ISSN: 2358-8829

APLICATIVO MOBILE PARA CAPTURA E DESCRIÇÃO DE IMAGENS COM INTELIGÊNCIA ARTIFICIAL PARA ESTUDANTES COM DEFICIÊNCIA VISUAL

Nayron Morais Almeida[1]
Ana Júlia da Silva Madeira[2]
Willyan Sousa da Silva[3]
Marina Laís Costa da Silva[4]

1 INTRODUÇÃO

No cotidiano escolar, imagens, tabelas e objetos ajudam a apresentar ideias, explorar conceitos e desenvolver atividades de aprendizagem. Para estudantes com deficiência visual, porém, o acesso a esses recursos pode ser limitado quando as informações são oferecidas apenas de forma visual. Tornar esse conteúdo acessível é, portanto, parte das condições necessárias para que esses estudantes participem das atividades e tenham acesso ao currículo. Essa compreensão está em consonância com a Lei Brasileira de Inclusão, que prevê recursos de acessibilidade e de tecnologia assistiva no contexto educacional (Brasil, 2015).

Diante dessa necessidade, o uso de dispositivos móveis associado à Inteligência Artificial (IA) abre possibilidades de apoio ao acesso à informação visual. A integração entre a câmera do celular, a geração de descrições em português e a reprodução sonora dessas descrições pode oferecer aos estudantes outra forma de conhecer o conteúdo de materiais e identificar elementos presentes em seu entorno.

Nesse contexto, este trabalho tem como objetivo desenvolver um aplicativo móvel para capturar imagens e gerar descrições em português, disponibilizadas em texto e áudio, voltado ao apoio de estudantes com deficiência visual, desta forma, ampliando as possibilidades de acesso aos materiais utilizados na escola. Para isso, apresenta-se o desenvolvimento concluído do protótipo e discutem-se os resultados de testes funcionais, considerando a integração entre seus componentes e as descrições produzidas. Esses resultados constituem uma etapa inicial da investigação, cuja continuidade deverá incluir a

[1] Professor do Instituto Estadual de Educação, Ciência e Tecnologia do Maranhão (IEMA). E-mail: nayronmorais@gmail.com.
[2] Estudante do Curso Técnico em Desenvolvimento de Sistemas do IEMA. E-mail: anajuliadasilvamadeira2@gmail.com.
[3] Estudante do Curso Técnico em Desenvolvimento de Sistemas do IEMA. E-mail: willyansousa305@gmail.com.
[4] Estudante do Curso Técnico em Desenvolvimento de Sistemas do IEMA. E-mail: marinalaiss2022@gmail.com.

ISSN: 2358-8829

avaliação com o público-alvo para compreender a experiência de uso, a acessibilidade e as possíveis contribuições da ferramenta às atividades escolares, bem como o uso da ferramenta fora do ambiente escolar.

2 METODOLOGIA

A pesquisa é aplicada, de desenvolvimento tecnológico, com abordagem qualitativa e exploratória. O percurso compreendeu cinco etapas: levantamento de requisitos, planejamento da interface, implementação do aplicativo, integração com o modelo de IA e testes funcionais.

Foram definidos como requisitos a captura sob demanda, a descrição em português, a apresentação em texto e áudio e o acionamento por voz. A interface foi planejada com poucos comandos e botões amplos, buscando facilitar a interação. Sua adequação ao público-alvo ainda será avaliada.

O aplicativo foi desenvolvido em Dart, com o *framework* Flutter e as bibliotecas *camera*, *flutter_tts* e *speech_to_text*, responsáveis pela captura, síntese de voz e reconhecimento de comandos falados. O servidor local utiliza Node.js, Express e Multer para encaminhar as imagens à API e devolver as descrições. Nos testes, foi utilizado o modelo gpt-4.1-mini, da OpenAI, com instruções para produzir descrições breves em português, considerando objetos, textos legíveis e posições dos elementos.

A análise qualitativa consistiu na comparação entre as descrições apresentadas e os elementos visíveis nas imagens, considerando a identificação dos conteúdos, a organização das informações e as relações espaciais mencionadas. Nesta etapa não houve a avaliação com estudantes ou coleta de dados pessoais de participantes. A avaliação com o público-alvo está prevista para a continuidade da pesquisa e deverá observar os procedimentos éticos aplicáveis, incluindo consentimento em formato acessível e apreciação por instância competente, quando pertinente.

3 REFERENCIAL TEÓRICO

O acesso aos materiais didáticos deve considerar a informação necessária à atividade de aprendizagem. Stangl et al. (2021) mostram que as necessidades de descrição variam conforme o contexto. Na escola, reconhecer a estrutura de uma página e ler um enunciado são tarefas distintas. Assim, o apoio tecnológico precisa se articular aos objetivos de aprendizagem e à mediação docente.

Também é necessário que o estudante consiga operar a ferramenta. As Diretrizes de Acessibilidade para Conteúdo Web (WCAG) 2.2 orientam aspectos como alternativas textuais, contraste e identificação dos controles (W3C, 2023). Esses princípios ajudam a

planejar a interface, mas sua adoção deve ser acompanhada de verificações com tecnologias assistivas e avaliações com os usuários.

Entre os trabalhos relacionados, Gurari et al. (2020) apresentam o VizWiz-Captions, com imagens produzidas por pessoas cegas. Ahsan et al. (2021) incorporam textos detectados à geração de legendas, enquanto Mohanbabu e Pavel (2024) exploram o contexto de páginas *web* para orientar descrições, com avaliação de pessoas cegas ou com baixa visão. Esses estudos fundamentam a investigação de descrições adequadas ao conteúdo e à tarefa.

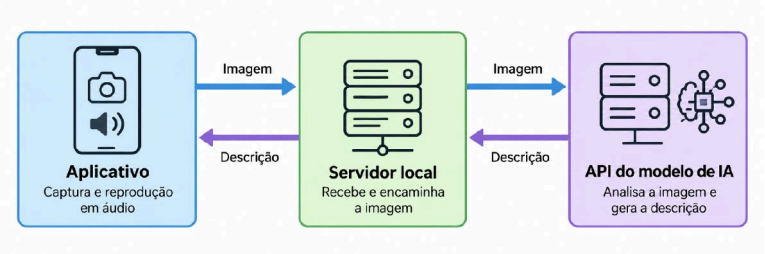
A fidelidade das respostas também exige atenção. Rohrbach et al. (2018) discutem alucinações nas quais os sistemas mencionam objetos ausentes da imagem.

4 RESULTADOS E DISCUSSÃO

4.1 Integração entre o aplicativo e o modelo de IA

O desenvolvimento resultou em um protótipo que integra captura de imagens, geração de descrições em português e apresentação das respostas em texto e áudio. A arquitetura reúne o aplicativo, o servidor local e a API do modelo de IA, conforme apresentado na Figura 2. Ao acionar a função de descrição, o aplicativo captura uma fotografia e a encaminha ao servidor, que solicita seu processamento ao modelo e devolve o texto produzido.

Figura 2 – Arquitetura do aplicativo



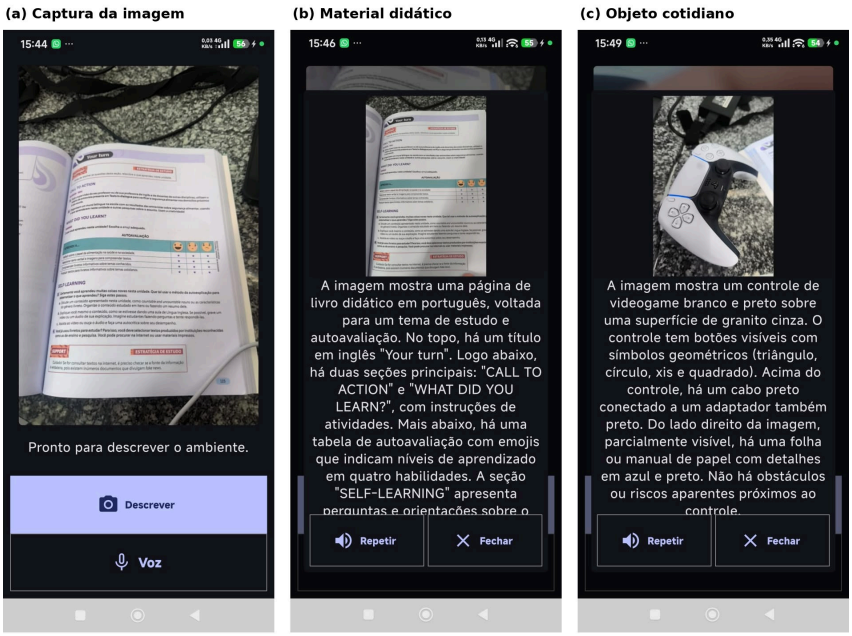
Fonte: elaborado pelos próprios autores (2026).

4.2 Interface e recursos de acessibilidade

A tela principal concentra os comandos “Descrever” e “Voz”, acompanhados da visualização da câmera e de uma mensagem de estado (Figura 3a). Após o processamento, uma janela apresenta a fotografia, a descrição e os botões “Repetir” e “Fechar” (Figuras 3b e 3c). Essa organização mantém as principais ações próximas ao conteúdo e permite consultar novamente a resposta sem realizar outra captura. Além disso, é possível controlar por comandos de voz em português, ao clicar no botão “voz”. Os comandos por voz funcionam por meio de palavras-chave: “descrever” ou “capturar” para descrever o ambiente; “repetir”

para repetir a descrição da foto atual; e, “ajuda” para ouvir os comandos por voz disponíveis. Esses recursos oferecem alternativas à interação exclusivamente visual.

Figura 3 – Protótipo e resultados dos testes funcionais



Fonte: elaborado pelos próprios autores (2026).

4.3 Testes funcionais

Em um teste com uma página de livro didático (Figura 3a), a resposta identificou o título *Your turn*, as seções *Call to action* e *What did you learn?*, uma tabela de autoavaliação com emojis e a seção *Self-learning* (Figura 3b). A descrição apresentou, portanto, uma visão geral da organização da página, contemplando elementos que poderiam orientar uma primeira aproximação ao material.

Do ponto de vista pedagógico, esse resultado sugere uma possibilidade de apoio à localização de conteúdos e ao reconhecimento das partes de uma atividade. Assim, o registro evidencia uma descrição estrutural, cuja utilidade para responder às atividades será investigada com os estudantes. Essa distinção dialoga com Stangl et al. (2021), ao considerar que a informação necessária depende da tarefa realizada.

No segundo caso apresentado (Figura 3c), o sistema descreveu um controle de videogame branco e preto, os símbolos geométricos dos botões e a superfície de granito cinza sobre a qual estava apoiado. Também mencionou um cabo e um adaptador na parte superior da imagem, além de um papel parcialmente visível à direita. A resposta contemplou características do objeto e relações espaciais entre os elementos enquadrados. O resultado indica que, nesse exemplo, a descrição ultrapassou a identificação do objeto principal e

incluiu informações sobre seu entorno. Já a afirmação de ausência de riscos aparentes faz parte do comando dado ao modelo IA.

4.4 Alcance dos resultados e implicações pedagógicas

Os dois casos apresentados documentam o funcionamento do protótipo e mostram possibilidades de descrição de materiais escritos e objetos. A análise também evidencia a importância de distinguir a produção de uma resposta compreensível da fidelidade e da suficiência dessa resposta para uma tarefa escolar. Conforme discutem Rohrbach et al. (2018), sistemas de descrição podem mencionar elementos ausentes da imagem, o que reforça a necessidade de verificar o conteúdo gerado.

5 CONSIDERAÇÕES FINAIS

A presente pesquisa concluiu um protótipo que integra captura sob demanda, descrição multimodal e apresentação em texto e áudio. Os testes identificaram elementos estruturais das situações testadas, sustentando a viabilidade funcional da integração. A acessibilidade e a utilidade pedagógica ainda precisam de avaliação.

Como trabalho futuro prioritário, prevê-se avaliar o aplicativo com estudantes cegos e com baixa visão, observando sucesso nas tarefas, necessidade de ajuda, tempo de resposta, clareza e fidelidade das descrições. Essa etapa deverá observar os procedimentos éticos aplicáveis e consentimento acessível. Os resultados orientarão ajustes de interface, modos de leitura textual e descrição de ambientes e estudo de processamento local.

REFERÊNCIAS

AHSAN, Hiba et al. Multi-modal image captioning for the visually impaired. *In:* CONFERENCE OF THE NORTH AMERICAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS: STUDENT RESEARCH WORKSHOP, 2021, Online. **Proceedings [...]**. Online: Association for Computational Linguistics, 2021. p. 53–60. DOI: 10.18653/v1/2021.naacl-srw.8. Disponível em: https://aclanthology.org/2021.naacl-srw.8/. Acesso em: 20 out. 2026.

BRASIL. **Lei nº 13.146, de 6 de julho de 2015**. Institui a Lei Brasileira de Inclusão da Pessoa com Deficiência (Estatuto da Pessoa com Deficiência). Brasília, DF: Presidência da República, 2015. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13146.htm. Acesso em: 20 out. 2026.

GURARI, Danna et al. Captioning images taken by people who are blind. *In:* EUROPEAN CONFERENCE ON COMPUTER VISION, 16., 2020, Glasgow. **Computer Vision – ECCV 2020**. Cham: Springer, 2020. pt. 17, p. 417–434. DOI: 10.1007/978-3-030-58520-4_25. Disponível em: https://doi.org/10.1007/978-3-030-58520-4_25. Acesso em: 20 out. 2026.

MOHANBABU, Ananya Gubbi; PAVEL, Amy. Context-aware image descriptions for web accessibility. *In:* INTERNATIONAL ACM SIGACCESS CONFERENCE ON COMPUTERS AND ACCESSIBILITY, 26., 2024, St. John's. **Proceedings [...]**. New York: Association for Computing Machinery, 2024. DOI: 10.1145/3663548.3675658. Disponível em: https://doi.org/10.1145/3663548.3675658. Acesso em: 20 out. 2026.

ROHRBACH, Anna et al. Object hallucination in image captioning. *In:* CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING, 2018, Brussels. **Proceedings [...]**. Brussels: Association for Computational Linguistics, 2018. p. 4035–4045. DOI: 10.18653/v1/D18-1437. Disponível em: https://aclanthology.org/D18-1437/. Acesso em: 29 out. 2026.

STANGL, Abigale et al. Going beyond one-size-fits-all image descriptions to satisfy the information wants of people who are blind or have low vision. *In:* INTERNATIONAL ACM SIGACCESS CONFERENCE ON COMPUTERS AND ACCESSIBILITY, 23., 2021, Virtual Event. **Proceedings [...]**. New York: Association for Computing Machinery, 2021. p. 1–23. DOI: 10.1145/3441852.3471233. Disponível em: https://doi.org/10.1145/3441852.3471233. Acesso em: 29 out. 2026.

WORLD WIDE WEB CONSORTIUM (W3C). **Web Content Accessibility Guidelines (WCAG) 2.2**. [S. l.]: W3C, 2023. Recomendação de 5 de outubro de 2023. Disponível em: https://www.w3.org/TR/2023/REC-WCAG22-20231005/. Acesso em: 20 out. 2026.