

Inteligência artificial para audiodescrição de imagens: uma análise da pessoa com deficiência visual

Luciana Tavares Perdigão¹, Felipe Vieira Monteiro², Bruno José Peixoto³, Vittorio Lo Bianco³, Ediclea Mascarenhas Fernandes²

¹Universidade Federal Fluminense (UFF) - Niterói - RJ

²Universidade do estado do Rio de Janeiro (UERJ) - Rio de Janeiro - RJ

³Fundação Cecierj - Rio de Janeiro - RJ

Resumo: *A audiodescrição é uma tecnologia assistiva que permite acessibilidade aos conteúdos visuais através da tradução verbal. O presente estudo analisa o uso da inteligência artificial para audiodescrição de imagens em ambientes virtuais da educação a distância. Utilizou como instrumentos o Google Lens e o ChatGPT. Das oito imagens selecionadas, metade não foi reconhecida pelas IAs; a outra metade passou pela revisão de um consultor com deficiência visual e os textos foram classificados em uma escala entre bom e muito ruim. O resultado aponta para uma ineficiência da inteligência artificial para audiodescrição de imagens da EAD.*

Abstract: *Audio description is an assistive technology that allows accessibility to visual content through verbal translation. The present study analyzes the use of artificial intelligence for audio description of images in virtual environments of distance education. Google Lens and ChatGPT were used as instruments. Of the eight selected images, half were not recognized by the AIs; the other half was reviewed by a visually impaired consultant and the texts were rated on a scale between good and very bad. The result points to an inefficiency of artificial intelligence for audio description of distance learning images.*

1. Introdução

Imagens são mais do que meros adornos visuais na educação, são ferramentas poderosas que podem transformar o aprendizado em uma experiência mais reflexiva e criativa. No contexto da educação a distância (EAD) elas ganham maior protagonismo ao serem exploradas nos ambientes virtuais de aprendizagem (AVA). Todavia, estudantes com perfis de aprendizagens diversos se sentem excluídos por falta de recursos que promovam o acesso a esse tipo de conteúdo. Nesse contexto a audiodescrição surge como uma tecnologia assistiva que permite o acesso ao conteúdo visual através da descrição verbal. Além disso, a acessibilidade das imagens por meio da audiodescrição beneficia não apenas os alunos com deficiência visual, mas toda a comunidade educacional. De acordo com Motta (2016, p. 119) a audiodescrição aumenta a fluência verbal, o senso de observação, o repertório cultural, o acervo de palavras. Ao adotar a audiodescrição didática, os educadores promovem uma cultura de inclusão e diversidade, construindo ambientes educacionais mais acessíveis e acolhedores para todos os alunos.

No entanto, a produção de um roteiro de audiodescrição requer formação especializada e habilidade dos profissionais envolvidos. Isso limita a disponibilidade de conteúdos audiodescritos e torna a técnica inacessível para alguns educadores, pela falta de tempo, por exemplo. Embora o uso de IAs generativas venham sendo promovidas por Big Techs como soluções tecnológicas que podem trazer maior produtividade, estudos que avaliem sua confiabilidade e eficiência para produzir recursos de acessibilidade como a audiodescrição ainda são escassos. Em um levantamento no Google Acadêmico utilizando as palavras-chave "inteligência artificial", audiodescrição, "tecnologias assistivas" e "deficiência visual" foram retornados 78 resultados. Desconsiderando os estudos duplicados e aqueles fora do escopo deste levantamento a partir da filtragem, nessa ordem, dos títulos; resumos; objetivos; resultados e considerações finais; foram identificados quatro estudos relevantes. Entretanto, nenhum deles envolveu a análise das mesmas IAs e a participação de um consultor em audiodescrição, o que caracteriza o ineditismo desta pesquisa.

O presente estudo tem como objetivo 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 do audiodescritor consultor com deficiência visual.

2. Revisão de Literatura

2.1. Inteligência artificial

A inteligência artificial (IA) é uma área da ciência da computação que busca desenvolver sistemas capazes de realizar tarefas onde, tradicionalmente, domina a inteligência humana. Segundo Teixeira (2019), a expressão “inteligência artificial” pode soar de maneira assustadora pelo fato de que uma máquina pode executar tarefas que até então eram atreladas aos seres humanos, pensando e agindo racionalmente, isto é, executando tarefas para as quais se supõe a necessidade da inteligência. A IA está associada à linguagem, raciocínio, aprendizagem e resolução de problemas, que por sua vez perpassam os vários domínios das ciências, desde a linguística e a psicologia até a filosofia e a epistemologia (DE CASTRO BARBOSA, 2020). Segundo David Carmona (2019), as IAs são projetadas para aprender, raciocinar, tomar decisões e solucionar problemas de forma autônoma, baseando-se em algoritmos complexos e análise de dados. Um algoritmo é uma sequência de passos ou instruções lógicas bem definidas que descrevem como resolver um problema ou realizar uma tarefa. Carmona (2019) afirma que a IA tem o potencial de transformar significativamente diversos setores, entre eles a educação, favorecendo o ambiente para a inovação. Dentre a diversidade de aplicações possíveis com o uso da Inteligência artificial, está o reconhecimento visual e a descrição de imagens.

A descrição automática de imagens ajuda todos os usuários a acessar conteúdo importante nelas, de uma foto que surge como resultado de uma busca a uma imagem inclusa em uma apresentação. O uso de legendas para descrição de uma imagem é especialmente importante para portadores [sic] de deficiência visual. (...) Sistemas de descrição de imagens geralmente são treinados com conjuntos de dados que contém imagens combinadas com sentenças que as descrevem. (RIGUES, 2020).

De acordo com Nabeel (2019), uma "tecnologia de reconhecimento facial" consiste em um software programado para identificar e reconhecer rostos humanos

específicos a partir de fotos ou vídeos. Por meio do uso de vastas bases de dados e conexões de internet de alta velocidade, as tecnologias de reconhecimento facial são capazes de identificar e catalogar os detalhes de cada indivíduo, a fim de processar imagens capturadas por computadores, smartphones ou câmeras de vigilância. Os dados processados podem ser empregados para uma ampla gama de propósitos. Em termos gerais, os sistemas de reconhecimento facial operam por meio da biometria, mapeando as características faciais de uma pessoa em uma fotografia ou vídeo e comparando essas informações com um banco de dados de rostos conhecidos, a fim de encontrar uma correspondência. Embora as técnicas utilizadas possam variar, os sistemas de reconhecimento facial seguem, em geral, etapas comuns, como descrito por Weschler (2007).

O Google Lens é um exemplo de ferramenta de reconhecimento visual que utiliza a tecnologia de inteligência artificial para analisar e identificar imagens capturadas pela câmera do smartphone ou tablet. O aplicativo utiliza algoritmos de processamento de imagem para identificar pessoas, objetos, locais, códigos de barras, textos e outros elementos presentes nas imagens. De acordo com Donatti (2022) a IA do Google Lens faz uma busca no banco de dados do Google, comparando com a imagem capturada pelo usuário e buscando similaridade, relevância, além dos metadados como títulos, legendas e descrições prévias para determinar os resultados mais prováveis. Embora a tecnologia possa usar modelos de linguagem em alguns de seus recursos, o Google Lens é uma combinação de técnicas de visão computacional e processamento de imagem.

Um modelo de linguagem é uma distribuição de probabilidade sobre sequências de palavras (DIGIAMPIETRI, 2022) e o ChatGPT é um exemplo. De acordo com a desenvolvedora OpenAI, foi projetado com base na arquitetura GPT-3.5 (Generative Pre-trained Transformer) “para conversar com pessoas e responder a uma ampla variedade de perguntas e tópicos” (OPENAI, 2023). Digiampietri (2022) esclarece que a arquitetura GPT-3.5 é um modelo de linguagem baseada em uma rede neural de transformadores pré-treinada em grandes quantidades de dados textuais. O autor define que Processamento de Língua Natural (PLN) é um campo interdisciplinar que envolve a inteligência artificial e linguística com o foco na interação entre computadores e línguas humanas naturais, resultando no processamento ou produção de dados em diversos formatos como textos, áudios, vídeos etc.

Como modelo de linguagem, o ChatGPT não possui a capacidade de visualizar ou analisar imagens diretamente. Ele é alimentado com uma grande quantidade de texto e treinado para gerar respostas e fornecer informações com base nessa entrada textual. Portanto, o ChatGPT só descreve imagens a partir de uma URL - Uniform Resource Locator (localizador uniforme de recursos) - ou seja, um endereço textual onde essa imagem está publicada.

2.2. Audiodescrição didática

Descrever é o ato de expressar verbalmente ou por escrito as características, aparência, funcionamento, contexto ou qualquer outra informação relevante sobre um objeto, pessoa, lugar, evento ou qualquer elemento que se queira comunicar. Diferentemente do ato natural de descrever, a audiodescrição é uma técnica que tem como objetivo proporcionar uma imagem clara e precisa do objeto em questão, permitindo que o receptor da descrição visualize, compreenda e forme uma ideia sobre aquilo que está

sendo descrito. De acordo com Motta e Romeu Filho (2010, p. 07) é uma modalidade de tradução intersemiótica, que transforma o visual em verbal, abrindo possibilidades de acesso à informação, contribuindo para a inclusão cultural, social e escolar.

Na educação a distância, a audiodescrição deve ser utilizada nas imagens dos materiais didáticos, das videoaulas, de recursos multimídia e no AVA. Trata-se da audiodescrição didática que tem como objetivo oferecer ao aluno cego o mesmo conhecimento oferecido ao aluno que enxerga (VERGARA-NUNES, 2016, p. 163). Apesar da audiodescrição ter o público de pessoas com deficiência visual como prioritário, atualmente entende-se que esse recursos de acessibilidade, como a audiodescrição, não atende um público específico, e sim estão disponíveis para todos. Por ser um segundo canal sensorial a ser utilizado na compreensão das informações visuais (QUEIROZ, 2010), já é comprovado que a audiodescrição didática beneficia também os alunos com deficiência intelectual, síndrome de Down, idosos, disléxicos e outros transtornos de aprendizagem. Vergara-Nunes (2016) aponta ainda que, no caso de materiais didáticos, cada imagem tem um objetivo de aprendizagem, e isto deve ser considerado para definir o tipo de audiodescrição a ser adotado. "A áudio-descrição não é determinada pela disponibilidade imediata do áudio, mas é construída na mediação da informação oral descritiva dos eventos visuais, estada na provisão de acessibilidade comunicacional que leve ao empoderamento das pessoas usuárias desse serviço" (LIMA, 2011).

A partir do entendimento das possibilidades de uso da inteligência artificial como instrumento para descrição de imagens utilizadas na educação a distância, o presente estudo apresenta a seguir o procedimento metodológico de uso do Google Lens e do ChatGPT.

3. Metodologia

Para realização do estudo foram utilizadas as imagens de um material disponibilizado na disciplina Artes Visuais na Educação do curso de Licenciatura em Pedagogia - Unirio Cederj. O critério de seleção foi por ser uma disciplina predominantemente visual que recebeu a inscrição de uma aluna com deficiência visual em 2023.2 onde ela encontrou barreiras de acessibilidade ao conteúdo. O arquivo no formato .pdf contém 8 imagens em preto e branco, baixa resolução e sem legendas, o que justificou o uso do Google Lens como instrumento para identificação da imagem.

Como o arquivo estava disponibilizado em um AVA de acesso restrito, não foi possível utilizar o Google Lens a partir do navegador Google Chrome, sendo necessário o uso da câmera de um smartphone para esse reconhecimento visual do app.

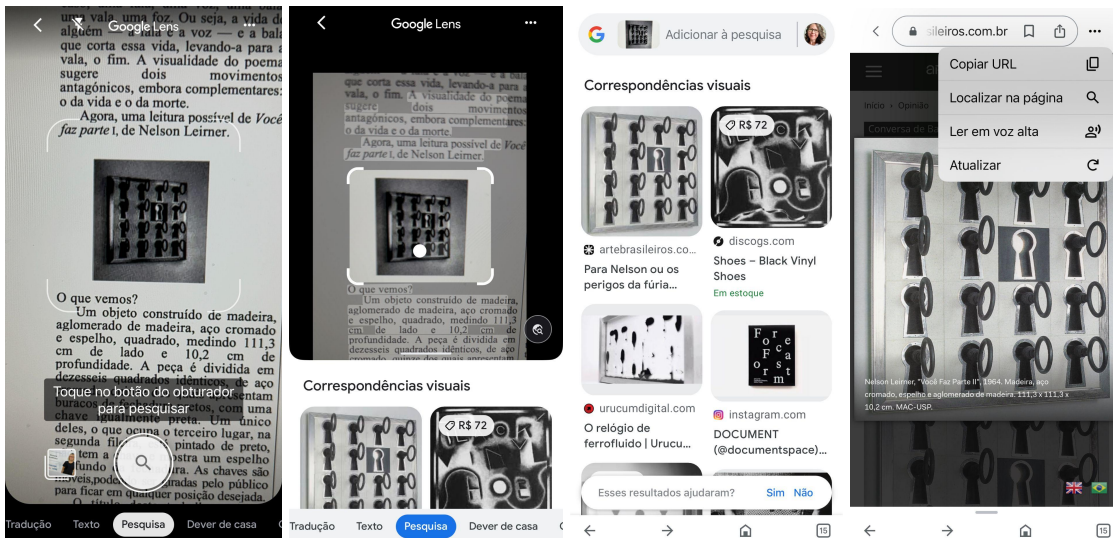


Figura 1 - Sequência de capturas do uso do Google Lens para identificação da imagem

A partir desta identificação o endereço da imagem identificada foi copiado para o ChatGPT¹ com a seguinte seguinte instrução: "descreva esta imagem: [link]."

🗨️ descreva esta imagem:
<https://www.baslibrary.org/sites/default/files/bsbr0601039001.jpg>

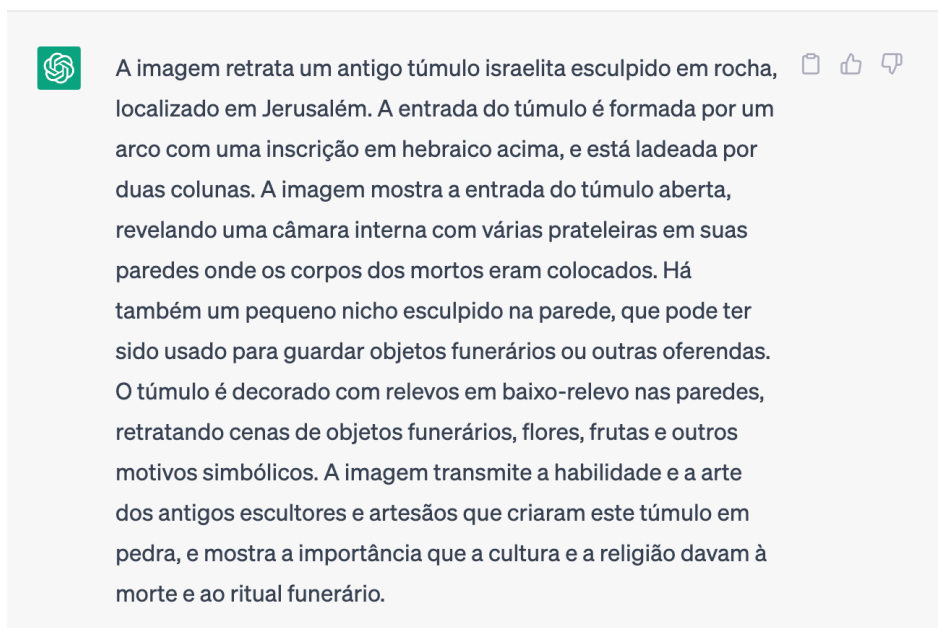


Figura 2 - Tela do ChatGPT com a instrução para descrição da imagem a partir do link.

Apesar de algumas imagens exploradas no documento serem mais conhecidas, a descrição a partir do link ao invés da instrução, por exemplo, "descreva o Discóbolo de Miron", foi uma opção para que a inteligência artificial identificasse acertadamente que se trata de uma fotografia da escultura do Discóbolo, como está no documento, e não da escultura propriamente dita. As descrições elaboradas pela IA foram encaminhadas para análise de um consultor com deficiência visual que revisa e valida roteiros de audiodescrição.




¹ O site de acesso ao ChatGPT é o <https://chat.openai.com/>

4. Resultados

Das oito imagens contidas no documento, duas não foram identificadas pelo Google Lens e outras duas tiveram descrições feitas pelo ChatGPT sem qualquer correspondência com a imagem original. As demais imagens que tiveram descrições que condizem com a imagem receberam a consultoria e uma classificação, conforme detalhado no quadro a seguir.

Quadro 1. Relação das imagens e o retorno das descrições pesquisadas.

IMAGEM	DESCRIÇÃO	AVALIAÇÃO
 <p>universo de significações possíveis de um obra</p> <p>x: intencionalidade do at</p> <p>y, h, w, n etc.: significados que podemos atribuir à obra, sem desrespeitar sua proposta</p> <p>o: significado arbitrário; que não pertence ao universo obra e que não podemos impor a ela</p>	<p>A imagem foi elaborada pelo autor do texto e não foi identificada pelo Google Lens, portanto não houve descrição a ser encaminhada para a consultoria.</p>	<p>Qualidade do roteiro da IMAGEM 1:</p> <p>SEM DESCRIÇÃO</p>
<p>uma vez uma fala uma foz uma vez uma bala uma fala uma voz uma foz uma vala uma bala uma vez uma voz uma vala uma vez'</p>	<p>A imagem retrata uma cena noturna em uma cidade, com uma rua iluminada por uma série de lâmpadas penduradas em postes de luz. A rua é estreita e pavimentada com paralelepípedos, e os prédios ao longo dela são altos e estreitos, com janelas iluminadas que parecem sugerir que há pessoas morando ali. O céu está escuro e estrelado, com uma lua brilhando acima dos prédios...</p>	<p>Qualidade do roteiro da IMAGEM 2:</p> <p>NÃO AVALIADA</p>
	<p>A imagem é uma pintura a óleo em tela com dimensões verticais. Ela retrata uma paisagem natural, com montanhas ao fundo, céu azul claro com nuvens brancas e um rio serpenteando pelo primeiro plano da imagem. A cena é iluminada por uma luz intensa que cria sombras longas e contrastes marcantes entre as cores. A vegetação no primeiro plano é composta principalmente de árvores altas e frondosas com folhas verdes e brilhantes, enquanto as montanhas ao fundo são mais escarpadas e escuras, sugerindo uma maior distância. A composição geral da imagem é harmônica e equilibrada, com a linha do rio funcionando como um elemento que guia o olhar do observador através da imagem.</p>	<p>Qualidade do roteiro da IMAGEM 3:</p> <p>NÃO AVALIADA</p>
	<p>A imagem é uma fotografia em preto e branco de uma escultura antiga intitulada "Discóbolo" ou "O Lutador com Disco", criada pelo escultor grego Mirón por volta de 450 aC. A escultura retrata um atleta masculino em pleno movimento de lançamento do disco, com o corpo tenso e musculoso em uma pose atlética. O atleta está nu, com a exceção de um pequeno pano enrolado em sua perna esquerda. A escultura é vista em um ângulo lateral, com o atleta virado para a direita e a perna esquerda puxada para trás, dando a sensação de movimento e dinamismo. A imagem transmite a força, beleza e habilidade atlética da figura representada.</p>	<p>Qualidade do roteiro da IMAGEM 4:</p> <p>BOM</p>
	<p>A imagem mostra um mosaico localizado na parede de uma basílica cristã bizantina, intitulada Basílica de São Vital, em Ravenna, Itália. O mosaico é composto por pequenos pedaços de vidro colorido, que foram cuidadosamente arranjados para criar uma imagem maior. A cena retratada é a figura do imperador bizantino Justiniano I, com um halo dourado em volta de sua cabeça, em pé no centro do mosaico, usando uma túnica roxa e segurando um cálice de ouro nas mãos. À</p>	<p>Qualidade do roteiro da IMAGEM 5:</p> <p>BOM</p>

	<p>esquerda do imperador está um grupo de clérigos e à direita está um grupo de militares. Ao fundo, há um edifício com uma cúpula e torres que representam a cidade de Ravena, enquanto no chão há um tapete luxuoso. A imagem transmite uma sensação de majestade e poder imperial, com a figura de Justiniano I destacada em meio a um corte de seguidores.</p>	
	<p>A imagem mostra um baixo-relevo esculpido em pedra, que representa uma cena bíblica do profeta Isaías. O baixo-relevo está localizado na fachada de um edifício em Paris, conhecido como Cité de l'architecture et du patrimoine. Na cena, o profeta Isaías é retratado de pé, com a mão direita levantada e segurando um rolo de pergaminho, enquanto sua mão esquerda segura um bastão. A imagem é altamente detalhada, mostrando como roupas e os adornos do profeta, bem como a textura do pergaminho. A cena é emoldurada por uma moldura decorativa, com várias figuras esculpidas em relevo ao redor da cena principal, incluindo anjos e animais. A imagem é uma obra de arte excepcionalmente bem executada e demonstra a habilidade e a técnica dos escultores que a criaram.</p>	<p>Qualidade do roteiro da IMAGEM 6: MÉDIO</p>
	<p>A imagem retrata uma folha de papel com várias cabeças humanas desenhadas à mão. Os desenhos foram feitos com lápis em tons de cinza e apresentam uma grande variedade de expressões e ângulos de cabeça. As cabeças retratadas parecem ser estudos anatômicos, possivelmente feitas por um artista como uma preparação para pinturas ou esculturas. Os desenhos foram executados com grande habilidade, mostrando a destreza do artista em capturar a forma humana e a emoção por trás de cada expressão. A imagem é um exemplo da habilidade do renomado artista e inventor Leonardo da Vinci em desenhar e estudar a anatomia humana em detalhes precisos.</p>	<p>Qualidade do roteiro da IMAGEM 7: MÉDIO PARA RUIM</p>
	<p>A imagem não foi identificada pelos pesquisadores nem pelo Google Lens, portanto não houve descrição a ser encaminhada para a consultoria.</p>	<p>Qualidade do roteiro da IMAGEM 8: SEM DESCRIÇÃO</p>

Como modelo de linguagem, o ChatGPT não visualiza imagens. Ele é alimentado com uma grande quantidade de dados descritivos e treinado para gerar respostas, fornecendo informações com base nessa entrada textual. Por isso que na IMAGEM 1 não foi possível alcançar nem uma descrição inicial, pois é uma ilustração que está disponível apenas no documento de acesso restrito e não compartilhado publicamente na web. Já a IMAGEM 2 está disponível em diversos documentos compartilhados na web, principalmente em apostilas online sobre poesia concreta. Entretanto, a IA ainda não tem dados suficientes para identificá-la, fornecendo uma descrição aleatória de outra imagem. O mesmo aconteceu com a IMAGEM 3: foram inseridos dois comandos de descrever a imagem a partir de endereços distintos de hospedagem, além de mais seis opções de “Regenerar resposta”, que seguiram apresentando descrições completamente diferentes da imagem original.

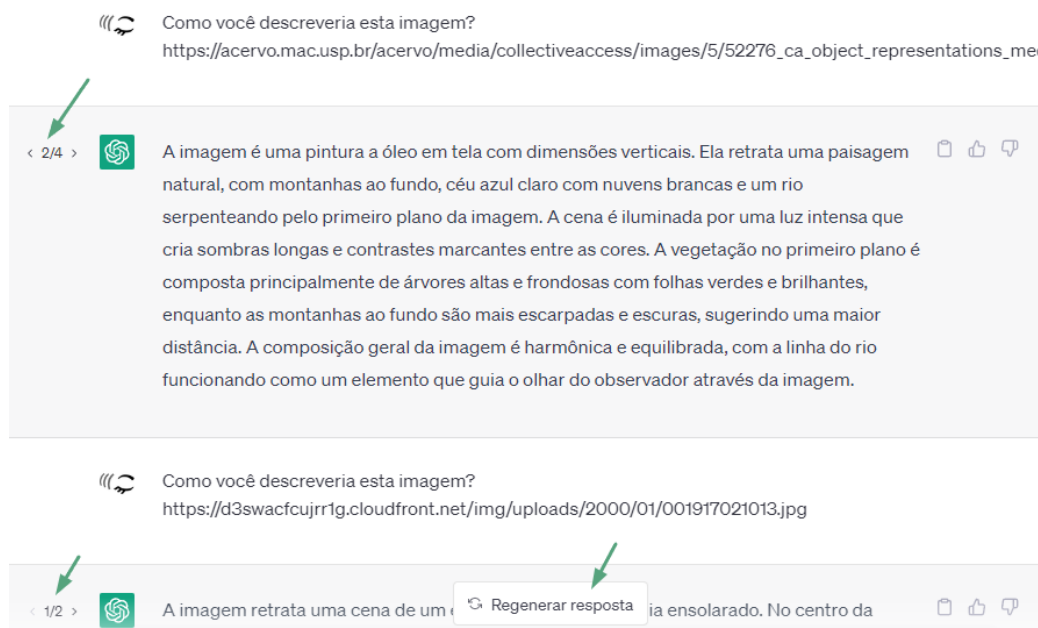


Figura 3. Tela do ChatGPT com duas instruções e indicação de seis respostas regeneradas.

Todas as respostas foram de imagens aleatórias, desde o "Abaporu" de Tarsila do Amaral, até pinturas a óleo de paisagens, retratos, etc. Ainda na IMAGEM 3, por ser uma obra conhecida, foi feito um terceiro experimento inserido o comando "Descreva a imagem: 'Você faz parte I', de Nelson Leirner". O ChatGPT retornou: "Peço desculpas, mas como sou um modelo de linguagem de texto, não tenho capacidade de visualizar imagens ou acessar conteúdo específico da internet. Portanto, não posso descrever a imagem 'Você faz parte I' de Nelson Leirner com precisão. No entanto, posso fornecer informações gerais sobre Nelson Leirner e seu estilo artístico, se desejar."

Nas imagens 4, 5, 6 e 7, as descrições geradas pelo ChatGPT iniciam com "a imagem mostra" ou "imagem de" ou "a imagem é"... Trata-se de um erro básico recorrente de descritores sem experiência em audiodescrição: se estamos tratando da descrição de uma imagem, não é necessário introduzir com essa informação, basta descrever o tipo e o formato da imagem.

A descrição da IMAGEM 4 foi classificada como BOA pelo consultor, mas com algumas considerações. Apesar de identificar que se trata de uma fotografia em preto e branco, a descrição não contempla a cor da escultura, que pode ser caracterizada com as tonalidades da escala de cinza. Também não foram descritas as características físicas do homem e do objeto, bem como o seu posicionamento. As descrições "EM PLENO MOVIMENTO DE LANÇAMENTO DE DISCO" e "POSE ATLÉTICA" não são específicas o suficiente para compreensão da ação por uma pessoa com cegueira congênita, por exemplo. É necessário descrever o posicionamento das pernas, braços, dorso e cabeça para dar clareza do movimento. Existe também um problema de ordenamento da descrição, que, após as notas proêmias², informa sobre o movimento, depois as características físicas, depois volta para o movimento e retoma para as características físicas.

² Notas proêmias ou notas introdutórias são informações iniciais e complementares sobre a obra a ser descrita, para maior fruição do usuário (TAVARES, 2013).

A descrição da IMAGEM 5 também foi considerada boa pelo consultor. Mas apresenta erros de redundância (basílica; mosaico; imperador) e inferência (cuidadosamente; para criar uma imagem maior; destacada) que afetam na concisão do roteiro. O texto oferece detalhes que deveriam estar nas notas introdutórias (A imagem transmite uma sensação de majestade e poder imperial...) e carece de descrição das características físicas dos personagens.

O consultor avaliou a descrição da IMAGEM 6 como média, pois apresenta muitos léxicos desnecessários, tornando a descrição longa e deixando de priorizar as características físicas do elemento principal da imagem, conforme apresentado no quadro 1:

Quadro 2. Apontamentos do consultor para a descrição da IMAGEM 5.

A imagem mostra um baixo-relevo esculpido em pedra, que representa uma cena bíblica do profeta Isaías.
@Desnecessário A IMAGEM MOSTRA UM. @Citar o tipo de imagem. @Citar a orientação da imagem.

O baixo-relevo está localizado na fachada de um edifício em Paris, conhecido como Cité de l'architecture et du patrimoine.

@Desnecessário O BAIXO RELEVO. @Desnecessário CONHECIDO COMO.

Na cena, o profeta Isaías é retratado de pé, com a mão direita levantada e segurando um rolo de pergaminho,

@Desnecessário NA CENA. @Descrever as CARACTERÍSTICAS FÍSICAS. @Descrever as VESTIMENTAS.

enquanto sua mão esquerda segura um bastão.

@Fazer aposto para PERGAMINHO. @Colocar um PONTO após o aposto de PERGAMINHO.

@Citar a cor do pergaminho. @Desnecessário ENQUANTO. @Desnecessário SUA. @Citar a cor do BASTÃO.

A imagem é altamente detalhada, mostrando como roupas e os adornos do profeta, bem como a textura do pergaminho.

@Desnecessário A IMAGEM É BASTANTE DETALHADA MOSTRANDO COMO.

A cena é emoldurada por uma moldura decorativa, com várias figuras esculpidas em relevo ao redor da cena principal, incluindo anjos e animais.

@Desnecessário A CENA. @Desnecessário EMOLDURADA. @Desnecessário CENA PRINCIPAL. Tem outra cena?

A imagem é uma obra de arte excepcionalmente bem executada e demonstra a habilidade e a técnica dos escultores que a criaram.

@Desnecessária essa última frase, pois é uma inferência. @O roteiro está médio.

A pior avaliação foi para a descrição da IMAGEM 7, pois, para o consultor, possui muitas inferências e poucas informações sobre as características físicas das cabeças.

5. Considerações Finais

O presente estudo apontou que O ChatGPT não pode ser considerado como uma ferramenta para audiodescrição de imagens. Das oito imagens levadas para o experimento, somente quatro receberam descrição, sendo todas classificadas de nível médio para ruim. É necessário que a audiodescrição como técnica seja mais amplamente disponibilizada nos conteúdos visuais compartilhados na web para que as bases de dados das IAs sejam mais robustas.

Uma possibilidade é o uso da inteligência artificial como instrumento para auxílio da elaboração de roteiros; ainda assim é necessário fazer um estudo com os audiodescritores roteiristas para analisar essa possível utilização. Outra perspectiva de

estudo é fazer testes de treinamento de máquina através do feedback para os resultados oferecidos pela IA. E mais estudos são necessários para aprofundamento da temática.

REFERÊNCIAS

- CARMONA, D. (2019) "The AI Organization". O'Reilly Media, Inc. Ebook Kindle.
- DE CASTRO BARBOSA, X. (2020) Breve introdução à história da Inteligência Artificial. Jamaxi, v. 4, n. 1.
- DIGIAMPIETRI, L.A. (2023) "O que é o ChatGPT?" EACH-USP.
http://www.each.usp.br/digiampietri/SIN5028/06_ChatGPT.pdf
- DONATTI, F. (2022) "Ensino de física e experimentos didáticos: explorando o aplicativo Google Lens como ferramenta de identificação de imagens". Universidade Estadual Paulista (Unesp). <http://hdl.handle.net/11449/242732>.
- LIMA, F. J. (2011) "Introdução aos estudos do roteiro para áudio-descrição: sugestões para a construção de um script anotado". Revista Brasileira de Tradução Visual, v. 7.
- MOTTA, L.M.V. e ROMEU FILHO, P. (2010) "Audiodescrição: Transformando Imagens em Palavras". Secretaria dos Direitos da Pessoa com Deficiência de SP.
- MOTTA, L.M.V. (2016) "A audiodescrição na escola: abrindo caminhos para leitura de mundo". Campinas, SP: Pontes Editores.
- NABEEL, F. (2019) "Regulating facial recognition technology in public places". Centre for Strategic and Contemporary Research.
https://www.academia.edu/39871139/Regulating_Facial_Recognition_Technology_in_Public_Places.
- OPENAI. (2023) "Apresentando o ChatGPT", 2023. <https://openai.com/blog/chatgpt>
- QUEIROZ, M.A. (2010) "Prefácio". In: MOTTA, L. M. V. .M. e FILHO, P. R. Audiodescrição: Transformando Imagens em Palavras. São Paulo, 13-20.
- RIGUES, R. (2020) "Nova IA da Microsoft descreve imagens tão bem quanto os humanos". Disponível em:
<https://olhardigital.com.br/2020/10/14/noticias/nova-ia-da-microsoft-descreve-imagens-tao-bem-quanto-os-humanos/>
- TEIXEIRA, J. (2019) "O que é inteligência artificial". E-galáxia. Disponível em:
<https://repositorio.ufsc.br/bitstream/handle/praxis/395/o%252%20que%20e%20inteligencia%20artificial.pdf?sequence=1> .
- TAVARES, L. B. (2013) "Notas Proêmias: Acessibilidade Comunicacional para Produções Culturais". Pernambuco: DVD Acessível, 2013.
https://edutec.unesp.br/images/stories/rededor2-ee-ei/1ed-ee-ei/Ebook/Notas_Proemias/index_textos.html.
- WECHSLER, H (2007). "Reliable face recognition methods: system design, implementation and evaluation". Springer.