] U. Singh, J. D. Venkatesh, A. Muraleedharan, K. S. Saluja, A. J. H., and P. Biswas, "Accessibility Analysis of Educational Websites Using WCAG 2.0," Digit. Gov.: Res. Pract., vol. 5, no. 3, Art. 32, pp. 1–28, Sep. 2024, doi: 10.1145/3696318.

[35] V. le Pochat, T. van Goethem, S. Tajalizadehkhoob, M. Korczyński, and W. Joosen, “TRANCO: A Research-Oriented Top Sites Ranking Hardened Against Manipulation,” in 26th Annual Network and Distributed System Security Symposium, NDSS 2019, The Internet Society, 2019. doi: 10.14722/ndss.2019.23386.

[36] W. Aljedaani, A. Habib, A. Aljohani, M. Eler, and Y. Feng, "Does ChatGPT Generate Accessible Code? Investigating Accessibility Challenges in LLM-Generated Source Code," Proc. 21st Int. Web All Conf. (W4A '24), New York, NY, USA, 2024, pp. 165–176, doi: 10.1145/3677846.3677854.

[37] W. Aljedaani, M. M. Eler, and P. D. Parthasarathy, "Enhancing Accessibility in Software Engineering Projects with Large Language Models (LLMs)," Proc. 56th ACM Tech. Symp. Comput. Sci. Educ. (SIGCSETS 2025), New York, NY, USA, 2025, pp. 25–31, doi: 10.1145/3641554.3701841.

[38] WebAIM : Skip navigation links. (2021, 22 septembre). https://webaim.org/techniques/skipnav/

[39] WebAIM, “The WebAIM Million - The 2024 report on the accessibility of the top 1,000,000 home pages,” Mar. 28, 2024. [Online]. Available: https://webaim.org/projects/million/. Last updated: Mar 31, 2025

[40] Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: writing a literature review. MIS Q 8–23 (2002)

[41] Wohlin, C.: Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, p. 38. ACM (2014)

[42] “ArticleS.UncleBob.PrinciplesOfOOD.” http://butunclebob.com/ArticleS.UncleBob.PrinciplesOfOod"

[43] “CÓDIGO DE CONDUTA — P. PORTO | Ensino superior público.” [Online]. Available: https://www.ipp.pt/sobre/transparencia-integridade-anticorrupcao/codigo-de-conducta

[44] “DIRETIVA (UE) 2016/2102 DO PARLAMENTO EUROPEU e DO CONSELHO,” Jornal Oficial Da União Europeia, Oct. 26, 2016. [Online]. Available: https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32016L2102&from=EN [Accessed: Feb. 02, 2025].

[45] “Decreto-Lei n.o 83/2018, de 19 de outubro,” Diário Da República, Oct. 19, 2018. [Online]. Available: https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32016L2102&from=EN [Accessed: Feb. 02, 2025].

[46] “European accessibility act,” Employment, Social Affairs and Inclusion, Nov. 29, 2024. [Online]. Available: https://employment-social-affairs.ec.europa.eu/policies-and-activities/social-protection-social-inclusion/persons-disabilities/union-equality-strategy-rights-persons-disabilities-2021-2030/european-accessibility-act_en

[47] “IEEE Code of Conduct.” [Online]. Available: https://www.ieee.org/about/corporate/governance/code-of-conduct.html

[48] “IEEE Code of Ethics.” [Online]. Available: https://www.ieee.org/about/corporate/governance/p7-8.html

[49] “Object-oriented analysis and design with applications: Booch, Grady: Free Download, Borrow,and Streaming: Internet Archive,” Internet Archive, 1994.https://archive.org/details/objectorientedan00booc

[50] “Software architecture patterns,” O’Reilly Online Learning.https://www.oreilly.com/library/view/software-architecture-patterns/9781491971437/ch01.html

[51] “Understanding success Criterion 1.1.1: Non-text content | WAI | W3C.” https://www.w3.org/WAI/WCAG22/Understanding/non-text-content.html

[52] “Web Content Accessibility Guidelines (WCAG) 2.2,” W3C Recommendation, Oct. 05, 2023. [Online]. Available: https://www.w3.org/TR/WCAG22/ [Accessed: Oct. 18, 2024].

9 Anexos

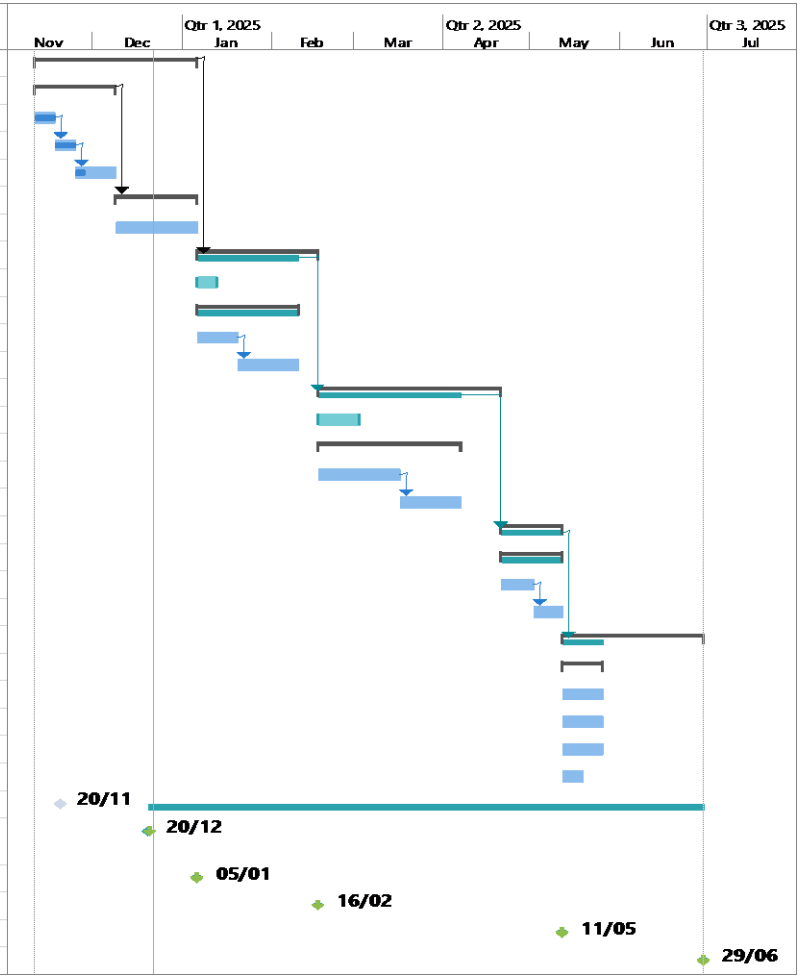
9.1 Anexo 1 – Procedimentos de monitorização e controlo

CRITÉRIO	*AGILE*	*WATERFALL*	*SCRUM*
Estrutura	Iterativa e incremental.	Linear e sequencial.	Estrutura definida dentro do Agile.
Flexibilidade	4 – Fácil adaptação a mudanças.	2 - Mudanças difíceis e de alto custo.	3 - Iterações curtas, porém estruturadas.
Foco no Cliente	4 - Com feedback contínuo do cliente.	2 - Com envolvimento do cliente no início e fim.	4 - Feedback fornecido no fim dos sprints.
Gestão de riscos	5 - Permite ajustes rápidos em resposta a riscos devido à sua natureza iterativa e adaptativa.	2 - Riscos precisam de ser identificados no início, com pouca flexibilidade para ajustes posteriores.	4 - Revisões regulares ajudam a identificar e mitigar riscos durante cada sprint.
Ciclo de Desenvolvimento	Iterações curtas e incrementais.	Etapas rígidas e sequenciais.	Baseado em sprints (1-4 semanas).

Legenda: 1 - Muito baixa; 2 – Baixa; 3 – Moderada; 4 – Alta; 5 – Muito Alta

9.2 Anexo 2 - Diagrama de Gantt do planeamento do projeto

Task Name	Duration	Start	Finish	Predecessors
Análise	**56 days**	**Mon 11/11/2**	**Sun 05/01/25**	
Identificação	**28 days**	**Mon 11/11/2**	**Sun 08/12/24**	
Identificar o âmbito do projeto	7 days	Mon 11/11/24	Sun 17/11/24	
Identificar os objetivos do projeto	7 days	Mon 18/11/24	Sun 24/11/24	3
Identificar as questões de investigação	14 days	Mon 25/11/24	Sun 08/12/24	4
Revisão da Literatura	**28 days**	**Mon 09/12/2**	**Sun 05/01/25**	2
Mapeamento do estudo	28 days	Mon 09/12/24	Sun 05/01/25	
Design	**42 days**	**Mon 06/01/2**	**Sun 16/02/25**	1
Desenvolvimento de competências	7 days	Mon 06/01/25	Sun 12/01/25	
Arquitetura	**35 days**	**Mon 06/01/2**	**Sun 09/02/25**	
Pesquisar modelos e frameworks relevantes	14 days	Mon 06/01/25	Sun 19/01/25	
Desenho da arquitetura	21 days	Mon 20/01/25	Sun 09/02/25	11
Desenvolvimento	**63 days**	**Mon 17/02/2**	**Sun 20/04/25**	8
Desenvolvimento de competências	14 days	Mon 17/02/25	Sun 02/03/25	
Protótipo	**49 days**	**Mon 17/02/2**	**Sun 06/04/25**	
Desenvolvimento do protótipo	28 days	Mon 17/02/25	Sun 16/03/25	
Teste e refinamento do protótipo	21 days	Mon 17/03/25	Sun 06/04/25	16
Avaliação	**21 days**	**Mon 21/04/2**	**Sun 11/05/25**	13
Avaliação do modelo em relação aos objetivos	**21 days**	**Mon 21/04/2**	**Sun 11/05/25**	
Testagem de usabilidade	11 days	Mon 21/04/25	Thu 01/05/25	
Análise Comparativa	10 days	Fri 02/05/25	Sun 11/05/25	20
Conclusões	**49 days**	**Mon 12/05/2**	**Sun 29/06/25**	18
Documentação	**14 days**	**Mon 12/05/2**	**Sun 25/05/25**	
Documentar descobertas	14 days	Mon 12/05/25	Sun 25/05/25	
Contribuição académica	14 days	Mon 12/05/25	Sun 25/05/25	
Contribuição prática	14 days	Mon 12/05/25	Sun 25/05/25	
Revisão	7 days	Mon 12/05/25	Sun 18/05/25	
Milestones	**0 days**			
Entrega da versão intermédia do trabalho preparatório do projeto	0 days	Fri 20/12/24	Fri 20/12/24	
Entrega do trabalho preparatório do projeto	0 days	Sun 05/01/25	**Sun 05/01/25**	
Entrega da proposta de arquitetura	0 days	Sun 16/02/25	**Sun 16/02/25**	
Entrega do protótipo desenvolvido	0 days	Sun 11/05/25	**Sun 11/05/25**	
Entrega do projeto	0 days	Sun 29/06/25	**Sun 29/06/25**	



9.3 Anexo 3 - Dicionário EAP / WBS

Dicionário EAP	Tipo	Descrição adicional / critérios de aceitação	Critérios de progresso
Análise	Fase	Constitui a primeira parte do projeto. Conclusão dependente da aprovação formal dos seus entregáveis.	Identificação – 50% Revisão da Literatura – 50%
Identificação	Entregável	Deverá ser feito o levantamento do âmbito, objetivos e questões de investigação do projeto.	Âmbito – 33% Objetivos – 33% Questões de investigação – 34%
Identificar o âmbito do projeto	Pacote de trabalho	Deverá ser levantado o contexto e problema e *stakeholders*.	Contexto e problema – 100%
Identificar os objetivos do projeto	Pacote de trabalho	Deverão ser levantados os objetivos, assim como as questões de investigação, identificando as métricas de avaliação e palavras-chave.	Questões de investigação – 20% por cada Métricas de avaliação – 10 % por cada Palavras-chave – 3% por cada
Revisão da Literatura	Entregável	Deverá ser feita uma revisão da literatura da área de investigação face à área de investigação do projeto.	Mapeamento do estudo – 100%
Mapeamento do Estudo	Pacote de trabalho	Deverá ser feito um levantamento dos documentos de investigação por forma a identificar lacunas e obter uma visão geral da área de pesquisa.	Mapeamento sistemático da literatura – 100%
Design	Fase	Constitui a segunda fase do desenvolvimento do projeto. Os conteúdos presentes e resultantes desta fase estão enquadrados na documentação da solução.	Análise de arquiteturas – 30 % Diagramas de arquitetura – 70%
Arquitetura	Entregável	Deverá ser feito o levantamento e documentação relativa à arquitetura da solução.	Modelo 4+1 – 100%
Pesquisa de modelos e *frameworks*	Pacote de trabalho	Deverá ser feita uma análise aos modelos e frameworks existentes por forma a proceder à escolha consciente e justificada das mesmas.	Análise de modelos e frameworks – 50%

Dicionário EAP	Tipo	Descrição adicional / critérios de aceitação	Critérios de progresso
			Escolha, documentação e comparação do modelo e framework -50%
Desenho da arquitetura	Pacote de trabalho	Deverá ser desenhada e documentada a arquitetura da solução, assim como gerados os artefactos relevantes para o efeito.	Modelo 4 + 1 – 100%
Desenvolvimento	Fase	Constitui a terceira fase do desenvolvimento do projeto. Os conteúdos presentes e resultantes desta fase estão enquadrados com a construção da solução propriamente dita.	Desenvolvimento do protótipo – 100%
Protótipo	Entregável	Deverá ser efetuada a construção da solução conforme o especificado no levantamento da arquitetura.	Desenvolvimento do protótipo – 90 % Refinamento do protótipo.
Desenvolvimento do protótipo	Pacote de trabalho	Deverá ser construído um protótipo da solução proposta.	Front-end – 40% Back-end -40% Comunicação e implementação da componente de IA – 20%
Refinamento do protótipo	Pacote de trabalho	Deverá ser feito o refinamento e ajuste ao protótipo por forma a obter uma solução o mais robusta e eficaz possível	Ajustes e refinamento – 100 %
Avaliação	Fase	Constitui a quarta fase do desenvolvimento. Nesta fase, pretende-se testar e comparar a solução produzida, fazendo o levantamento de resultados	Avaliação do modelo em relação aos objetivos – 100%
Avaliação do modelo em relação aos objetivos	Entregável	Deverão ser conduzidos testes ao modelo relativamente aos objetivos por forma a aferir resultados	Testagem de usabilidade – 50 % Análise comparativa – 50 %
Testagem de usabilidade	Pacote de trabalho	Deverá ser levada a cabo a testagem da solução a nível de usabilidade, comparativamente com ferramentas tradicionais.	Testes de usabilidade – 100%
Análise comparativa	Pacote de trabalho	Deverá ser efetuada a comparação e análise do modelo face às ferramentas e outros métodos tradicionais de solucionamento de violações de acessibilidade.	Análise comparativa -100 %

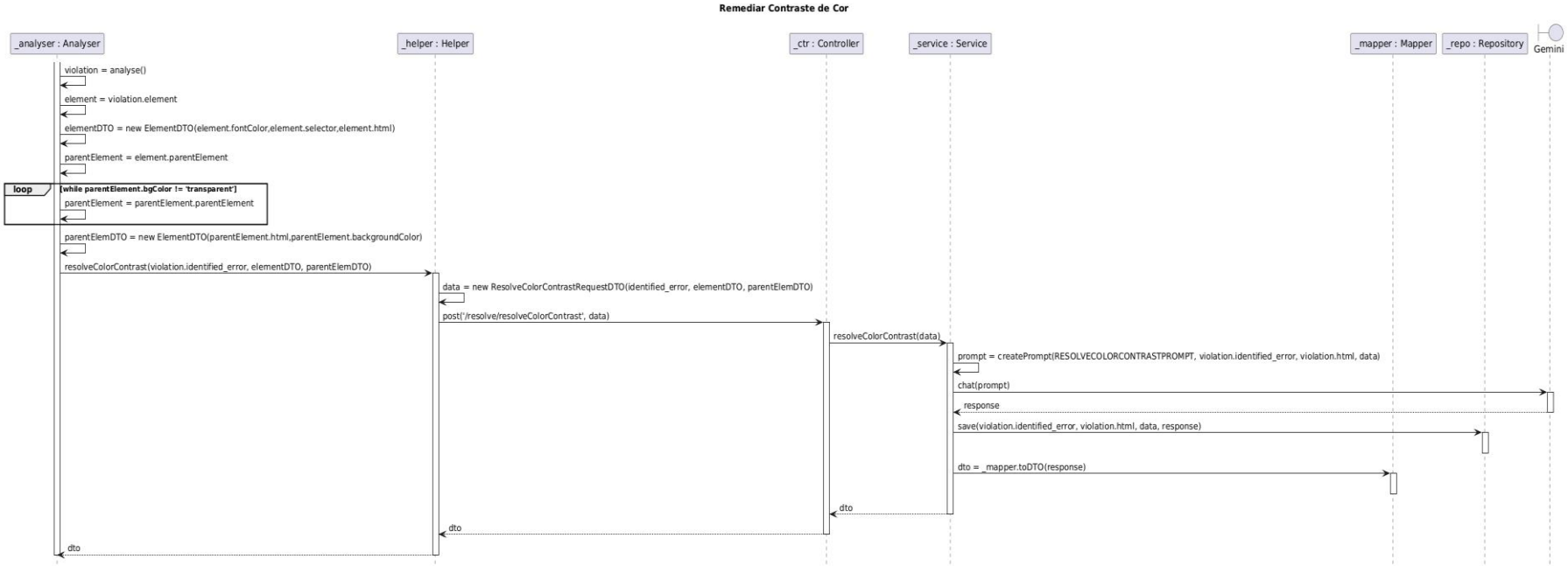
Dicionário EAP	Tipo	Descrição adicional / critérios de aceitação	Critérios de progresso
Conclusões	Fase	Constitui a última fase do desenvolvimento do projeto. Nesta fase, pretende-se fazer o levantamento das descobertas, contribuições, e uma revisão final do documento.	Documentação – 100%
Documentação	Entregável	Deverá ser feito o levantamento de quaisquer descobertas e contribuições feitas durante o desenvolvimento do projeto.	Documentação de descobertas – 30% Contribuições académicas – 30% Contribuições práticas – 30% Revisão – 10%
Documentação de descobertas	Pacote de trabalho	Deverá ser feito o levantamento das descobertas feitas durante o desenvolvimento do projeto, providas dos respetivos dados que suportem as teses.	Levantamento das descobertas – 60% Documentação de descobertas – 40%
Contribuições académicas	Pacote de trabalho	Deverá ser realizado o levantamento de quaisquer contribuições académicas feitas durante o desenvolvimento do projeto, assim como os dados que as suportam.	Levantamento das contribuições académicas – 60% Documentação das contribuições académicas – 40%
Contribuições práticas	Pacote de trabalho	Deverá ser realizado o levantamento de quaisquer contribuições práticas, isto é, para a área de investigação, feitas durante o desenvolvimento do projeto, assim como os dados que as suportam	Levantamento das contribuições práticas – 60% Documentação das contribuições práticas – 40%
Revisão do documento	Pacote de trabalho	Deverá ser conduzida uma análise e correção de eventuais erros ao documento.	Revisão do documento – 100%

9.4 Anexo 4 - Plano de ação

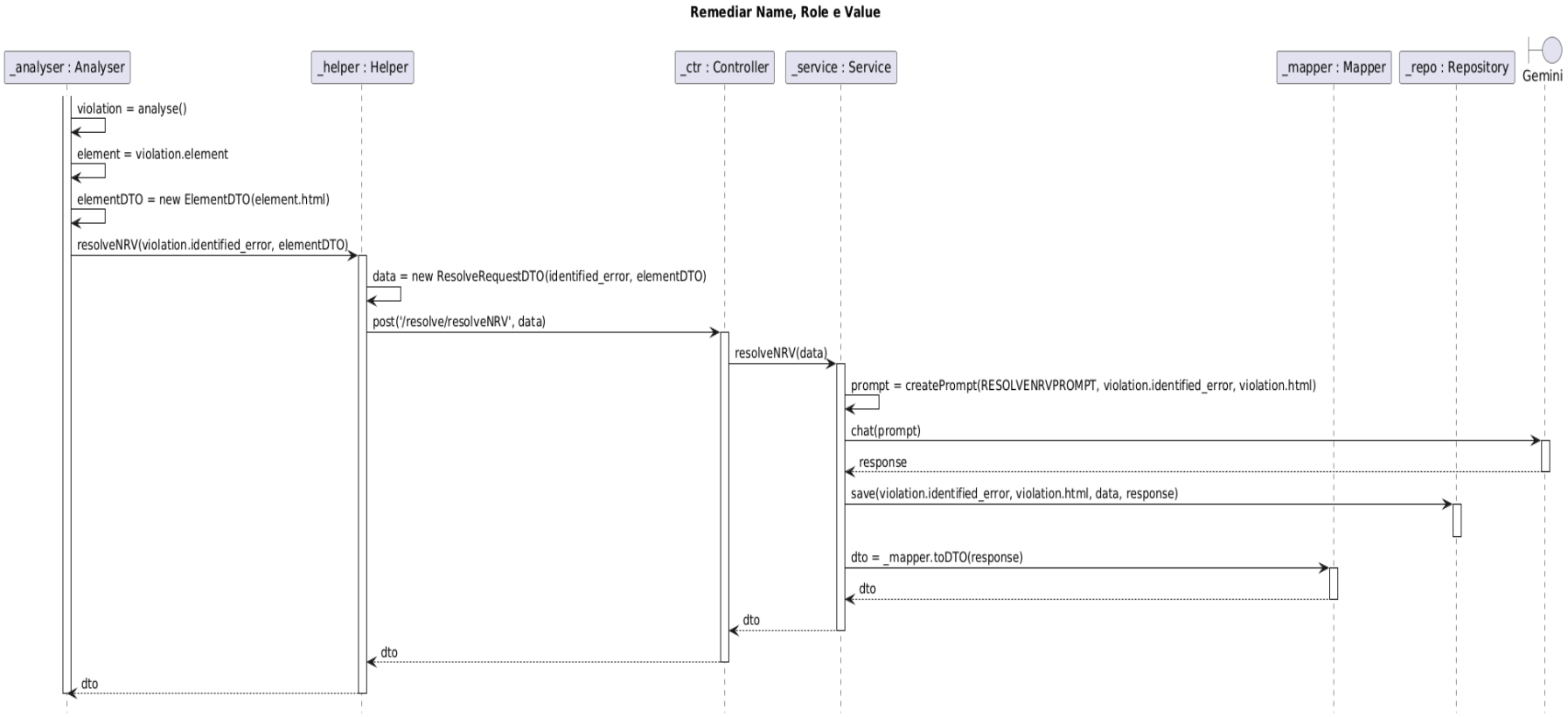
COMPETÊNCIA	OBJETIVO	AÇÕES	AVALIAÇÃO DO SUCESSO
Análise e resolução de problemas.	Desenvolver a habilidade de identificar e definir problemas de maneira clara e precisa, praticando com situações reais ou simuladas pelo menos uma vez por semana durante os próximos três meses.	Selecionar um problema por semana, dentro ou fora do ambiente de trabalho, e tentar definir claramente a sua origem, impacto e as partes envolvidas; Usar técnicas de definição de problemas, como o 5 Porquês (5 *Whys*) [17], que ajuda a encontrar a causa raiz.	Rever os problemas definidos semanalmente, verificando se a identificação foi clara e precisa, procurando *feedback* de terceiros para avaliar a clareza das definições.
	Melhorar a capacidade de análise crítica, através da prática de avaliação de causas e efeitos para problemas específicos, duas vezes por mês, durante os próximos três meses.	Aplicar técnicas de análise, por exemplo a análise SWOT; Investigar cada problema assumindo perspetivas distintas, tentando prever potenciais consequências das soluções.	Documentar cada análise e avaliar a profundidade das interpretações, além de analisar como estas análises influenciaram a tomada de decisão final.
Gestão de tempo	Estabelecer um cronograma detalhado para os próximos três meses do projeto, definindo metas intermediárias semanais e prazos para cada etapa.	Dividir o projeto em etapas e definir prazos específicos para cada uma delas; Reservar tempo semanalmente para rever o progresso do cronograma e fazer ajustes conforme necessário; Utilizar ferramentas de gestão de projetos, para visualizar o cronograma; Criar e seguir um plano diário de tarefas prioritárias semanalmente;	Rever os problemas definidos semanalmente, verificando se a identificação foi clara e precisa, procurando *feedback* de terceiros para avaliar a clareza das definições.
Gestão de tempo	Implementar um sistema de revisão diário e semanal do planeamento para ajustar metas e prioridades, ao longo de três meses.	Reservar 10 minutos no final de cada dia para rever o que foi cumprido e ajustar o plano para o dia seguinte; Fazer uma revisão mais aprofundada no final de cada semana, analisando o	Ao final dos três meses, avaliar a consistência na revisão e o impacto dessa prática no cumprimento das metas.

COMPETÊNCIA	OBJETIVO	AÇÕES	AVALIAÇÃO DO SUCESSO
Comunicação escrita	Escrever textos claros e objetivos, livres do uso de frases longas e palavras desnecessárias na escrita de documentos nos próximos três meses.	progresso nas metas de curto e longo prazo; Ajustar o plano semanal e mensal com base nas lições aprendidas e nos novos requerimentos que surgirem. Praticar a escrita de frases curtas e diretas, evitando jargões ou linguagem complexa desnecessária; Rever todos os textos para identificar e cortar palavras e frases redundantes; Pedir feedback a terceiros para avaliar se a mensagem está clara e direta.	Comparar os textos antes e depois do período de três meses, avaliando a redução de frases longas e o aumento da clareza.
Comunicação escrita	Ampliar o vocabulário e enriquecer a comunicação escrita, aprendendo e aplicando pelo menos cinco novas palavras ou expressões por semana, durante três meses.	Ler textos variados para identificar novas palavras e expressões; Anotar palavras e expressões interessantes e praticar a sua utilização em textos; Rever o próprio texto e procurar um vocabulário variado e adequado ao contexto.	Ao final dos três meses, avaliar a consistência na revisão e o impacto dessa prática no cumprimento das metas.

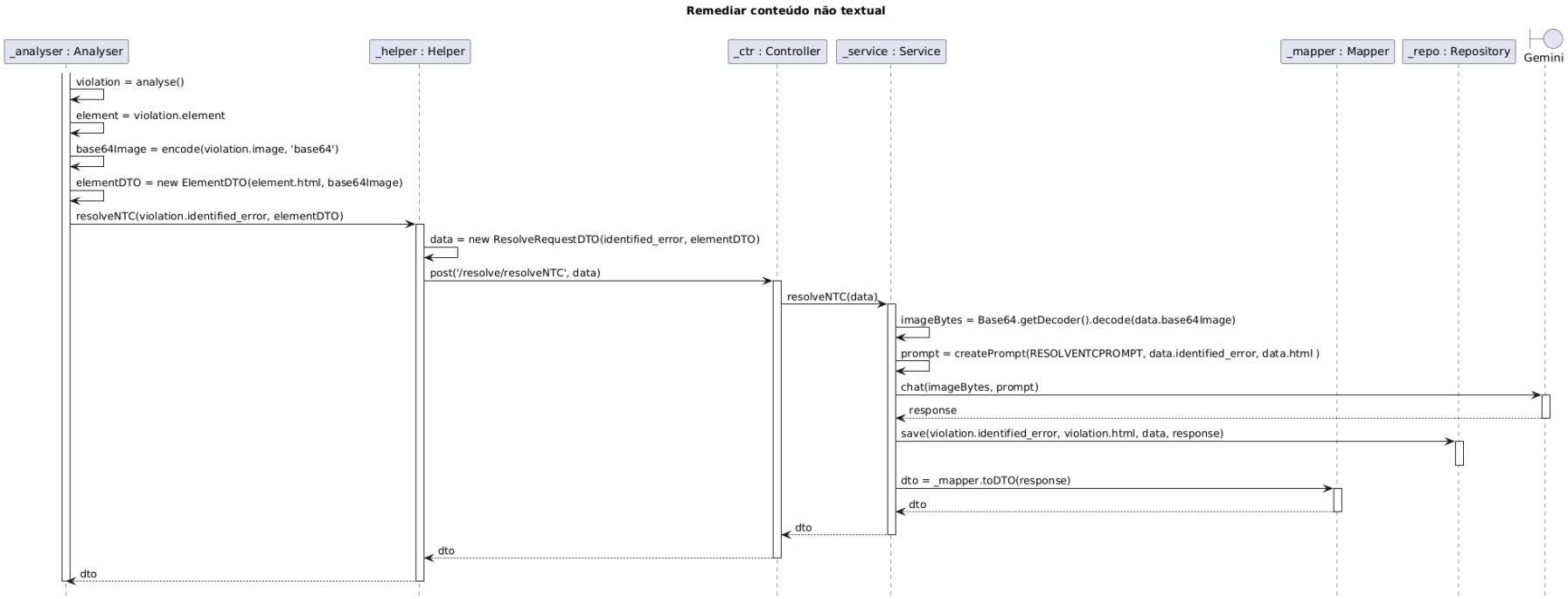
9.5 Anexo 5 – Remediar Contraste de Cor (Nível 2)



9.6 Anexo 6 – Remediar *Name, Role* e *Value* (Nível 2)



9.7 Anexo 7 – Remediar conteúdo não textual (Nível 2)



9.8 Anexo 8 – Remediar *links* (Nível 2)