Revisão

EDIÇÃO ESPECIAL
EAD COM QUALIDADE, INCLUSÃO E ACESSIBILIDADE

EaD em foco
Revista Científica em Educação a Distância

ISSN 2177-8310
DOI: 10.18264/eadf.v15i2.2593

Revisão Sistemática de Literatura sobre Uso da Inteligência Artificial para Audiodescrição de Imagens da EaD

Systematic Literature Review on the Use of Artificial Intelligence for Audio Description of EaD images

Luciana PERDIGÃO*[1]
Sergio PINTO[1]
Caio Dias FERREIRA[1]

[1] Universidade Federal Fluminense – Niterói BRASIL

*lucianaperdigao@id.uff.br

Resumo

Com os avanços recentes das tecnologias digitais, incorporar a inteligência artificial à educação a distância já se apresenta como uma possibilidade muito mais viável e acessível. O uso da Inteligência Artificial (IA) para elaboração de audiodescrição das imagens estáticas pode ser um potencial instrumento de acessibilidade para a EaD. O objetivo desta Revisão Sistemática de Literatura – RSL – é identificar o estado da arte sobre o uso da inteligência artificial para audiodescrição de imagens estáticas. A execução da RSL foi apoiada com a ferramenta Parsifal, e, através do protocolo especificado, foram encontrados 62 artigos. Foram analisados o título, resumo e palavras-chave e, por meio dos critérios de inclusão, exclusão e de qualidade, foram aceitos 22 estudos. Após a leitura completa dos estudos aceitos, foram elencados 86 algoritmos e ferramentas de IA utilizadas, além das 69 bases de dados e outras 9 ferramentas úteis no desenvolvimento de instrumentos para descrição de imagens. Esses resultados apontam para potenciais possibilidades de utilização de instrumentos prontos para desenvolvimento de uma ferramenta para audiodescrição de imagens estáticas para EaD.

Palavras-chave: Inteligência artificial. Audiodescrição. Educação a distância.

Recebido 23/05/2025
Aceito 24/06/2025
Publicado 16/07/2025

Editores responsáveis:
Márcia Denise Pletsch
Andrea Velloso
Klaus Schlunzen Junior

COMO CITAR ESTE TRABALHO

ABNT: PERDIGÃO, L.; PINTO, S.; FERREIRA . C. D. Revisão Sistemática de Literatura sobre Uso da Inteligência Artificial para Audiodescrição de Imagens da EaD. **EaD em Foco**, v. 15, n. 2, e2593, 2025.doi: https://doi.org/10.18264/eadf.v15i2.2593

Systematic Literature Review on the Use of Artificial Intelligence for Audio Description of EaD images

Abstract

With recent advances in digital technologies, incorporating artificial intelligence into distance education now presents itself as a much more viable and accessible possibility. The use of Artificial Intelligence (AI) to prepare audio description of static images can be a potential accessibility instrument for EaD. The objective of this Systematic Literature Review – RSL – is to identify the state of the art on the use of artificial intelligence for audio description of static images. The execution of the RSL was supported with the Parsifal tool, and, through the specified protocol, 62 articles were found. The title, abstract and keywords were analyzed and, using the inclusion, exclusion and quality criteria, 22 studies were accepted. After fully reading the accepted studies, 86 algorithms and AI tools used were listed, in addition to 69 databases and 9 other tools useful in the development of instruments for describing images. These results point to potential possibilities for using ready-made instruments to develop a tool for audio description of static images for distance learning.

***Keywords**: Artificial intelligence. Audio description. Distance education.*

1. Introdução

O ano de 2023 marcou um momento significativo para a Inteligência Artificial (IA), quando ela se tornou acessível e utilizada pelo grande público em atividades profissionais, pessoais e acadêmicas. Embora o termo "Inteligência Artificial" tenha sido cunhado na década de 1950 por John McCarthy (Dartmouth College, 1956), quando surgiram as primeiras respostas para o aprendizado de máquina, o ChatGPT, desenvolvido pela OpenAI, foi o pioneiro em trazer a IA para o centro das atenções. Com essa ferramenta, as pessoas puderam fazer perguntas e se surpreender com as respostas geradas por essa IA.

O salto significativo na IA deve-se ao surgimento da inteligência artificial generativa. Diferentemente da IA preditiva, que analisa dados existentes, a IA generativa cria conteúdos novos e originais. Essa tecnologia emergiu com potencial para inovações em áreas como entretenimento, comunicação e educação. Além dos *chatbots*, um programa de computador que simula uma conversa humana, seja por texto seja por voz, para interagir com usuários, e geradores de imagens, a nova geração de ferramentas baseadas em IA generativa inclui criação de vídeos, músicas e síntese de voz, tornando-a acessível e democrática para todos.

Dentre os recursos de acessibilidade que têm potencial para serem desenvolvidos pela IA, um deles é a descrição de imagens. A descrição de imagens é uma das Diretrizes de Acessibilidade de Conteúdo da *Web* – WCAG – desenvolvida pelo Consórcio *World Wide Web* –W3C (*World Wide Web Consortium*, 2018). O documento apresenta essa recomendação como "texto alternativo para conteúdo não textual" e aponta para a necessidade de descrição dos dados representados em gráficos, diagramas e ilustrações.

A maioria das ferramentas *online* e redes sociais já disponibilizam um campo para inserção de texto alternativo atribuído às imagens (Perdigão, Monteiro, Fernandes, 2021). Entretanto, assim como nos requisitos da WCAG, as orientações para preenchimento desse campo não se relacionam com as diretrizes da audiodescrição (Perdigão, Lima, 2017). A audiodescrição é uma técnica de tradução das imagens em palavras de forma clara, coesa, concisa, específica, vívida e ordenada (Lima e Tavares, 2010 e Perdigão,

2023). É uma tecnologia assistiva que busca principalmente a inclusão e o empoderamento da pessoa com deficiência visual; além disso, este recurso pode ampliar as possibilidades de inserção social e acesso à informação/comunicação às pessoas com deficiência intelectual, disléxicos e idosos em diversos contextos sociais (Lima e Tavares, 2010). Essa tecnologia assistiva pressupõe a necessidade de validação por um consultor com deficiência visual. Experimentos iniciais de uso da Inteligência Artificial para audiodescrição de imagens foram realizados em estudo anterior (Perdigão, 2023) cujo resultado apontou para uma ineficiência do ChatGPT e do Google Lens (um aplicativo de reconhecimento de imagens desenvolvido pelo Google), pela falta de qualidade da descrição das imagens utilizadas. Desde então, novas ferramentas foram introduzidas no mercado, apontando para novas possibilidades.

Para aprofundar os conhecimentos sobre o que vem sendo pesquisado em relação ao uso da inteligência artificial generativa para audiodescrição de imagens, foi realizada uma Revisão Sistemática de Literatura – RSL. Trata-se de um estudo com o objetivo de identificar, analisar e interpretar as evidências, em fontes primárias e secundárias, acerca da temática (Kitchenham, Charters, 2007). Foi utilizado o Parsifal, uma ferramenta *online* que auxilia no planejamento e condução da revisão seguindo protocolos pré-definidos, com a possibilidade de geração de relatórios.

2. Metodologia

O Parsifal é uma ferramenta que auxilia na organização da RSL através de uma interface subdividida em Revisão, Planejamento, Condução e Relatórios. A aba Revisão é destinada para inserir as informações básicas da RSL, como título, descrição e autores. A seguir serão detalhados o Planejamento e Condução da RSL.

2.1. Planejamento

A aba Planejamento é organizada em três sessões: Protocolo, *Checklist* de análise de qualidade e Formulário de extração de dados. O protocolo é plano que descreve a condução de uma proposta de RSL, conforme detalhado no quadro a seguir:

Quadro 1: Protocolo da Revisão Sistemática de Literatura.

PROTOCOLO DA RSL
Objetivos:
Avaliar estudos anteriores sobre o uso da Inteligência Artificial para Audiodescrição de Imagens
PICOC
População: Imagens estáticas utilizadas na EaD **Intervenção:** Usar a Inteligência Artificial para descrição das imagens **Comparação:** Algoritmos, ferramentas e bases de dados utilizadas **Outcome (Resultado):** Lista de instrumentos úteis para criação de ferramenta de IA **Contexto:** Conteúdos didáticos da plataforma Moodle Cederj
Questões de pesquisa:
Os artigos relacionam a IA com descrição de imagens estáticas? Os artigos relacionam a IA para descrição de imagens e acessibilidade? Os artigos relacionam a IA com descrição de imagens no contexto didático? Os artigos apresentam a descrição de imagens com IA seguindo as diretrizes da audiodescrição? Quais ferramentas foram abordadas nos trabalhos selecionados?

Palavras-chave	Sinônimos	Relacionado a
Imagens estáticas	Fotografia, Gráfico, Ilustração	População
Audiodescrição	Descrição de imagens, Texto alternativo	Intervenção
Inteligência Artificial	LLM, Machine Learning	Intervenção
Instrumentos de IA	Algoritmos, Bases de dados, Ferramentas	Comparação
Lista de instrumentos	Acervo	Resultado
String de busca		
("descrição de imagens" OR audiodescrição OR audiodescription OR "audio description" OR "image description" OR "image caption" OR "visual caption") AND ("inteligência artificial" OR "artificial intelligence" OR "learning machine" OR "deep learning" OR "language model")		
Fontes		
IEEExplore: https://ieeexplore-ieee-org.ez24.periodicos.capes.gov.br/ Springer: https://link-springer-com.ez24.periodicos.capes.gov.br/ Google scholar: https://scholar.google.com		
Critérios de seleção		
Critérios de inclusão palavras da string e sinônimos no resumo relaciona a IA com descrição de imagens e acessibilidade relata uso de IA na descrição de imagens	Critérios de exclusão artigo duplicado documentos protegidos não está em inglês, português ou espanhol não é artigo	

Para a análise de qualidade dos artigos pré-selecionados, foi estabelecido um *checklist* de questões:

- O artigo relaciona a IA com descrição de imagens estáticas?
- O artigo relaciona a IA à descrição de imagens e acessibilidade?
- O artigo apresenta a descrição de imagens com IA seguindo as diretrizes da audiodescrição?
- O artigo apresenta qual ferramenta é utilizada para descrição de imagem?

As respostas “SIM” receberam nota 1,0; “PARCIALMENTE” nota 0,5 e “NÃO” nota 0. A pontuação de corte foi 2.

Ainda na etapa de planejamento foram estabelecidos os campos para o formulário de extração de dados dos estudos selecionados: *link*, autores, ano da publicação, revista onde foi publicado, repositório, algoritmos e ferramentas de IA citadas, bases de dados e outras ferramentas exploradas e os resultados.

2.1. Condução

A aba de Condução destina-se a Pesquisa a partir da *string* estabelecida no planejamento; Importação dos estudos nos repositórios pré-definidos, Seleção de acordo com os critérios de inclusão e exclusão e Avaliação segundo o *checklist* de questões. Essa primeira etapa da condução foi realizada a partir da leitura do título e resumo de todos os 62 artigos retornados a partir da *string* nos respectivos repositórios. Ainda na etapa de condução, foi feita a leitura completa dos artigos selecionados para extração e análise dos dados, conforme detalhado na seção a seguir.

3. Resultados e Discussão

A partir da leitura dos 24 artigos selecionados, dois artigos foram descartados segundo os critérios de qualidade. Dos selecionados, um estudo foi publicado em 2017 (Campos, Gonçalves, Araújo, 2017), um em 2020 (Karagöz; Koruyan, 2020), um em 2021 (Vrindavanam *et al*, 2021), 3 em 2022 (dos Santos; Colombini;

Avila, 2022), (Angrave; Li; Zhong, 2022), (Soldan *et al*, 2022), 13 em 2023 (Xin *et a*l, 2023), (Oion *et al*, 2023), (Han *et al*, 2023), (Shahira; Pulkit; Lijiya, 2023), (Shanthi *et al*, 2023), (Shen *et al*, 2023), (Dittakan *et al*, 2023), (Sucharitha *et al*, 2023), (Perdigão *et al*, 2023), (Ferreira *et a*l, 2023), (Alam; Islam; Hoque, 2023), (Kruszielski *et al*, 2023), (Pitcher-Cooper *et al,* 2023) e 3 em 2024 (Oncescu *et al*, 2024), (Han *et al*, 2024), (Priya *et al,* 2024), o que converge com a afirmativa do início deste texto. Esse panorama evidencia o aumento recente do interesse e da produção científica sobre o uso da inteligência artificial na audiodescrição de imagens, especialmente a partir de 2023, refletindo o impacto da popularização da IA generativa nesse contexto.

Entre os repositórios, treze estudos foram retornados do IEEE Xplore, dois do Springer Link e sete retornaram do Google Acadêmico, a partir de repositórios diversos (ACM Digital Library; DergiPark Akademik; ERIC; MDPI; PMC - PubMed Central; SET; SOL - SBC Open Lib). A diversidade das fontes retornadas reforça o caráter multidisciplinar da temática.

No total, foram identificados 83 algoritmos e ferramentas de IA distintas nos estudos analisados. Dentre os mais citados, destacam-se as *Convolutional Neural Networks (CNNs)* que são as Redes Neurais Convolucionais utilizadas em tarefas de processamento de imagens e reconhecimento visual, presentes em oito estudos. Modelos de linguagem de grande porte também foram amplamente empregados, como, por exemplo:

- GPT - *Generative Pre-trained Transformer* – Transformador Pré-Treinado Generativo - uma arquitetura de modelo de linguagem que utiliza aprendizado profundo para gerar texto, com base em uma vasta quantidade de dados pré-treinados;
- Chat GPT – um *chatbot* que utiliza o modelo GPT; e
- BERT – *Bidirectional Encoder Representations from Transformers*, um modelo de inteligência artificial desenvolvido pelo Google para melhorar a compreensão da linguagem natural, e suas variantes.

Esse resultado demonstra a aplicabilidade em tarefas de geração textual automatizada. Além disso, modelos multimodais, como *CLIP – Contrastive Language-Image Pre-training* e *BLIP – Bootstrapping Language-Image* –, que integram linguagem natural com representação visual, foram identificados como soluções promissoras na geração de audiodescrição automatizada. Xin *et al*. (2023) apresentam uma visão geral dos avanços do que eles chamam de "legendagem visual", e propõem uma taxonomia sobre o assunto a partir de estudos realizados entre 2014 e 2018, baseados em *deep learning*, um subconjunto do aprendizado de máquina que utiliza redes neurais artificiais com múltiplas camadas para analisar dados e aprender padrões complexos, de forma semelhante à forma como o cérebro humano processa informações.

Outras abordagens recorrentes foram o uso de modelos sequenciais, como *LSTM – Long Short-Term Memory* (Memória de Curto e Longo Prazo), usada em redes neurais para processar e prever sequências de dados. Também a *RNN – Recurrent Neural Network* (Rede Neural Recorrente), que é um tipo de modelo de aprendizado profundo projetado para processar dados sequenciais, como texto e fala. Além das arquiteturas do tipo *Transformer*, especialmente aplicadas na geração de legendas para imagens. As ferramentas de conversão de texto em fala (*text-to-speech*), como Amazon Polly, Google TTS e Microsoft Azure Speech Services foram utilizadas para transformar descrições textuais em conteúdo acessível para pessoas com deficiência visual.

Foram também mapeadas 69 bases de dados distintas utilizadas no treinamento e validação dos modelos de IA. Entre as mais recorrentes destacam-se:

- *MS COCO (Microsoft Common Objects in Context):* conjunto de dados de grande escala utilizado para tarefas de detecção de objetos, segmentação de imagens e legendagem.
- Flickr8K: conjunto de dados com mais de 8.000 imagens, cada uma com 5 legendas, utilizado para treinar e avaliar o modelo de geração de legendas descritivas.

- *ImageNet*: base de dados extensa e amplamente utilizada para treinamento e avaliação de algoritmos de aprendizado de máquina, especialmente em tarefas de visão computacional.

- *LSMDC (Large Scale Movie Description Challenge):* Conjunto de dados que contém clipes de vídeo e descrições associadas, com o objetivo de desenvolver e avaliar métodos que possam gerar descrições automáticas para vídeos.

- *MAD (Movie Audio Description dataset):* Conjunto de dados utilizado para treinamento e avaliação de modelos de geração de descrição de áudio. Contendo mais de 384.000 sentenças em linguagem natural ancoradas em mais de 1.200 horas de vídeo, esse banco de dados escaláveis é focado em tarefas de ancoragem e comparação de linguagem em vídeos.

Esses bancos de dados, assim como o *CC3M (Conceptual Captions),* são utilizados para o treinamento de modelos de IA, permitindo que esses sistemas adquiram a capacidade de gerar descrições textuais detalhadas de imagens estáticas. Ressalta-se a presença de bases em português, como o *dataset* #PraCegoVer, o que evidencia um esforço de adaptação linguística e cultural no desenvolvimento de tecnologias assistivas voltadas ao contexto brasileiro.

Além de os algoritmos e bases de dados, a revisão destacou o uso de nove ferramentas adicionais, como o *JAWS (Job Access With Speech)* e o *NVDA (NonVisual Desktop Access)*, tecnologias assistivas de leitura de tela para pessoas com deficiência visual. Além de o *SeeChart*, uma ferramenta capaz de gerar resumos textuais de gráficos, o que seria particularmente útil em contextos educacionais e em ambientes virtuais de aprendizagem, como a plataforma *Moodle (Modular Object-Oriented Dynamic Learning Environment)*, promovendo maior acessibilidade em conteúdos visuais complexos.

Embora diversos estudos concentrem-se na descrição de vídeos, as audiodescrições de filmes apresentados em nove dos artigos selecionados são produzidas pelos sistemas a partir de *frames* de imagens estáticas; portanto, são também úteis para esse formato. Estudos como o de Campos, Gonçalves e Araújo (2017) evidenciam que, apesar de alguns sistemas gerarem audiodescrições adequadas dentro dos parâmetros de tempo e quantidade de palavras recomendados, a qualidade da descrição nem sempre é suficiente para garantir uma experiência satisfatória para o usuário final. A formulação das sentenças com expressões genéricas, como “a imagem mostra”, foi considerada inadequada frente às diretrizes da audiodescrição, por não oferecer informações claras ou específicas (Perdigão, 2017). Nesse sentido, a necessidade de validação por consultores com deficiência visual continua sendo fundamental, pois apenas essa validação pode garantir que as descrições geradas por IA atendam às exigências de clareza e utilidade para as pessoas com deficiência visual.

Verificou-se que, apesar de a maioria dos estudos estabelecer relação entre IA e acessibilidade, poucos atendem de forma rigorosa às diretrizes internacionais da audiodescrição, como aquelas estabelecidas pelas WCAG (W3C, 2018). A ausência de validação das descrições geradas por pessoas com deficiência visual ou consultores especializados foi um ponto crítico observado, comprometendo a efetividade das soluções propostas.

4. Conclusão

A análise dos 22 estudos selecionados revelou uma variedade de 83 Algoritmos e ferramentas de IA, 69 Bases de dados distintas e outras nove ferramentas utilizadas no processo de descrição de imagens com inteligência artificial, evidenciando diversas possibilidades de reuso de recursos disponíveis para o desenvolvimento de ferramentas automatizadas de audiodescrição.

A maioria dos estudos relaciona a IA com a descrição de imagens dinâmicas, mas através de um mapeamento das informações contidas em *frames* estáticos. A maioria dos estudos também apresenta a relevância da conexão entre IA e acessibilidade. A revisão revelou que, embora existam ferramentas e al-

goritmos promissores, ainda são necessários avanços, particularmente em relação à qualidade e à conformidade com as diretrizes de audiodescrição. A necessidade de validação por consultores com deficiência visual permanece fundamental, garantindo que as descrições sejam claras, coesas e úteis.

Portanto, os achados desta RSL indicam um campo em expansão, com ampla diversidade de recursos, algoritmos e bases de dados disponíveis para o desenvolvimento de sistemas de audiodescrição. No entanto, reforça-se a necessidade de avanços metodológicos e validação prática das soluções, a fim de garantir sua efetiva contribuição para a acessibilidade na Educação a Distância.

Biodados e contatos dos autores

Após o artigo aprovado, os autores serão solicitados a incluir seus Biodados, conforme o modelo abaixo. Essa área do artigo é opcional, mas caso haja interesse, todos os autores deverão consentir a autorização do uso de sua imagem (foto 3x4). Recomenda-se que, ao incluir o dados e resumo do autor, seja citado qual foi a sua participação na pesquisa ou redação do artigo.

PERDIGÃO, L. é Coordenadora do Núcleo de Acessibilidade e Inclusão da Fundação Cecierj. Completou o seu doutorado no Programa de Pós-graduação em Ciências, Tecnologias e Inclusão da Universidade Federal Fluminense, onde prossegue com o Pós Doc. Seus interesses de pesquisa incluem INCLUSÃO NO ENSINO SUPERIOR, POLÍTICAS PÚBLICAS e TECNOLOGIAS ASSISTIVAS com destaque para AUDIODESCRIÇÃO.

Atuou na coleta de dados, análises estatísticas e redação final deste artigo.

ORCID: 0000-0002-5662-212X
Contato: +55 21 998399055
E-mail: lperdigao@cecierj.edu.br

PINTO, S. é Professor Associado na Universidade Federal Fluminense UFF. Docente credenciado como membro permanente do Programa de Doutorado em Ciência, Tecnologia e Inclusão da UFF. Líder do grupo de pesquisa CNPq: TeCEADI+: Tecnologias Computacionais no ensino e aprendizagem na ótica da Diversidade, Inclusão e Inovação.

Atuou na orientação para essa pesquisa.

ORCID: 0000-0001-6914-2398
Contato: +55 22 992153838
E-mail: screspo@id.uff.br

FERREIRA, C. D. é graduando em Ciência da Computação na Universidade Federal Fluminense, campus Rio das Ostras – RJ. Bolsista PIBIC no projeto sobre Uso de Inteligência Artificial para Acessibilidade de conteúdos EAD na Plataforma Moodle.

Atuou na análise de dados para essa pesquisa.

ORCID: 0009-0006-3625-7133
Contato: +55 21 994756640
E-mail: caio_ferreira@id.uff.br

Referências Bibliográficas

ALAM, M. Z. I.; ISLAM, S.; HOQUE, E. **SeeChart: enabling accessible visualizations through interactive natural language interface for people with visual impairments.** In: INTERNATIONAL CONFERENCE ON INTELLIGENT USER INTERFACES (IUI), 28., 2023. Proceedings [...]. New York: ACM, 2023. p. 46–64. DOI: 10.1145/3581641.3584099.

ANGRAVE, L.; LI, J.; ZHONG, N. **Creating TikToks, memes, accessible content, and books from engineering videos?** First solve the scene detection problem. Grantee Submission, 2022.

CAMPOS, V. P.; GONÇALVES, L. M.; ARAÚJO, T. M. **Applying audio description for context understanding of surveillance videos by people with visual impairments.** In: IEEE INTERNATIONAL CONFERENCE ON ADVANCED VIDEO AND SIGNAL BASED SURVEILLANCE, 14., 2017, Lecce. Proceedings [...]. IEEE, 2017. p. 1–5. DOI: 10.1109/AVSS.2017.8078530

Dittakan, K., Prompitak, K., Thungklang, P. *et al*. **Image caption generation using transformer learning methods: a case study on Instagram image.** Multimedia Tools and Applications, v. 83, p. 46397–46417, 2024. DOI: 10.1007/s11042-023-17275-9

DOS SANTOS, G. O.; COLOMBINI, E. L.; AVILA, S. **#PraCegoVer: a large dataset for image captioning in Portuguese.** Data, v. 7, n. 2, p. 13, 2022. DOI: 10.3390/data7020013

FERREIRA, L. *et al*. **Presenter-centric image collection and annotation: enhancing accessibility for the visually impaired.** In: SIBGRAPI CONFERENCE ON GRAPHICS, PATTERNS AND IMAGES, 36., 2023, Rio Grande. Proceedings [...]. IEEE, 2023. p. 199–204. DOI: 10.1109/SIBGRAPI59091.2023.10347135

HAN, T. *et al*. **AutoAD II: the sequel – who, when, and what in movie audio description.** In: IEEE/CVF INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV), 2023, Paris. Proceedings [...]. IEEE, 2023. p. 13599–13609. DOI: 10.1109/ICCV51070.2023.01255.

HAN, T. *et al*. **AutoAD: movie description in context.** In: CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2023, Vancouver. Proceedings [...]. IEEE/CVF, 2023. p. 18930–18940. DOI: 10.1109/CVPR52729.2023.01815.

KARAGÖZ, E.; KORUYAN, K. **Design of audio description system using cloud based computer vision.** Mehmet Akif Ersoy Üniversitesi Uygulamalı Bilimler Dergisi, v. 4, n. 1, p. 74–85, 2020. DOI: 10.31200/makuubd.651261.

KITCHENHAM, B. ; CHARTERS, S. **Guidelines for performing Systematic Literature Reviews in Software Engineering** (EBSE 2007-001). Keele University and Durham University Joint Report, 2007.

KRUSZIELSKI, L. F. *et al*. **The Use of Artificial Intelligence Enabling Scalable Audio Description on Brazilian Television: A Workflow Proposal.** SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING, [S. l.], v. 9, n. 1, 2025. Disponível em: https://revistaeletronica.set.org.br/ijbe/article/view/271

LIMA, F. J. de; TAVARES, F. S. S. **Subsídios para a construção de um código de conduta do áudio-descritor**. Revista Brasileira de Tradução Visual (RBTV), 2010. Disponível em http://www.associadosdainclusao.com.br/enades2016/sites/all/themes/berry/documentos/07-subsidios-para-a-construcao-de-um-codigo-de-conduta.pdf - Acesso em setembro de 2016.

OION, M. S. R. *et al.* **A machine learning based image to object detection and audio description for blind people.** In: INTERNATIONAL CONFERENCE ON COMPUTER AND INFORMATION TECHNOLOGY (ICCIT), 26., 2023, Cox's Bazar. Proceedings [...]. IEEE, 2023. p. 1–6. DOI: 10.1109/ICCIT60459.2023.10441089

ONCESCU, A.-M. *et al.* **A sound approach: using large language models to generate audio descriptions for egocentric text-audio retrieval.** In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEE-

CH AND SIGNAL PROCESSING (ICASSP), 2024, Seoul. Proceedings [...]. IEEE, 2024. p. 7300–7304. DOI: 10.1109/ICASSP48485.2024.10448486

PARISFAL. **Review Planning**. 2021. Disponível em: https://parsif.al/help/#planning - Acesso em: 15 maio 2025.

PERDIGÃO, L. T. **Conteúdos audiovisuais acessíveis na educação a distância: a audiodescrição didática como pós-produção.** 2023. 150 f. Tese (Doutorado em Ciências, Tecnologias e Inclusão) – Universidade Federal Fluminense, Niterói, 2023.

PERDIGÃO, L. T.; LIMA, N. R. W. **Vendo com outros olhos: a audiodescrição na educação a distância.** Niterói: [s. n.], 2017. 100 p. Il. Ebook. Disponível em: https://educapes.capes.gov.br/bitstream/capes/429946/4/GuiaPDFfinal_2020.pdf - Acesso em: 15 maio 2025.

PERDIGÃO, L. T.; MONTEIRO, F. V.; FERNANDES, E. M. **Audiodescrição como texto alternativo nas principais redes sociais digitais.** In: CONGRESSO INTERNACIONAL INTERDISCIPLINAR EM SOCIAIS E HUMANIDADES, 10., 2021, Niterói. Anais [...]. Niterói: Programa de Pós-Graduação, 2021. Disponível em: https://www.even3.com.br/anais/xc22021/435829-AUDIODESCRICAO-COMO-TEXTO-ALTERNATIVO--NAS-PRINCIPAIS-REDES-SOCIAIS-DIGITAIS - Acesso em: 21 out. 2024.

PERDIGÃO, L. *et al*. **Inteligência artificial para audiodescrição de imagens: uma análise da pessoa com deficiência visual.** In: CONGRESSO SOBRE TECNOLOGIAS NA EDUCAÇÃO, 8., 2023, Santarém. Anais [...]. Porto Alegre: SBC, 2023. p. 182–191. DOI: 10.5753/ctrle.2023.232651

PITCHER-COOPER, C. *et al.* **You described, we archived: a rich audio description dataset.** Journal on Technology and Persons with Disabilities, v. 11, p. 192–208, 2023.

REVATHY P. *et al*. **Seeing with sound: automatic image captioning with auditory output for the visually impaired.** In: INTERNATIONAL CONFERENCE ON RESEARCH METHODOLOGIES IN KNOWLEDGE MANAGEMENT, ARTIFICIAL INTELLIGENCE AND TELECOMMUNICATION ENGINEERING (RMKMATE), 2023, Chennai. Proceedings [...]. IEEE, 2023. p. 1–5. DOI: 10.1109/RMKMATE59243.2023.10368610

SHANTHI, N. *et al*. **Deep learning based audio description of visual content by enhancing accessibility for the visually impaired.** In: INTERNATIONAL CONFERENCE ON SUSTAINABLE COMMUNICATION NETWORKS AND APPLICATION (ICSCNA), 2023, Theni. Proceedings [...]. IEEE, 2023. p. 1234–1240. DOI: 10.1109/ICSCNA58489.2023.10370414

SHEN, X. *et al*. **Fine-grained audible video description.** In: CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2023, Vancouver. Proceedings [...]. IEEE/CVF, 2023. p. 10585–10596. DOI: 10.1109/CVPR52729.2023.01020

SUCHARITHA, V. *et al*. **Imaging description production by means of deeper neural networks.** In: INTERNATIONAL CONFERENCE ON ADVANCES IN COMPUTING, COMMUNICATION AND APPLIED INFORMATICS (ACCAI), 2023, Chennai. Proceedings [...]. IEEE, 2023. p. 1–5. DOI: 10.1109/ACCAI58221.2023.10200618

VRINDAVANAM, J. *et al*. **Machine learning based approach to image description for the visually impaired.** In: ASIAN CONFERENCE ON INNOVATION IN TECHNOLOGY (ASIANCON), 2021, Pune. Proceedings [...]. IEEE, 2021. p. 1–6. DOI: 10.1109/ASIANCON51346.2021.9544867

WORLD WIDE WEB CONSORTIUM. **Web Content Accessibility Guidelines (WCAG) 2.1**, 2018. Disponível em: https://www.w3.org/TR/WCAG21/ - Acesso em: 15 maio 2025.

XIN, B. *et al*. **A comprehensive survey on deep-learning-based visual captioning.** Multimedia Systems, v. 29, p. 3781–3804, 2023. DOI: 10.1007/s00530-023-01175-x