Orientações para Descrição de Fotografias Adaptadas para Inteligência Artificial no Contexto de Mídias Sociais

Carolina Sacramento[1,2], Simone Bacellar Leal Ferreira[2], Priscilla Fonseca de Abreu Braz[3], Sara Lobato[2,4], João Marcelo dos Santos Marques[2,5]

[1]Casa de Oswaldo Cruz
Fundação Oswaldo Cruz (Fiocruz)
Rio de Janeiro – RJ – Brasil

[2]Departamento de Informática Aplicada
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Rio de Janeiro – RJ – Brasil

[3]Departamento de Informática e Ciência da Computação
Universidade do Estado do Rio de Janeiro (UERJ)
Rio de Janeiro – RJ – Brasil

[4]Instituto de Comunicação e Informação Científica e Tecnológica em Saúde
Fundação Oswaldo Cruz (Fiocruz)
Rio de Janeiro – RJ – Brasil

[5]Seção de Serviços de Informática
Instituto Brasileiro de Geografia e Estatística (IBGE)
Maceió – AL – Brasil

carolina.sacramento@fiocruz.br, simone@uniriotec.br

priscilla.abreu@ime.uerj.br, sara.lobato@edu.unirio.br

joao.marques@ibge.gov.br

Abstract. Introduction: *The evolution of generative artificial intelligence (AI) systems has enabled significant advances in digital accessibility, particularly in image description, making visual content more accessible to people with visual impairments.* ***Objective:*** *In this study, we proposed a set of guidelines to support the description of photographs in the context of online social media, based on previous works but focused on generative AI models.* ***Methodology:*** *The guidelines were developed in previous work and adapted for use with a widely adopted AI model. Two AI-generated photo descriptions, created based on these guidelines, were evaluated by two individuals with total visual impairment. Participants assessed the ease of understanding and compared the descriptions generated using the proposed guidelines with those produced using basic instructions.* ***Results:*** *Findings from the users' assessment enabled an initial refinement of the guidelines.*
Keywords *Accessibility, Image description, Alternative text, Generative AI, People with visual impairments*

Resumo. Introdução: *A evolução dos sistemas de inteligência artificial (IA)*

generativa tem permitido avanços significativos na acessibilidade digital, especialmente na descrição de imagens, tornando conteúdos visuais mais acessíveis a pessoas com deficiência visual. ***Objetivo:*** *Neste estudo, propõe-se um conjunto de orientações para apoiar a descrição de fotografias no contexto de mídias sociais online, com base em trabalhos anteriores, mas focadas em modelos de IA generativa.* ***Metodologia:*** *As orientações foram desenvolvidas em trabalhos prévios e adaptadas para uso com um modelo de IA amplamente adotado. Duas descrições de fotografias, geradas com base nessas orientações, foram avaliadas por duas pessoas com deficiência visual total. As pessoas voluntárias avaliaram a facilidade de compreensão e compararam as descrições geradas a partir das orientações adaptadas com descrições produzidas usando instruções básicas.* ***Resultados:*** *As avaliações empreendidas possibilitaram um refinamento inicial das orientações propostas.*
Palavras-Chave *Acessibilidade, Descrição de imagens, Texto alternativo, IA generativa, Pessoas com deficiência visual*

1. Introdução

As mídias sociais *online* são plataformas digitais que possibilitam a criação e o compartilhamento de conteúdo gerado pelos próprios usuários, como vídeos, fotos, *podcasts*, entre outros [Kaplan e Haenlein 2010]. Em 2023, a Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD Contínua) constatou que 83,5% das pessoas acessaram a internet para utilizar essas plataformas [IBGE 2024]. Pessoas com deficiência visual também estão ativamente presentes nesses ambientes [Pedrosa 2015, Moraes 2018], participando das interações, bem como consumindo e produzindo conteúdo [Wu e Adamic 2014].

A deficiência visual abrange diferentes condições que comprometem a percepção visual e é comumente classificada em dois tipos: cegueira, que pode variar desde a ausência total de visão até a percepção de vultos [IFPB 2018], e baixa visão, uma condição intermediária entre a cegueira e a visão plena, que não pode ser corrigida por tratamento clínico, cirúrgico ou com o uso de óculos convencionais [IFPB 2018, Gasparetto 2007].

Esta deficiência também pode ser definida quanto à sua origem: congênita, quando a perda da visão ocorre antes dos cinco anos de idade, e adquirida, quando ocorre após essa idade [Nunes e Lomônaco 2008]. A delimitação da cegueira adquirida baseia-se em estudos que observaram a presença de memória visual em pessoas que perderam a visão após os cinco anos, em contraste com as que perderam antes dessa fase [Amiralian 1997].

Tanto a cegueira, quanto a baixa visão requerem recursos de acessibilidade específicos para assegurar a navegação autônoma e a participação plena em ambientes digitais, como as mídias sociais *online*. Um desses recursos, utilizado por pessoas cegas e, em alguns casos, por pessoas com baixa visão, é o *software* leitor de telas, que interpreta o conteúdo exibido na tela e, com o auxílio de sintetizadores de voz, converte as informações visuais em áudio [Ferreira et al. 2006].

Apesar da existência desses recursos de acessibilidade, a interação de pessoas com deficiência visual com as mídias sociais nem sempre é bem-sucedida, especialmente no que diz respeito ao conteúdo visual, um dos principais tipos de conteúdo compartilhados

nessas plataformas [Mohanbabu e Pavel 2024]. Usuários de leitores de tela dependem de descrições para compreender as imagens, uma vez que esses *softwares* ainda não são capazes de gerar descrições de qualidade satisfatória [Jandrey et al. 2021]. Além da limitação dos leitores de tela, há ainda um outro problema: a ausência ou baixa qualidade das descrições fornecidas em imagens disponíveis *online*. De acordo com o relatório *WebAIM Million* de 2025, que realizou uma avaliação de acessibilidade das páginas iniciais de 1.000.000 de sites, 55,5% das páginas analisadas apresentaram imagens sem texto alternativo, e 13,4% das que possuíam esse recurso continham descrições vagas ou genéricas, como "imagem", "gráfico" ou nomes de arquivos [WebAIM 2025].

Especificamente no contexto das mídias sociais, diversas estratégias têm sido adotadas para ampliar a acessibilidade de conteúdos visuais para usuários de leitores de tela, variando desde descrições manuais (humanas) feitas por pessoas videntes [Morash et al. 2015, Salisbury et al. 2017], até descrições geradas automaticamente pelas próprias plataformas [Gleason et al. 2020, Wu et al. 2017], além de abordagens híbridas que combinam esforços humanos e automação [Mack et al. 2021, Singh et al. 2024].

A inteligência artificial (IA) também vem sendo utilizada pelas pessoas com deficiência para acessibilizar imagens [Berton et al. 2024, Furuya 2024]. Em especial, a inteligência artificial generativa: ramo da IA capaz de produzir novos conteúdos a partir de dados de entrada, como textos, imagens ou áudios [Bommasani et al. 2022]. Ferramentas como *GPT-4*[1], *Gemini*[2] e *LLaVa*[3] são exemplos de modelos de inteligência artificial generativa que permitem converter imagens em descrições textuais detalhadas sob demanda. Nesses sistemas, as descrições de imagens podem ser acionadas a partir de *uploads* de imagens diretamente nas interfaces conversacionais ofertadas pelas ferramentas, a partir de integração com outros *softwares*, como o *BeMyEyes*[4] ou mesmo integrados a leitores de tela, como no caso do plugin do leitor de telas NVDA (*NonVisual Desktop Access*) com *OpenAI* [Clause 2025].

Apesar de úteis na promoção da acessibilidade, tais ferramentas ainda enfrentam limitações quanto à oferta de descrições adequadas, uma vez que muitas delas geram conteúdo sem considerar o contexto em que a imagem está inserida, o que compromete a relevância e a precisão das informações fornecidas [Mohanbabu e Pavel 2024]. Além disso, especialistas destacam que, embora modelos de inteligência artificial generativa sejam capazes de produzir descrições com rapidez, eles não garantem a precisão e a adequação exigidas em contextos mais complexos, tornando necessária a validação humana [Berton et al. 2024].

Diante dessas limitações, considerar estratégias consolidadas como a audiodescrição (AD) pode agregar qualidade às descrições de imagem geradas por inteligência artificial. A AD é uma técnica muito popular no contexto audiovisual que consiste na conversão de informações visuais em linguagem verbal, com a participação de um roteirista com formação na técnica (audiodescritor), responsável pela elaboração da descrição, e de um consultor com deficiência visual, também com conhecimentos na técnica, que valida o conteúdo produzido. O objetivo da AD é ampliar a compreensão de

[1]https://openai.com/index/gpt-4/
[2]https://gemini.google.com/
[3]https://llava-vl.github.io/
[4]https://www.bemyeyes.com/pt-br/

imagens estáticas ou em movimento, textos e sons fora de contexto, especialmente para pessoas que não utilizam a visão como principal canal de percepção [ABNT 2016].

No contexto da IA generativa, fornecer orientações claras e bem estruturadas aos modelos pode ser útil para garantir resultados mais relevantes, uma vez que esses sistemas respondem de maneira mais eficaz quando recebem instruções precisas e contextualizadas. A *OpenAI* destaca, em suas diretrizes para criação de instruções para GPTs (do inglês *Generative Pre-trained Transformer*) personalizados, que instruções bem formuladas são fundamentais para obter respostas mais alinhadas aos objetivos do usuário [OpenAI 2025].

Este artigo apresenta a continuidade de uma pesquisa, que vem sendo desenvolvida há alguns anos, com o propósito de investigar estratégias que favoreçam a produção de descrições de imagens no contexto das mídias sociais *online*, considerando as necessidades e preferências de pessoas com deficiência visual. Neste momento, busca-se aprofundar a compreensão sobre as preferências e formas de interação desse público com descrições geradas por um modelo de IA generativa, dado a crescente popularidade dessas ferramentas entre pessoas usuárias de leitores de tela.

Assim, o objetivo deste estudo foi analisar se descrições de fotografias, geradas por um modelo de inteligência artificial popular, com base em orientações estruturadas previamente desenvolvidas, são adequadas à compreensão por pessoas com deficiência visual total. Buscou-se ainda investigar a preferência deste público entre descrições geradas com base nas orientações e descrições produzidas a partir de instruções básicas. Com base nos resultados, foram propostas orientações para descrição de fotografias adaptadas à inteligência artificial no contexto de mídias sociais.

Este estudo dialoga diretamente com os Grandes Desafios de Interação Humano-Computador (IHC), o GranDIHC-BR, especialmente com o desafio "*GC6 (Implicações da Inteligência Artificial na IHC: Uma Discussão sobre Paradigmas, Ética e Diversidade, Equidade e Inclusão)*", ao investigar o uso da IA na descrição de imagens no contexto de mídias sociais. A proposta de um conjunto de orientações adaptadas à IA para descrever fotografias nesse contexto está alinhada às ações de curto prazo da agenda de pesquisa indicada no referido desafio, ao "*fomentar um ecossistema de IA no campo da IHC que seja inclusivo, equitativo e considere especificidades culturais e contextos sociais brasileiros*" [Duarte et al. 2024].

2. Trabalhos Relacionados

2.1. Iniciativas na Literatura

O trabalho de [Jandrey et al. 2021] apresentou uma revisão sistemática de literatura, utilizando a técnica de *snowballing*, para identificar limitações nas descrições de imagens voltadas a pessoas com deficiência visual. O estudo destacou que, embora essas descrições sejam essenciais para promover acessibilidade, muitas vezes elas são inexistentes, genéricas, imprecisas ou insatisfatórias para o público-alvo. Foram identificados treze tipos de problemas recorrentes em descrições de imagens, entre eles textos insuficientes, errados, que não descrevem apropriadamente as características, expressões faciais, corporais ou ações das pessoas, entre outros. Os autores também discutiram em recomendações para a melhoria para alguns dos problemas encontrados,

como a necessidade de mídias sociais fornecerem instruções sobre como descrever uma imagem e o uso de perguntas estruturadas para apoiar a descrição de imagens.

Considerando a perspectiva do uso de questões estruturadas, a pesquisa de [Salisbury et al. 2017] buscou identificar um conjunto de perguntas que refletissem as demandas de pessoas com deficiência visual na descrição de imagens. Para isso, foi conduzido um experimento com trabalhadores videntes da plataforma *Amazon Mechanical Turk*, divididos em dois grupos: o primeiro simulava deficiência visual e deveria formular perguntas sobre imagens descritas automaticamente; o segundo grupo respondia às perguntas por meio de uma interface de conversação. A partir da análise das interações, os autores propuseram um conjunto de oito questões canônicas, que capturavam elementos essenciais da imagem. Essas perguntas foram posteriormente validadas por participantes cegos, que confirmaram sua pertinência ao reconhecer nelas dúvidas comuns entre usuários reais, indicando sua utilidade como apoio à geração de descrições acessíveis.

No contexto da inteligência artificial, [MacLeod et al. 2017] investigaram como pessoas cegas experienciam descrições automáticas de imagens em mídias sociais, com ênfase na confiança do usuário. Foram realizadas observações contextuais e entrevistas com seis usuários, para melhor entendimento das experiências vivenciadas pelo público com deficiência e, posteriormente um experimento *online* com 100 pessoas. Os autores concluíram que os participantes do experimento frequentemente aceitavam descrições imprecisas como corretas, preenchendo lacunas com base em suposições pessoais. A pesquisa revelou os riscos de confiança excessiva em sistemas automatizados sem mecanismos de validação.

A pesquisa de [Mohanbabu e Pavel 2024] apresentou um sistema de geração de descrições de imagens sensíveis ao contexto, voltado para páginas web, integrando *GPT-4* com uma extensão para o navegador *Chrome*. As descrições foram geradas a partir de elementos como título da página, URL e trechos textuais adjacentes à imagem. As descrições geradas pelo sistema foram avaliadas por doze pessoas cegas ou com baixa visão, por meio de um questionário que considerava quatro critérios principais: qualidade, relevância, plausibilidade e facilidade de imaginar a imagem. As descrições contextuais foram, em geral, preferidas em relação às tradicionais, destacando a importância do contexto em descrições geradas por IA generativa.

Um outro trabalho de destaque, porém com foco em imagens educacionais, foi o de [Perdigão et al. 2023]. O objetivo do artigo foi analisar os resultados obtidos ao instruir a IA *ChatGPT* a descrever imagens de um material didático de curso superior a distância, a partir da perspectiva de um consultor em audiodescrição com deficiência visual. Das descrições das imagens, nenhuma obteve avaliação plenamente positiva, com classificações variando de média a ruim. Os autores concluíram que o *ChatGPT* não pode ser considerado uma ferramenta confiável para audiodescrição, embora possam existir perspectivas de uso complementar da IA na elaboração de roteiros.

2.2. Estudos Anteriores

Esta pesquisa foi desenvolvida em continuidade a trabalhos anteriores [Sacramento e Ferreira 2022, Sacramento et al. 2022, Nardi 2021]. Em [Sacramento e Ferreira 2022] buscou-se aprofundar o comportamento adotado por

pessoas cegas congênitas no consumo e na publicação de conteúdo visual em mídias sociais, bem como suas preferências em relação aos formatos das descrições. Participaram onze pessoas autodeclaradas cegas congênitas, que detalharam diversos aspectos de uso e interação com as mídias sociais, com destaque às seguintes observações sobre preferências de descrição de imagens:

- Dificuldade de alguns voluntários em diferenciar descrições feitas por humanos ou por *softwares*, bem como a autoria da descrição, sendo relevante indicá-la.
- Relevância da adoção de uma abordagem em que os detalhes são fornecidos progressivamente no contexto das mídias sociais. Apesar de a maioria preferir uma abordagem mais simples e direta, dependendo do tipo de conteúdo visual e interesse em saber mais sobre a imagem, acessar uma descrição detalhada pode ser necessário.
- Preferência da maioria pela oferta de descrição em formato textual (para consumo de leitores de tela);
- Interesse na oferta de outros recursos que possam contribuir com a composição das descrições, por exemplo, sons ambientes.

No trabalho [Sacramento et al. 2022], foram realizados exercícios de descrição com a participação das mesmas pessoas cegas do estudo anterior e de onze participantes videntes. O objetivo foi identificar particularidades, semelhanças e diferenças na forma como ambos os grupos elaboram descrições que poderiam servir como alternativas ao conteúdo visual em mídias sociais *online*. O estudo investigou se a perspectiva de pessoas cegas congênitas, que possuem maneiras específicas de interagir com o mundo, poderia influenciar a forma como descrevem elementos visuais, representados por fotografias.

No exercício, solicitou-se aos participantes cegos e videntes que fizessem descrições de: uma *selfie* de rosto; uma situação específica (ida à uma praia); características utilizadas na descrição de uma pessoa e de um local desconhecidos; e quatro elementos concretos (laranja, escova de dentes, edifício e fumaça de cigarro). As descrições feitas pelos voluntários foram submetidas à análise de conteúdo, que resultou em uma lista de categorias (ou atributos) que compunham a descrição.

Os atributos gerados por cegos e videntes foram então submetidos a uma comparação estatística. Os resultados demonstraram pouca diferença entre os perfis na descrição de *selfie* e de praia, onde os cegos tenderam a descrever da mesma forma que recebem descrições nas mídias sociais. Já na descrição de local novo/desconhecido, os cegos fizeram descrições mais próxima da realidade de interação deles com o mundo. O mesmo em relação à descrição de uma pessoa desconhecida. Nos demais elementos, o único que apresentou diferença nas descrições foi fumaça de cigarro. Os atributos obtidos em [Sacramento et al. 2022] foram então comparados com um conjunto de questões propostas por [Salisbury et al. 2017] para descrição de imagens, resultando em uma adaptação das questões para o público em estudo.

Por fim, o trabalho [Nardi 2021] trouxe um conjunto de orientações baseadas nas questões adaptadas, considerando também o conteúdo de formações sobre audiodescrição e uma revisão de literatura *ad hoc* sobre boas práticas para descrição de imagens. As questões propostas consideraram duas abordagens: uma simples e outra detalhada, inspirada na proposta de [Morris et al. 2018], que propôs novas formas de apresentação

de descrições de imagens, incluindo detalhes progressivos, que no estudo em questão foi a mais bem avaliada quando submetida a testes com usuários cegos.

Este trabalho incluiu uma avaliação das orientações criadas, com a participação de cinco pessoas videntes e duas pessoas cegas (uma congênita e outra adquirida), nenhum deles especialista em audiodescrição. Este recorte foi feito por considerar que as questões seriam úteis para pessoas leigas na atividade de descrever imagens.

A avaliação dos videntes consistiu na descrição de uma fotografia postada pelo próprio voluntário em uma mídia social *online* a partir de duas abordagens: livre e orientada por questões. As descrições criadas foram então apresentadas aos voluntários com deficiência visual, que avaliaram se houve melhor entendimento quando a alternativa foi gerada a partir das orientações propostas ou de forma livre.

O resultado da avaliação revelou não existir uma abordagem de preferência unânime entre os videntes. Já as pessoas cegas preferiram as descrições feitas na abordagem livre, por julgarem mais naturais, apesar de sentirem falta de informações importantes nas descrições. As discussões apresentadas na avaliação permitiram concluir que o uso exclusivo de questões norteadoras pareceu mais relevante para pessoas que não tenham habilidade ou não saibam nem como começar uma descrição. E que as questões deveriam servir como suporte às descrições livres e não em substituição. Além disso, o resultado permitiu identificar uma série de ajustes nas orientações propostas.

Uma contribuição importante da avaliação feita pelas voluntárias com deficiência foi a reprovação de experiências sensoriais como parte de uma descrição textual. Apesar de a avaliação ter envolvido apenas duas pessoas, trabalhos anteriores já haviam apontado uma tendência de a pessoa com deficiência descrever elementos e situações da mesma forma com que recebem as descrições [Sacramento e Ferreira 2022]. Outra contribuição das pessoas com deficiência, que já havia sido destacada no trabalho [Sacramento e Ferreira 2022], é que a descrição detalhada deve ser um recurso configurável, a ser acionado de acordo com o interesse da pessoa.

3. Método de Pesquisa

A presente pesquisa, de natureza exploratória e com abordagem qualitativa, é uma continuidade dos trabalhos anteriores apresentados na Subseção 2.2. Ela foi composta pela avaliação de descrições de imagens geradas por inteligência artificial, com participação de pessoas com deficiência visual total, que analisaram descrições de imagens geradas a partir de questões estruturadas, incluindo uma análise comparativa com descrições geradas por instruções básicas. Com os resultados obtidos, foram propostos um conjunto de orientações para descrição de imagens por modelos de IA. As etapas envolvidas na pesquisa foram:

1. **Definição do sistema de IA generativa:** onde o modelo de inteligência artificial *ChatGPT 4 plus*, da *Open AI*[5], foi escolhido para uso na pesquisa. A escolha foi norteada pela percepção, em estudo anterior [Jacques et al. 2025], de que esta ferramenta apresentou descrições mais adequadas, em comparação com outras estudadas. Além disso, o *ChatGPT* foi considerado por pessoas com deficiência visual e especialistas como uma ferramenta potencial para descrição

[5]https://chatgpt.com/

de imagens [Berton et al. 2024], bem como trata-se do modelo de IA utilizado pelo *BeMyEyes*, aplicativo muito popular entre pessoas com deficiência visual para a tarefa de descrever imagens [Furuya 2024]. Utilizou-se o recurso de GPT personalizado via interface do usuário.

2. **Adaptação das orientações de [Nardi 2021] para uso em modelo de inteligência artificial:** onde foram feitos ajustes nas orientações apresentadas em [Nardi 2021], tornando-as mais direcionadas ao consumo de modelos de inteligência artificial. Para tal, foram utilizadas diretrizes da *OpenAI* para a escrita de instruções destinadas a GPT personalizados [OpenAI 2025]. A adaptação está detalhada na Seção 4.
3. **Definição das imagens para avaliação:** onde foram definidas as imagens utilizadas na avaliação por pessoas com deficiência visual. Optou-se por fotografias contendo elementos baseados no exercício proposto em [Sacramento et al. 2022], tais como: a *selfie* de uma pessoa, uma praia e uma pessoa em local não conhecido. Apesar de [Sacramento et al. 2022] também incluir a descrição de objetos, optou-se por não os considerar na presente pesquisa por dois motivos: tornar a análise por pessoa com deficiência visual mais extensa, o que poderia levar à fadiga das pessoas voluntárias, interferindo no resultado da avaliação [Barbosa et al. 2021] e por considerar que a fotografia de um objeto não apresenta a mesma complexidade e quantidade de elementos que as demais apresentam. As fotografias foram obtidas de um repositório de imagens (o *Freepik*[6]), que permite uso de imagem gratuitamente, desde que sejam feitas atribuições ao autor. A Figura 1 apresenta uma montagem com as fotografias utilizadas;
4. **Definição do perfil e recrutamento dos participantes do estudo:** optou-se por recrutar pessoas com deficiência visual total adquirida (cegos adventícios) com memória visual preservada. Apesar de ser um perfil diferente dos trabalhos anteriores [Sacramento e Ferreira 2022, Sacramento et al. 2022], que privilegiaram cegos congênitos, decidiu-se trabalhar com cegos adventícios devido à intenção de empreender uma investigação inicial com pessoas que demandam descrição de imagens, mas que ainda conseguem formar um modelo mental das situações representadas nas fotografias. Durante o período de recrutamento foram enviados convites para diversas pessoas da rede de colaboração dos pesquisadores pessoalmente ou via *WhatsApp*, incluindo participantes de estudo anterior [Sacramento et al. 2020].
5. **Avaliação de descrições geradas por IA com participação de pessoas com deficiência visual:** composta por entrevistas semiestruturadas com a participação de pessoas com deficiência visual total, que utilizam leitores de telas para interagir com imagens digitais. As fotografias selecionadas, foram acompanhadas de duas descrições geradas pela IA: a primeira, a partir de instruções básicas ("*descreva a fotografia para uma pessoa com deficiência visual*") e, a segunda, utilizando como instrução, as orientações propostas em [Nardi 2021], adaptadas para uso por modelos de IA, que foram apresentadas aos voluntários, entrevistados na sequência. Esta etapa está detalhada na Seção 5.
6. **Proposição de Orientações para Descrição de Imagens por Sistemas de**

[6]https://www.freepik.com/

Inteligência Artificial: a partir dos ajustes e correções identificados nas avaliações realizadas, as orientações são apresentadas na Seção 6.

Figura 1. Fotografias utilizadas na análise: (a) Registro de uma paisagem. Crédito: lifeforstock via Freepik; (b) Registro de uma selfie. Crédito: wayhomestudio via Freepik; e (c) Registro de uma pessoa em local não conhecido. Crédito: Freepik.

3.1. Cuidados Éticos

Esta pesquisa adotou cuidados éticos na condução da entrevista, como o envio antecipado e a leitura de um Termo de Consentimento Livre e Esclarecido (TCLE) aos voluntários. O TCLE contemplou o objetivo e o procedimento do estudo, a garantia de participação voluntária, com possibilidade de desistir a qualquer momento, os riscos e benefícios da participação, a garantia de anonimato e de confidencialidade dos dados dos participantes.

3.2. Limitações do Método

Dentre as limitações do método apresentado está a quantidade de participantes na avaliação: mais participantes certamente contribuiriam com o reforço dos pontos abordados e novas discussões não exploradas. Além disso, a validação das descrições geradas por IA foram feitas por pessoas não especialistas em descrição de imagens. A participação de especialistas (consultores) poderia ter contribuído com outras perspectivas na discussão dos resultados. Deve-se destacar que essa limitação é imposta pela dificuldade de se conseguir voluntários com deficiência visual em participar dos testes, principalmente com a especificidade do perfil definido para o estudo. Muitos convites foram feitos durante período de recrutamento, com número muito pequeno de aceites recebidos, em alguns casos por conflito de agenda dos voluntários.

No que se refere ao perfil das pessoas participantes, o estudo foi direcionado a pessoas com cegueira adquirida e memória visual preservada, não incluindo outros grupos, como pessoas cegas congênitas, que podem apresentar necessidades e percepções diferentes em relação às descrições de imagens. A opção por este perfil, neste momento, teve como objetivo incluir participantes com referências visuais prévias.

Outro aspecto limitante foi o uso de fotografias de diferentes naturezas na avaliação. Enquanto o trabalho [Nardi 2021] utilizou fotografias autorais dos voluntários videntes, com intuito de permitir a utilização de outros sentidos na descrição das imagens, nesta avaliação optou-se por fotografias de bancos de dados de imagens *online*. A escolha foi feita por considerar que o apoio da IA na descrição das imagens não permite a vivência dos outros sentidos explorados na primeira avaliação. Ainda assim, por não estarem diretamente relacionadas ao interesse dos participantes, as imagens selecionadas podem ter influenciado a percepção sobre a relevância prática das descrições.

Cabe ainda mencionar o uso de apenas um modelo de inteligência artificial para gerar as descrições. Embora esse modelo seja amplamente utilizado pelo público-alvo da pesquisa, a inserção de outros modelos poderia trazer novas perspectivas para o estudo.

4. Adaptação das Orientações para Inteligência Artificial

As orientações apresentadas em [Nardi 2021] foram adaptadas às diretrizes da *OpenAI* [OpenAI 2025] especialmente no que diz respeito a:

- **Reestruturação das instruções iniciais:** as orientações foram divididas em etapas, no intuito de promover maior granularidade e, consequentemente, simplificá-las para execução via modelo de inteligência artificial;
- **Reformulação para uma abordagem positiva:** quando possível, as instruções foram reescritas utilizando linguagem positiva. As diretrizes da *OpenAI* [OpenAI 2025] orientam evitar negações, pois estas poderiam causar ambiguidade ou confusão ao modelo;
- **Uso de marcadores para melhorar a legibilidade:** as seções foram identificadas com o caractere adequado (#) e os passo-a-passos foram padronizados em listas numeradas utilizando algarismos arábicos (1, 2, 3 etc.).

Outra adaptação realizada foi a inclusão do contexto das mídias sociais nas instruções, considerando a importância dele ao descrever imagens [Mohanbabu e Pavel 2024]. Além disso, foi preciso explicitar a necessidade de produzir dois tipos de descrição: uma simples (baseada nas questões iniciais) e uma detalhada (baseada em todas as questões), pois em [Nardi 2021] isso é dito de maneira implícita ("*Esteja atento ao nível de detalhe sugerido*").

Por fim, foi necessário incluir nas "Dicas adicionais" um item indicando que, caso a resposta à uma das questões de nível principal ou detalhado não fizesse sentido para a imagem, não seria necessário incorporá-la à descrição. Isso se tornou necessário após testes com algumas fotografias, onde percebeu-se que o modelo mencionava na descrição "*Não há elementos famosos na fotografia*", em resposta ao Item 4 de nível principal: "*Esta imagem retrata um sujeito famoso ou conhecido?*". Esse exemplo também foi incorporado ao texto das orientações. As orientações adaptadas estão disponíveis em: http://nau.uniriotec.br/index.php/orientacoes-nardi2021.

5. Avaliação por Pessoas com Deficiência Visual

A avaliação contou com a participação de duas pessoas cegas. Elas foram convidadas pessoalmente à participação e receberam uma versão em PDF acessível do TCLE na véspera da entrevista, via *WhatsApp*. Ambos os participantes indicaram ter lido o termo

e concordado em participar da pesquisa em mensagem de texto. Ainda assim, na ocasião da entrevista, o TCLE foi lido, com reforço do caráter voluntário da participação.

O roteiro da entrevista contou com algumas perguntas para identificação do perfil dos participantes e um conjunto de questões vinculadas às descrições das fotografias (o mesmo conjunto para cada foto). As primeiras quatro questões foram baseadas no trabalho de [Mohanbabu e Pavel 2024] e feitas somente para as descrições geradas com base nas orientações de [Nardi 2021]: tanto as simples quanto as detalhadas, consideradas no Estilo 1. Posteriormente foi feita uma pergunta comparando as descrições baseadas em [Nardi 2021] com as descrições geradas com base em instruções básicas ("*descreva a fotografia para uma pessoa com deficiência visual*"), consideradas no Estilo 2. A Figura 2 apresenta a ordem em que as descrições e perguntas foram feitas. As questões do roteiro vinculadas às descrições das fotografias foram as seguintes:

- Em uma escala de 1 a 5, onde 1 (um) representa "Muito ruim" e 5 (cinco) "Muito boa", como você avalia a **Descrição [simples | detalhada]** do **Estilo 1**? Por quê?
- Em uma escala de 1 a 5, em que 1 (um) significa "Com muita dificuldade" e 5 (cinco) "Com muita facilidade", como você avalia a capacidade de formar um modelo mental da fotografia da com base na **Descrição [simples | detalhada]** do **Estilo 1**? Por quê?
- Em uma escala de 1 a 5, onde 1 (um) significa "Capta muito pouco" e 5 (cinco) "Capta muito bem", como você avalia a capacidade da **Descrição [simples | detalhada]** do **Estilo 1** em captar os aspectos mais relevantes da fotografia? Por quê?
- Em uma escala de 1 a 5, em que 1 (um) significa "Pouca confiança" e 5 (cinco) "Muita confiança" o quanto você confia que a **Descrição [simples | detalhada]** do **Estilo 1** apresenta descrições corretas? Por quê?
- Qual dos estilos de descrição você preferiu? O **Estilo 1**, com subdivisão de descrições: simples e detalhada ou o **Estilo 2**? Por quê?
- Você tem algum comentário ou observação adicional sobre os estilos apresentados?

As descrições utilizadas podem ser consultadas nas Tabelas 1 (Estilo 1 - descrição simples), 2 (Estilo 1 - descrição detalhada) e 3 (Estilo 2). Durante a entrevista, cada descrição, conforme ordem do roteiro, foi enviada pelo *WhatsApp* aos participantes, que fizeram a leitura por meio do leitor de telas. Optou-se por esta abordagem para proporcionar a mesma experiência na interação com as descrições de imagem que eles têm no dia a dia, o que não aconteceria se a leitura fosse feita pela pesquisadora responsável pela condução da atividade.

As entrevistas aconteceram em dias diferentes, no primeiro semestre de 2025, com duração de cerca de 40 minutos cada. Para garantir a confidencialidade das pessoas voluntárias, elas serão apresentadas por P1 e P2.

A participante P1, mulher de 44 anos com ensino superior completo, que se declarou com cegueira total, adquirida na fase adulta e que, portanto, possui memória visual, lembrando-se perfeitamente de cores, objetos, formas, pessoas etc. Não fez cursos sobre audiodescrição de imagens, sendo apenas usuária frequente do serviço. Quando questionada sobre o uso de sistemas de IA para descrever imagens, ela informou que faz uso diário e de duas formas: via celular, a partir dos aplicativos *BeMyEyes*

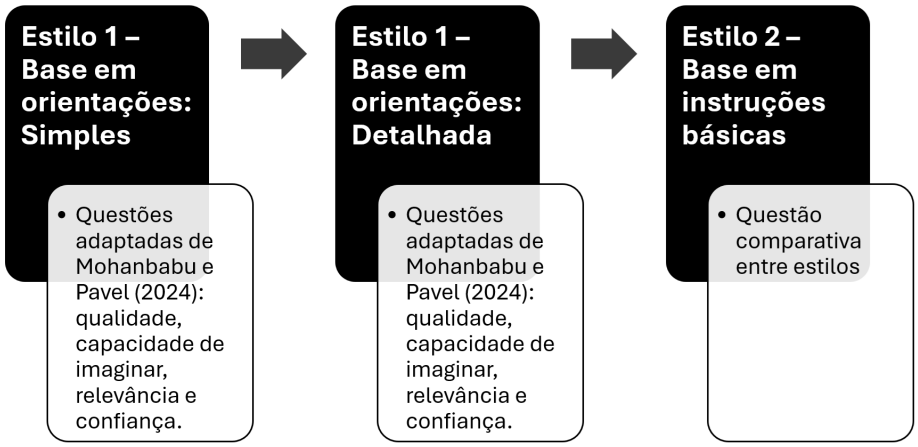

Figura 2. Ordem de leitura das descrições e perguntas durante avaliação. Crédito: dos autores.

(principalmente) e, mais recentemente, do *PiccyBot*, por este também permitir descrição de vídeos e via computador tradicional, onde tem o hábito de utilizar o *ChatGPT* ou o *Gemini*, este último incorporado ao navegador.

Já o participante P2, homem de 40 anos e ensino superior incompleto com cegueira adquirida na adolescência e memória visual preservada. É usuário de serviços de audiodescrição, sem nunca ter feito cursos para atuar como consultor. Relatou utilizar recursos baseados em IA para descrição de imagens e vídeos, prática que se intensificou e passou a fazer parte de sua rotina diária nos últimos dois anos. Indicou o *BeMyEyes* e o *PiccyBot* como ferramentas mais utilizadas para esta finalidade.

Antes do início da avaliação, os participantes foram informados que o primeiro estilo de descrição possuía dois níveis: descrição simples e descrição detalhada e que o segundo continha apenas uma breve descrição e lembrados que o contexto das descrições seria redes sociais. Além disso, explicou-se que no caso do primeiro estilo (baseado nas orientações de [Nardi 2021]), a descrição simples seria automaticamente lida ao interagir com a imagem e a descrição detalhada acionada pela pessoa usuária de leitor de telas no caso de interesse por mais detalhes. Esta abordagem considerou os detalhes progressivos apresentados no trabalho de [Morris et al. 2018], aprovados por pessoas com deficiência consultadas em trabalho anterior [Sacramento e Ferreira 2022].

5.1. Avaliação e comparação das descrições geradas com base em [Nardi 2021]

Para a pergunta em que eles deveriam avaliar a qualidade percebida da descrição simples, para todas as fotografias, os dois participantes indicaram nota máxima (cinco). Quando questionados sobre o conceito atribuído, ambos declararam que as descrições eram muito boas pela simplicidade e por conter elementos essenciais, apresentados de forma objetiva. Para a *selfie* (fotografia 2), o participante P2 comentou que a IA indicou a presença da pessoa e de detalhes básicos como cor dos objetos, roupa e onde a pessoa está, aspectos importantes para a descrição de uma *selfie*. A voluntária P1 limitou-se a dizer que estava

Tabela 1. Descrições simples do Estilo 1, utilizadas na avaliação

Fotografia	Descrições simples geradas por IA, a partir de [Nardi 2021]
Praia	Fotografia de uma praia ao entardecer. O principal foco da imagem está nas palmeiras altas à direita, com o mar calmo à esquerda e o sol se pondo no horizonte. A cena ocorre em um ambiente natural ao ar livre. As palmeiras estão inclinadas em direção ao mar, e a praia se encontra vazia e tranquila. O sol está prestes a desaparecer no horizonte, iluminando suavemente o céu com tons alaranjados e rosados.
Selfie	Fotografia de uma mulher jovem tirando uma selfie ao ar livre, próxima a um porto com diversos barcos atracados. Ela está sorrindo, usa um moletom azul claro e segura um celular lilás com a mão direita. Ao fundo, aparecem palmeiras, mastros de veleiros e o céu nublado, sugerindo um dia fresco.
Pessoa em lugar desconhecido	Fotografia de uma mulher caminhando sorridente em uma rua urbana de pedestres, cercada por prédios altos e elegantes. Ela usa jeans azul claro, blusa branca e uma camisa bege aberta, e segura um celular nas mãos. A cena acontece durante o dia, com diversas pessoas ao fundo passeando ou sentadas.

muito bom, pois era simples ou por ter conseguido entender a imagem.

Em relação à qualidade percebida das descrições detalhadas, com exceção da fotografia de uma pessoa em local desconhecido (fotografia 3), que recebeu nota três da participante P1, as demais imagens obtiveram nota máxima de ambos os voluntários. P1 justificou a avaliação intermediária pela grande quantidade de elementos presentes na descrição. Segundo ela, inicialmente conseguiu compreender a imagem, mas acabou perdendo o foco diante do excesso de informações. Embora o participante P2 tenha atribuído conceito máximo a todas as fotografias, comentou que as descrições detalhadas apresentaram um número elevado de detalhes.

Na pergunta seguinte, que abordou a facilidade da descrição simples em formar uma imagem mental da fotografia, a participante P1 atribuiu nota máxima a todas as imagens, indicando que a simplicidade das descrições permitiu imaginar os cenários sem dificuldades. Já o participante P2 concedeu nota máxima apenas à fotografia da praia (fotografia 1). Para a fotografia 2, atribuiu nota três justificando que a ausência de detalhes sobre a pessoa retratada dificultava a formação da imagem mental. Apesar de ter considerado a descrição muito boa no quesito anterior (qualidade percebida), o participante comentou que somente seria possível imaginar a pessoa caso a conhecesse previamente, uma vez que a descrição não forneceu características físicas da pessoa. Na foto da pessoa em local desconhecido (fotografia 3), P2 atribuiu nota quatro comentando que, por se tratar de um cenário cotidiano, foi possível imaginar o ambiente de maneira geral, ainda que a ausência de informações contextuais mais específicas tenha limitado a formação completa da cena em sua mente.

Quando questionados sobre a capacidade de formar um modelo mental das fotografias a partir da descrição detalhada, P1 atribuiu as notas 4, 4 e 2, enquanto P2 classificou em 4, 3 e 3 para as fotografias 1, 2 e 3, respectivamente. As principais queixas

Tabela 2. Descrições detalhadas do Estilo 1, utilizadas na avaliação

Fotografia	**Descrições detalhadas geradas por IA, a partir de [Nardi 2021]**
Praia	Fotografia de uma praia tropical durante o pôr do sol. O principal destaque da imagem são as palmeiras altas que se inclinam da direita para o centro da imagem, acompanhadas de arbustos verdes na parte inferior direita. À esquerda, o mar apresenta pequenas ondas que se quebram suavemente na areia clara da praia. Ao fundo, uma vegetação densa cobre uma pequena elevação de terra que avança sobre o mar. O céu, em degradê de tons quentes como laranja, rosa e dourado, apresenta nuvens esparsas e suaves, criando um clima sereno e relaxante. A imagem possui um estilo visual natural com iluminação quente e suave típica do final da tarde, transmitindo uma sensação de paz e tranquilidade. Não há elementos incomuns na cena, apenas a beleza simples e exuberante de uma paisagem litorânea tropical.
Selfie	Fotografia de uma mulher jovem com pele castanha e cabelo crespo, vestindo um moletom azul claro com capuz e uma alça azul no ombro esquerdo, tirando uma selfie com um celular lilás. Ela está posicionada levemente à esquerda da imagem, sorri de forma natural e parece estar em um momento descontraído. A cena acontece ao ar livre, em uma área portuária com vários barcos e iates atracados ao fundo, cercados por mastros altos. À esquerda e ao centro do fundo, há palmeiras, e o céu está coberto por nuvens, indicando um clima nublado, provavelmente durante o dia. O estilo visual da imagem é nítido e moderno, com cores naturais e boa iluminação, sem uso de filtros aparentes. A expressão da mulher transmite alegria e leveza, e o ambiente sugere uma atmosfera tranquila, comum em passeios próximos ao mar. Há carros estacionados mais ao fundo, reforçando o contexto urbano e turístico da paisagem.
Pessoa em desconhecido	Fotografia de uma mulher de cabelos longos e grisalhos, pele clara, que caminha com um sorriso leve pelo centro de uma rua de pedestres em uma área urbana de arquitetura clássica europeia. Ela veste uma calça jeans azul clara, blusa branca ajustada e uma camisa bege aberta, além de tênis brancos. A mulher está posicionada no centro da imagem e segura um celular com as duas mãos, olhando para a frente. A rua é pavimentada com pedras claras que refletem a luz do dia, sugerindo que a foto foi tirada pela manhã ou no início da tarde. Ao fundo, há prédios de vários andares com varandas, janelas com persianas e fachadas bem preservadas. Há também diversos postes pretos com floreiras penduradas. Muitas pessoas caminham, conversam ou estão sentadas ao longo da via, em clima de lazer e tranquilidade. A imagem é bem iluminada, com cores naturais e sem filtros, transmitindo uma sensação de leveza e bem-estar.

Tabela 3. Descrições do Estilo 2, utilizadas na avaliação

Fotografia	Descrições geradas por IA
Praia	Essa imagem mostra uma paisagem de praia ao entardecer, com o sol se pondo no horizonte sobre o mar calmo. O céu tem tons suaves de rosa e laranja, típicos do pôr do sol. Há diversas palmeiras altas e inclinadas à direita da imagem, com vegetação tropical densa ao fundo. A areia clara, pedras espalhadas na beira da água e o mar com pequenas ondas completam o cenário tranquilo e paradisíaco.
Selfie	Essa imagem mostra uma jovem sorridente tirando uma selfie com um celular roxo, provavelmente em um porto ou marina. Ela veste um moletom azul claro e carrega uma mochila no ombro. Ao fundo, vê-se palmeiras, diversos mastros de barcos atracados e um céu nublado, compondo um cenário urbano costeiro bastante agradável.
Pessoa em lugar desconhecido	Esta imagem mostra uma mulher sorridente caminhando por uma rua de pedestres em uma cidade com arquitetura europeia clássica. Ela veste uma blusa branca, camisa bege aberta, jeans flare e tênis brancos, e segura um celular. A rua está movimentada, com várias pessoas andando ou sentadas, e é decorada com postes de luz antigos e vasos de flores. As fachadas dos prédios têm janelas com venezianas e sacadas, típicas de centros históricos europeus.

dos voluntários referiram-se à presença excessiva de detalhes nas descrições. No caso da fotografia 3, que recebeu a avaliação mais baixa de P1, a participante esboçou a expressão "*Socorro!*" antes de justificar que a quantidade de informações a deixou confusa quanto à composição da fotografia. Ela comentou que não gosta de descrições muito detalhadas, especialmente no contexto das redes sociais. Para P1, é difícil conceber a interação com uma imagem no *Instagram* com tanto detalhamento e, mesmo sendo um recurso acionado sob demanda, provavelmente não o utilizaria.

O participante P2, por sua vez, comentou que precisaria de mais tempo e de múltiplas leituras para conseguir formar o modelo mental, dada a quantidade excessiva de informações. No entanto, destacou que gostaria que fosse implementado um recurso que permitisse esse nível de detalhamento, pois acredita que em uma foto de seu interesse poderia utilizá-lo.

Para a pergunta sobre a capacidade da descrição simples de capturar os aspectos mais relevantes da fotografia, os dois voluntários atribuíram nota máxima nas três imagens. A participante P1 comentou na fotografia 1 que, embora fosse difícil ter certeza que todos os elementos relevantes foram contemplados, a descrição apresentou componentes muito precisos em relação ao que ela espera de uma imagem de praia, especialmente por ser um local que ela gosta e frequenta muito.

Para a captura de aspectos relevantes na descrição detalhada, as fotografias 2 e 3 receberam nota quatro da participante P1, novamente devido à quantidade excessiva de informações. A voluntária comentou que há tantos detalhes que é questionável a relevância de todos eles, além de dificultar a percepção do que realmente é importante, pois o leitor acaba perdendo o foco. Já P2 atribuiu nota máxima na descrição de todas as

fotografias, argumentando que nesta abordagem todos os detalhes são relevantes e válidos.

Na pergunta sobre a confiança quanto à corretude da descrição, os voluntários tiveram posturas diferentes. A participante P1 atribuiu nota quatro a todas as descrições, mas argumentou que é algo difícil de avaliar e que tende a confiar parcialmente em descrições geradas por IA. Ela relatou que, quando precisa ter certeza sobre a corretude de uma descrição, costuma compará-la entre diferentes modelos de IA. Já P2 atribuiu nota cinco a todas as imagens e destacou que este é um ponto sensível. Ainda assim, P2 afirmou que costuma confiar na corretude da descrição até que alguém prove o contrário. Para ele, os modelos de IA estão cada vez mais evoluídos, errando menos nas descrições. Como estratégia de verificação, P2 compartilhou que, ao experimentar um novo modelo, costuma tirar uma foto, solicitar que a IA descreva, depois pede para uma pessoa vidente descrever a mesma imagem e compara os resultados obtidos.

Por fim, os participantes foram questionados sobre qual estilo de descrição preferiram. Em todas as imagens, a voluntária P1 preferiu o Estilo 2, mesmo quando comparado apenas à descrição simples do Estilo 1. Na fotografia 1, comentou que o Estilo 2 era ainda mais curto que a própria descrição simples, o que considerou um diferencial positivo. Na fotografia 2, afirmou ter gostado da descrição simples, mas ainda assim optaria pelo Estilo 2, pois trouxe mais detalhes apesar de ser uma descrição curta. Já na fotografia 3, reforçou sua preferência pelo Estilo 2 ao considerar o contexto das mídias sociais, e afirmou que, mesmo que houvesse a possibilidade de ativar uma descrição detalhada na interface de uma mídia social, provavelmente não o faria, dado seu gosto por descrições mais diretas e objetivas.

O participante P2 preferiu o Estilo 1 na fotografia 1 e, nas demais, o Estilo 2. Justificou sua escolha na primeira fotografia, afirmando que a descrição simples fazia mais sentido, estava mais organizada e poderia ser complementada, caso necessário, pela descrição detalhada. Já em relação às demais fotografias, argumentou que o Estilo 2, mesmo curto, trazia detalhes que não estavam presentes na descrição simples do Estilo 1. Como exemplo, na fotografia 3, comentou que o Estilo 2 mencionava elementos como "*arquitetura europeia clássica*" e "*centro histórico europeu*" que, no Estilo 1, só poderiam ser acessados via descrição detalhada. Em um comentário bem-humorado, apelidou o Estilo 2 da fotografia 3 como "*descrição simples bombada*", por conta do acréscimo de informações contextuais ausentes na descrição mais simples do Estilo 1.

5.2. Discussões

Apesar do número reduzido de voluntários, os resultados da avaliação permitiram um refinamento inicial da proposta e o direcionamento de trabalhos futuros sobre o uso de orientações estruturadas para instruir modelos de IA na tarefa de descrever imagens. Os principais pontos de atenção identificados estão listados a seguir:

Interesse por descrições objetivas e curtas: uma das observações mais significativas foi a preferência por descrições simples e objetivas no contexto das mídias sociais, o que já havia sido apontado em estudo anterior com pessoas cegas congênitas [Sacramento e Ferreira 2022] e agora reforçado por participantes com cegueira adquirida. Os participantes preferiram o Estilo 2 em praticamente todas as imagens, por este ter combinado objetividade com a apresentação de aspectos relevantes. Embora o Estilo 1 não tenha sido o preferido dos participantes na maioria das fotografias analisadas,

a qualidade percebida das descrições simples foi avaliada de forma muito positiva, justamente por sua objetividade e por incluírem os elementos considerados essenciais para a compreensão da imagem.

Ausência de elementos relevantes na descrição simples: apesar de a descrição simples do Estilo 1 ter sido bem avaliada quanto à sua objetividade, ela foi alvo de críticas por omitir informações contextuais consideradas relevantes, como características físicas de pessoas ou o estilo arquitetônico presente nas imagens. O Estilo 2 foi preferido por apresentar esses detalhes adicionais de forma objetiva, o que sugere que o Estilo 1 pode estar omitindo elementos visuais relevantes para a compreensão da imagem.

Preferência pelo que é familiar, em detrimento da técnica: Outro achado relevante foi a preferência dos participantes pelo Estilo 2, mesmo que o Estilo 1 tenha sido elaborado com base em questões estruturadas e em algumas diretrizes da audiodescrição. Essa preferência pode ter influência do hábito que os voluntários possuem de interagir com descrições *default* feitas pelo modelo de IA estudado, o que pode ter contribuído para uma maior familiaridade e aceitação desse formato. Embora não seja possível identificar com precisão os parâmetros utilizados pelo modelo de IA na geração das descrições a partir de instruções básicas (Estilo 2), um aspecto deste estilo que poderia ser questionado por especialistas em audiodescrição é a ausência de uma identificação explícita do tipo da imagem. Por exemplo, todas as descrições iniciam com "*Esta imagem mostra*", sem indicar que se trata de uma fotografia ou uma ilustração. A ausência de críticas a detalhes técnicos da audiodescrição aconteceu devido ao fato de os participantes não possuírem formação específica ou atuação como consultores, sendo essencial o envolvimento de especialistas em novas avaliações para uma análise mais técnica das descrições geradas.

Excesso e ausência de detalhes nas descrições: Apesar de as descrições detalhadas do Estilo 1 também terem sido consideradas, na maioria dos casos, como muito boas em relação à qualidade percebida, a quantidade excessiva de informações impactou negativamente na formação de um modelo mental das fotografias pelos dois participantes. Esse achado aponta para a necessidade de investigar com mais profundidade qual é o nível adequado de detalhamento a ser adotado nesse tipo de contexto. Mesmo no caso das descrições simples do Estilo 1, essa investigação permanece pertinente, já que duas das justificativas para a escolha do Estilo 2 foram justamente o fato dele apresentar mais detalhes relevantes, como características físicas da pessoa ou o estilo arquitetônico do local retratado, não incorporados na descrição simples do Estilo 1.

Necessidade de adequação a um contexto real de uso: as críticas em relação ao detalhamento das fotografias podem ter sido influenciadas pelo fato das imagens não serem de interesse direto dos participantes, além das fotografias terem sido apresentadas em um contexto artificial de leitura. Os resultados poderiam ser diferentes caso fosse investigada uma situação real de uso. Por exemplo, se houvesse uma interação espontânea com uma imagem de interesse dos voluntários, a partir de recurso efetivamente implementado em uma rede social, em que os participantes pudessem acessar descrições simples por padrão e, se desejassem, acionar versões mais detalhadas. A exploração dessa interação em trabalhos futuros pode trazer novas conclusões sobre a pertinência de detalhamentos no contexto das mídias sociais. O contexto artificial da atividade também pode ter dificultado a percepção de relevância dos conteúdos descritos, uma vez que os participantes não estavam inseridos em uma situação em que a compreensão da imagem

tivesse um propósito claro. Investigações futuras em cenários mais realistas, com imagens significativas para os voluntários podem contribuir para uma avaliação mais precisa da percepção de relevância do conteúdo.

Corretude, confiança e validação das descrições: Apesar de ambos os participantes terem avaliado positivamente todas as descrições em relação à confiança, as percepções individuais revelaram abordagens distintas. Enquanto uma voluntária demonstrou maior cautela e tendência a conferir as descrições por diferentes modelos, o outro relatou confiar nas descrições até que se prove o contrário. Essa variação de comportamento reforça a importância de oferecer transparência sobre a origem das descrições, tal como apontado em [MacLeod et al. 2017], que menciona a necessidade de identificar a autoria da descrição, além da necessidade de disponibilizar formas de validação da descrição, para fortalecer a confiança do usuário.

Embora os participantes não tenham apontado inconsistências e erros nas descrições, até pela dificuldade imposta pela deficiência, este é um aspecto relevante a ser explorado em trabalhos futuros. Por exemplo, na fotografia 2 (*selfie*) é dito no Estilo 1 que o celular retratado é lilás e no Estilo 2, que é roxo. Outros exemplos, na mesma imagem, ilustram erros de descrição: no Estilo 1 (descrição simples) é dito que a mulher "*segura um celular lilás com a mão direita*", porém o celular está na mão esquerda e, no Estilo 2 consta que ela "*carrega uma mochila no ombro*", contudo nenhuma mochila é exibida na imagem. O erro na identificação do lado do corpo pode ter sido causado por uma das dicas adicionais, que orientava o uso do ponto de vista do observador como referência direcional. Embora esta seja uma diretriz da audiodescrição, quando se trata do corpo de uma pessoa ou de um animal, deve-se indicar o lado real do sujeito retratado (por exemplo, "*mão esquerda*" ou "*pata direita*"), independentemente da posição em que apareçam na imagem em relação ao observador [Matsushita 2019].

Cabe ressaltar que a verificação da corretude das informações não foi objeto de investigação nesta pesquisa, mas trata-se de um aspecto essencial a ser considerado em estudos futuros, uma vez que impacta diretamente na credibilidade e na confiança atribuída às descrições geradas por inteligência artificial.

6. Orientações Adaptadas para Descrições de Imagem com Inteligência Artificial

Ainda que as orientações propostas não tenham sido a escolha dos participantes na maioria das imagens analisadas, os achados da pesquisa permitiram um refinamento inicial das orientações, caracterizado pelos seguintes ajustes:

- Indicação explícita, nas orientações iniciais, que a descrição simples deve ser objetiva e curta;
- Incorporação das características dos sujeitos reportados na imagem na descrição simples; ção da dica sobre posicionamento: ao descrever partes do corpo de alguém, deve-se indicar o lado real correspondente (por exemplo, "*mão esquerda*" ou "*perna direita*"), independentemente da posição em que apareçam na imagem em relação ao observador.

A seguir são apresentadas as orientações adaptadas após estudo inicial com envolvimento de pessoas cegas adquiridas e com memória visual.

#Orientações

1. Descreva a imagem para o contexto das redes sociais
2. Faça duas descrições para a imagem. A primeira deve ser objetiva e curta e você vai chamar de "Descrição Simples"usando apenas as questões do nível principal como referência (de 1 a 8, na seção "Nível principal"), e a segunda mais detalhada, que você vai chamar de "Descrição Detalhada", contemplando todas as questões: de 1 a 12 (das seções "Nível principal"e "Nível detalhado".
3. É essencial utilizar as orientações da seção "Dicas adicionais"para produzir as descrições.
4. Apresente as descrições em texto corrido.

#Nível principal

1. Qual é o tipo de imagem que está sendo descrita?
   A primeira coisa a ser incluída na descrição é o tipo de imagem que ela aborda. Exemplo: Fotografia.
2. Quem são os principais sujeitos da imagem?
   Qual o foco principal da imagem? Indique quem ou o que está sendo descrito (pessoas, animais, objetos, paisagem etc.). Exemplos: Fotografia de uma mulher negra, de um grupo de crianças, de um cachorro, de uma praia, de uma bicicleta.
3. Onde está este conjunto?
   Indique a localização do elemento que está sendo descrito, caso seja relevante ou seja possível identificar. Exemplos: em um restaurante, ao ar livre, em uma escola.
4. Esta imagem retrata um sujeito famoso ou conhecido?
   Caso positivo, indique a pessoa/animal/local/objeto conhecido na descrição. Exemplo: Fotografia da Artista X, da praia de Copacabana, da medalha de ouro das Olimpíadas 2016.
5. Como os sujeitos estão se portando na imagem?
   Indique a posição, expressões e outras informações relacionadas ao comportamento do sujeito. Exemplo: está sorrindo, chorando, está sentado, do lado esquerdo, pisca os olhos.
6. O que os sujeitos da imagem estão fazendo?
   Indique as ações e intenções dos sujeitos, caso possível ou se aplique à imagem. Exemplo: jogando futebol, bebendo cerveja, usando um computador, rezando.
7. Em que momento a imagem foi registrada?
   Se possível, indique informações que possam contextualizar o momento em que a imagem foi registrada. Exemplo: manhã, tarde, noite, momento específico.
8. Quais são as características físicas do sujeito?
   Detalhe as características físicas do sujeito. Exemplo: características notáveis, como cor dos olhos, cabelo, roupas (entre outros, para descrição de pessoas), itens da natureza/arquitetura (e posições relativas), composição, formatos, dimensões etc.

#Nível detalhado

9. Quais são as características do plano de fundo?
   Detalhe as características do plano de fundo, os itens (e posições relativas), indicando se existem pessoas e, caso necessário, condições climáticas (faz sol, está chovendo).

10. Existem aspectos notáveis no estilo visual da imagem?
    Caso a imagem possua um estilo visual que fuja do padrão, indique-o. Exemplo: se está em preto-e-branco, em sépia, se possui algum tipo de filtro, técnica de fotografia etc.
11. Que emoção esta imagem evoca?
    Se possível e pertinente, indique as emoções dos sujeitos presentes na imagem.
12. Você está retratando um objeto/elemento incomum?
    Caso esteja, indique sua utilidade e informações sobre como utilizá-lo. Se possível, compare-o com um objeto/elemento similar.

#Dicas adicionais

1. Escreva no tempo verbal presente e em linguagem simples e objetiva;
2. Evite regionalismos e gírias;
3. Mencione as cores
4. Procure descrever os elementos de cima para baixo, da esquerda para a direita, exceto se o foco estiver em outro local;
5. Ao indicar posicionamento (esquerda, direita), o ponto de vista deve sempre ser o de quem observa a imagem, no caso o seu. No entanto, ao descrever partes do corpo de uma pessoa ou animal, deve-se indicar o lado real correspondente (por exemplo, “mão esquerda” ou “pata direita”), independentemente da posição em que apareçam na imagem
6. Evite indicar termos como: “a imagem mostra”, “olhando para a câmera”;
7. Descreva apenas o que consta na imagem. Exemplos: Se a imagem somente retrata paisagem, evite indicar a ausência de pessoas.
8. Caso a resposta à uma das questões nível principal ou detalhado não faça sentido para a imagem, não é necessário incorporá-la à descrição. Por exemplo: Se a imagem não retrata uma paisagem ou pessoa famosa, não mencione na descrição "Não há elementos famosos"ou similar

7. Conclusão

Neste artigo, apresentou-se a avaliação inicial de um conjunto de orientações para a descrição de fotografias no contexto de redes sociais, desenvolvidas em trabalho anterior e adaptadas para uso com inteligência artificial. As orientações estruturadas foram avaliadas por duas pessoas com deficiência visual total e memória visual preservada, que opinaram sobre aspectos como qualidade percebida, facilidade para formar um modelo mental das imagens, relevância dos elementos descritos e confiança nas descrições. Adicionalmente, os voluntários compararam as descrições geradas a partir das orientações adaptadas, chamadas de Estilo 1 (que contempla descrições simples e detalhadas), com descrições produzidas a partir de uma instrução simples, o Estilo 2.

Nesta investigação a preferência dos participantes deu-se majoritariamente pelas descrições do Estilo 2. Alguns achados de estudos anteriores foram reforçados, como o interesse por descrições objetivas e breves no contexto de mídias sociais e a tendência a confiar em descrições geradas por inteligência artificial. Novas perspectivas também emergiram, como a dificuldade de compreensão da imagem em descrições com muitos detalhes, a preferência por formato de descrição familiar, em detrimento de descrições mais alinhadas tecnicamente e a importância da criação de estratégias e soluções que permitam a validação das descrições em relação à corretude.

Com base nos resultados, foram feitos ajustes nas descrições, como a inserção de características da pessoa/ambiente retratado na descrição simples, o reforço quanto a necessidade de objetividade na descrição simples e a correção de instruções relacionadas à indicação espacial, como a identificação correta do lado do corpo retratado, independentemente da posição em que aparece na imagem. Percebeu-se, ainda, a importância de validar as orientações propostas em contextos reais de uso, a fim de obter resultados mais aderentes às demandas práticas dos usuários.

As limitações apresentadas na Subseção 3.2 são reconhecidas e há intenção de atenuá-las em trabalhos futuros. Para tal, pretende-se ampliar a avaliação, aplicando as orientações em contextos mais próximos da experiência cotidiana - como a escolha de imagens de interesse das pessoas voluntárias, e com um número maior de participantes com deficiência visual, incluindo outros perfis, como pessoas com cegueira congênita.

Outras vertentes de investigação incluem a participação de especialistas em audiodescrição (roteiristas videntes e consultores cegos) na validação das orientações e o uso das orientações em diferentes modelos de inteligência artificial generativa, em um estudo comparativo com base nos critérios de avaliação adotados nesta pesquisa (preferência, capacidade de formar modelo mental da fotografia, captação de aspectos relevantes e confiança), além da investigação de aspectos relacionados à corretude das descrições.

8. Agradecimentos

Agradecemos às pessoas com deficiência que participaram da pesquisa e às pessoas revisoras, pelas considerações e sugestões para melhoria do trabalho. Além disso, informamos o uso do modelo de inteligência artificial generativa *ChatGPT 4 plus* na revisão ortográfica e gramatical de parte do texto, na descrição das imagens e na tradução do resumo para o inglês. A tradução também contou com o uso do *Google Tradutor* para revisão do inglês.

Referências

ABNT (2016). Abnt nbr 16452: Acessibilidade na comunicação – audiodescrição. Disponível em: https://www.abntcatalogo.com.br/norma.aspx?ID=359735. Acesso em: 08 ago. 2025.

Amiralian, M. L. T. (1997). *Compreendendo o cego: uma visão psicanalítica da cegueira por meio de desenhos-estórias*. Casa do Psicólogo, São Paulo. Disponível em: http://www.deficienciavisual.pt/txt-compreendendo-cego.htm. Acesso em: 20 ago. 2025.

Barbosa, S. D. J., da Silva, B. S., Silveira, M., Gasparini, I., Darin, T., e Barbosa, G. (2021). *Interação humano-computador e experiência do usuário [E-book]*. Leanpub. Disponível em: https://leanpub.com/ihc-ux. Acesso em: 08 ago. 2025.

Berton, E., Molina, L., Júnior, O., e Santana, W. (2024). Ia avança nas descrições de imagens, mas ainda não substitui a revisão humana. Movimento Web para Todos. Disponível em: https://mwpt.com.br/ia-avanca-nas-descricoes-de-imagens-mas-ainda-nao-substitui-a-revisao-humana. Acesso em: 08 ago. 2025.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., e et al. (2022). On the opportunities and risks of foundation models. Disponível em: https://arxiv.org/abs/2108.07258. Acesso em 08 ago. 2025.

Clause, A.-A. (2025). Open ai nvda add-on. Github. Disponível em: https://github.com/aaclause/nvda-OpenAI/. Acesso em: 08 ago. 2025.

Duarte, E. F., Porto, G. L. P. M. B., Nascimento, A., Palomino, P. T., dos Santos Portela, C., Aguiar, Y. P. C., Falcão, T. P., Ribeiro, D. F., Souza, M., Gasparotto, A. M. S., e Toda, A. M. (2024). Grandihc-br 2025-2035 - gc6: Implications of artificial intelligence in hci: A discussion on paradigms ethics and diversity equity and inclusion. In *IHC 24: Proceedings of the XXIII Brazilian Symposium on Human Factors in Computing Systems*, pages 1–19. Association for Computing Machinery. Disponível em: https://dl.acm.org/doi/10.1145/3702038.3702059. Acesso em: 08 ago. 2025.

Ferreira, S. B. L., Chauvel, M. A., e do Amaral Ferreira, M. G. (2006). e-acessibilidade: tornando visível o invisível. In *30º Encontro da ANPAD*. Disponível em: http://nau.uniriotec.br/index.php/publicacoes/artigos-de-congressos-ou-conferencias/98-e-acessibilidade-tornando-visivel-o-invisivel. Acesso em: 08 ago. 2025.

Furuya, B. (2024). Como a ia já está ajudando pessoas com deficiência. Olhar Digital. Disponível em: https://olhardigital.com.br/2024/07/09/pro/como-a-ia-ja-esta-ajudando-pessoas-com-deficiencia/. Acesso em: 08 ago. 2025.

Gasparetto, M. E. R. F. (2007). A pessoa com visão subnormal e seu processo pedagógico. In Masini, E. F. S. e Gasparetto, M. E. R. F., editors, *Visão subnormal: um enfoque educacional*, chapter 2. Vetor, São Paulo, 1 edition.

Gleason, C., Pavel, A., McCamey, E., Low, C., Carrington, P., Kitani, K. M., e Bigham, J. P. (2020). Twitter a11y: A browser extension to make twitter images accessible. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 1–12. Association for Computing Machinery. Disponível em: https://dl.acm.org/doi/10.1145/3313831.3376728. Acesso em: 08 ago. 2025.

IBGE (2024). Pesquisa nacional por amostra de domicílios contínua: Acesso à internet e à televisão e posse de telefone móvel celular para uso pessoal 2023. Disponível em: https://biblioteca.ibge.gov.br/visualizacao/livros/liv102107_informativo.pdf. Acesso em: 08 ago. 2025.

IFPB (2018). Cegueira x baixa visão. Instituto Federal da Paraiba. Disponível em: https://www.ifpb.edu.br/assuntos/fique-por-dentro/cegueira-x-baixa-visao. Acesso em: 08 ago. 2025.

Jacques, E. G., Sacramento, C., Gouveia, Y., Silva, W. N., Barros, Y. S., e Ferreira, S. B. L. (2025). Preservação da memória com acessibilidade digital: um plugin para descrição de imagens com ia generativa. In *Anais Estendidos do Simpósio Brasileiro de Sistemas de Informação (SBSI). Trilha Indústria e Inovação em Sistemas de Informação*, pages 157–161. SBC. Disponível em: https://sol.sbc.org.br/index.php/sbsi_estendido/article/view/34596. Acesso em: 08 ago. 2025.

Jandrey, A. H., Ruiz, D. D. A., e Silveira, M. S. (2021). Image descriptions' limitations for people with visual impairments: Where are we and where are we going? In *IHC 21: Proceedings of the XX Brazilian Symposium on Human Factors in Computing Systems*. Association for Computing Machinery. Disponível em: https://dl.acm.org/doi/10.1145/3472301.3484356. Acesso em: 08 ago. 2025.

Kaplan, A. M. e Haenlein, M. (2010). Users of the world, unite! the challenges and opportunities of social media. *Business Horizons*, 53:59–68. Disponível em: https://www.sciencedirect.com/science/article/abs/pii/S0007681309001232. Acesso em: 08 ago. 2025.

Mack, K., Cutrell, E., Lee, B., e Morris, M. R. (2021). Designing tools for high-quality alt text authoring. In *ASSETS 2021 - 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. Association for Computing Machinery, Inc. Disponível em: https://dl.acm.org/doi/10.1145/3441852.3471207. Acesso em: 08 ago. 2025.

MacLeod, H., Bennett, C. L., Morris, M. R., e Cutrell, E. (2017). Understanding blind people's experiences with computer-generated captions of social media images. In *Conference on Human Factors in Computing Systems - Proceedings*, volume 2017-May, pages 5988–5999. Association for Computing Machinery. Disponível em: https://doi.org/10.1145/3025453.3025814. Acesso em: 08 ago. 2025.

Matsushita, R. (2019). Curso de introdução à audiodescrição: diretrizes gerais de ad. All Dubbing Group, Rio de Janeiro.

Mohanbabu, A. G. e Pavel, A. (2024). Context-aware image descriptions for web accessibility. In *ASSETS 2024 - Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility*, volume 17. Association for Computing Machinery, Inc. Disponível em: https://dl.acm.org/doi/pdf/10.1145/3663548.3675658. Acesso em: 08 ago. 2025.

Moraes, C. P. (2018). Cego também usa facebook: #pracegover. Monografia (Bacharel em Publicidade e Propaganda). Universidade de Passo Fundo, Passo Fundo, RS. Disponível em: http://repositorio.upf.br/handle/riupf/1505. Acesso em: 08 ago. 2025.

Morash, V. S., Siu, Y. T., Miele, J. A., Hasty, L., e Landau, S. (2015). Guiding novice web workers in making image descriptions using templates. In *ACM Transactions on Accessible Computing (TACCESS)*, volume 7. ACMPUB27New York, NY, USA. Disponível em: https://dl.acm.org/doi/pdf/10.1145/2764916. Acesso em: 08 ago. 2025.

Morris, M. R., Johnson, J., Bennett, C. L., e Cutrell, E. (2018). Rich representations of visual content for screen reader users. In *Conference on Human Factors in Computing Systems - Proceedings*, volume 2018-April. Association for Computing Machinery. Disponível em: https://dl.acm.org/doi/10.1145/3173574.3173633. Acesso em: 08 ago. 2025.

Nardi, C. C. D. S. (2021). Diretrizes para produção de alternativas ao conteúdo visual em mídias sociais online sob a perspectiva de pessoas com deficiência visual. Tese (Doutorado em Informática). Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, RJ. Disponível em: http://nau.uniriotec.br/index.php/orientacoes/doutorado/307-diretrizes-para-producao-

de-alternativas-ao-conteudo-visual-em-midias-sociais-online-sob-a-perspectiva-de-pessoas-com-deficiencia-visual. Acesso em: 08 ago. 2025.

Nunes, S. D. S. e Lomônaco, J. F. B. (2008). Desenvolvimento de conceitos em cegos congênitos: caminhos de aquisição do conhecimento. *Psicologia Escolar e Educacional*, 12:119–138. Disponível em: https://www.scielo.br/j/pee/a/zvVp8FNBfyxH9b3FwJYskPx/?lang=pt. Acesso em: 08 ago. 2025.

OpenAI (2025). Key guidelines for writing instructions for custom gpts | openai help center. Disponível em: https://help.openai.com/en/articles/9358033-key-guidelines-for-writing-instructions-for-custom-gpts. Acesso em: 08 ago. 2025.

Pedrosa, L. (2015). Inclusão: quais são as redes sociais mais populares entre deficientes visuais? Portal EBC. Disponível em: https://memoria.ebc.com.br/cidadania/2015/05/pedagoga-cega-analisa-melhor-rede-social-na-opiniao-de-pessoas-com-deficiencia. Acesso em: 08 ago. 2025.

Perdigão, L. T., Monteiro, F. V., Peixotto, B. J., Bianco, V. L., e Fernandes, E. M. (2023). Inteligência artificial para audiodescrição de imagens: uma análise da pessoa com deficiência visual. In *Congresso sobre Tecnologias na Educação (Ctrl+E)*, pages 182–191. SBC. Disponível em: https://sol.sbc.org.br/index.php/ctrle/article/view/25797. Acesso em: 08 ago. 2025.

Sacramento, C. e Ferreira, S. B. L. (2022). Accessibility on social media: exploring congenital blind people's interaction with visual content. In *IHC 22: Proceedings of the 21st Brazilian Symposium on Human Factors in Computing Systems*. Association for Computing Machinery. Disponível em: https://dl.acm.org/doi/10.1145/3554364.3559140. Acesso em: 08 ago. 2025.

Sacramento, C., Ferreira, S. B. L., e Remedios, S. (2022). Um estudo sobre descrição de imagens em mídias sociais online na perspectiva de pessoas com cegueira congênita. In *Anais do XIII Workshop sobre Aspectos da Interação Humano-Computador na Web Social (WAIHCWS)*, pages 63–70. SBC. Disponível em: https://sol.sbc.org.br/index.php/waihcws/article/view/22577. Acesso em: 08 ago. 2025.

Sacramento, C., Nardi, L., Ferreira, S. B. L., e Marques, J. M. D. S. (2020). Pracegover: Investigating the description of visual content in brazilian online social media. In *IHC 2020 - Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems*. Association for Computing Machinery, Inc. Disponível em: https://dl.acm.org/doi/10.1145/3424953.3426489. Acesso em: 08 ago. 2025.

Salisbury, E., Kamar, E., e Morris, M. R. (2017). Toward scalable social alt text: Conversational crowdsourcing as a tool for refining vision-to-language technology for the blind. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 5, pages 147–156. AAAI Press. Disponível em: https://ojs.aaai.org/index.php/HCOMP/article/view/13301. Acesso em: 08 ago. 2025.

Singh, N., Wang, L. L., e Bragg, J. (2024). Figura11y: Ai assistance for writing scientific alt text. In *IUI '24: Proceedings of the 29th International Conference on Intelligent*

User Interfaces, pages 886–906. Association for Computing Machinery. Disponível em: https://dl.acm.org/doi/pdf/10.1145/3640543.3645212. Acesso em: 08 ago. 2025.

WebAIM (2025). Webaim: The webaim million - the 2025 report on the accessibility of the top 1,000,000 home pages. Disponível em: https://webaim.org/projects/million/wcag. Acesso em: 08 ago. 2025.

Wu, S. e Adamic, L. (2014). Visually impaired users on an online social network. In *Conference on Human Factors in Computing Systems - Proceedings*, pages 3133–3142. Association for Computing Machinery. Disponível em: https://dl.acm.org/doi/10.1145/2556288.2557415. Acesso em: 08 ago. 2025.

Wu, S., Wieland, J., Farivar, O., e Schiller, J. (2017). Automatic alt-text: Computer-generated image descriptions for blind users on a social network service. *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, pages 1180–1192. Disponível em: https://doi.org/10.1145/2998181.2998364. Acesso em: 08 ago. 2025.