

DiagramLens: Transformando Modelos Visuais de Engenharia de Software em Descrições Narrativas Assistidas por IA

Almerindo da Paixão Junior¹, Lucas Santos de Oliveira², Alex Ferreira dos Santos^{1,3}

¹Universidade Estadual de Feira de Santana (UEFS)
Av. Transnordestina, Feira de Santana - BA, 44036-900, Brasil

²Universidade Estadual do Sudoeste da Bahia (UESB)
Av. José Moreira Sobrinho, Jequié - BA, 45205-490, Brasil

³Universidade Federal do Recôncavo da Bahia (UFRB)
Av. Centenário, Feira de Santana - BA, 44042-280, Brasil

almerindopaixao@gmail.com, lsoliveira@uesb.edu.br,
alex.ferreira@ufrb.edu.br

Abstract. *The use of video lessons in digital education has broadened access to learning. However, in the context of Software Engineering, the lack of adequate descriptions of diagrams, such as UML and ERD, still limits the autonomy of visually impaired students. This article presents DIAGRAMLENS, an audio description tool based on Artificial Intelligence. The solution uses the YOLO 11n model to detect diagrams and Gemini 2.5 Flash-Lite to generate textual descriptions, converted into audio via the Web Speech API. The results highlight the educational potential of the tool, confirmed by expert evaluation, as well as its compliance with web accessibility guidelines, attesting to its effectiveness in promoting autonomy.*

Resumo. *O uso de videoaulas na educação digital ampliou o acesso ao ensino. Entretanto, no contexto da Engenharia de Software, a ausência de descrições adequadas de diagramas, como UML e DER, ainda limita a autonomia de estudantes com deficiência visual. Este artigo apresenta o DIAGRAMLENS, uma ferramenta de audiodescrição baseada em Inteligência Artificial. A solução utiliza o modelo YOLO 11n para detectar diagramas e o Gemini 2.5 Flash-Lite para gerar descrições textuais, convertidas em áudio por meio da Web Speech API. Os resultados evidenciam o potencial educacional da ferramenta, confirmado pela avaliação de especialistas, além da conformidade com diretrizes de acessibilidade web, atestando sua eficácia na promoção da autonomia.*

1. Introdução

O cenário de baixa representatividade de pessoas com deficiência nas Instituições de Ensino Superior (IES) brasileiras contrasta com os princípios estabelecidos pela Lei nº 13.146/2015, que assegura igualdade de oportunidades e proíbe qualquer forma de discriminação [Rocha and Miranda 2009]. Apesar desses avanços legais, persistem barreiras de acesso e permanência no ensino superior e, em particular, na área de Computação. Em 2022, estudantes com deficiência representaram menos de 1% das

matrículas de graduação no país [Marco et al. 2025]. Além disso, dados do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) [Inep 2026] indicam a existência de 95.572 estudantes com deficiência, superdotação ou transtorno global do desenvolvimento no ensino superior, dos quais 22,8% possuem deficiência visual (DV), evidenciando a relevância desse público e a necessidade de estratégias efetivas de inclusão acadêmica.

Esse cenário torna-se ainda mais desafiador na área de computação, cujas disciplinas fundamentais dependem fortemente de representações visuais para a compreensão de conceitos abstratos. Em Banco de Dados, a modelagem conceitual e lógica utiliza diagramas Entidade-Relacionamento (DER) para representar relações entre dados e constitui um elemento essencial para o projeto de sistemas consistentes [Franck et al. 2021]. De forma semelhante, no desenvolvimento de sistemas, os diagramas da Linguagem de Modelagem Unificada (UML) são amplamente empregados para especificar, visualizar e documentar artefatos de software e representam aspectos funcionais e estruturais dos sistemas [Gavric et al. 2024].

Para a parcela significativa de estudantes com DV que ingressam no ensino superior, especificamente em cursos de computação, a exigência de interpretar representações complexas, como os diagramas UML e DER, constitui uma barreira formativa expressiva. Apesar de alguns esforços na implementação de tecnologias assistivas para auxiliar esse público [Kennel 1996, Luque et al. 2016, Kour et al. 2025], essas ferramentas ainda estão distantes de atender plenamente às suas necessidades educacionais [Magalhães and Neto 2010]. Esse cenário de exclusão é agravado pela crescente disseminação de conteúdos em formato de vídeo (videoaulas) na educação digital para fins didáticos e de ensino [Barrére 2014]. Se, por um lado, as videoaulas trouxeram melhorias para a escala do ensino, por outro, acentuaram as barreiras para os discentes com DV [Lima and Fonseca 2016]. Os leitores de tela tradicionais não conseguem interpretar o conteúdo dinâmico da videoaula, a disposição espacial dos nós nos diagramas UML ou as relações e cardinalidades em um DER, tornando essas representações visuais inacessíveis [Verde et al. 2023, Klein et al. 2025]. Além disso, a apresentação de diagramas sem uma descrição clara ou completa nos vídeos impede a compreensão total dos conceitos, o que força uma dependência de mediadores humanos para descrever essas abstrações e reduz drasticamente a autonomia do estudante [Pavão and Pavão 2020].

Nesse contexto, este trabalho propõe a **DIAGRAMLENS**, uma ferramenta que utiliza Inteligência Artificial (IA) para a audiodescrição automática de diagramas em videoaulas de computação, com o objetivo de devolver a autonomia do aluno em seu processo de aprendizagem. Essa solução identifica e descreve diagramas utilizados na representação de sistemas computacionais, como os diagramas UML e DER, em videoaulas por meio de modelos de IA. Para isso, utiliza o modelo de visão computacional *You Only Look Once* (YOLO), em sua variação 11n, e o *Large Language Model* (LLM) Gemini 2.5 Flash-Lite para a geração de descrições semânticas textuais desses diagramas, as quais são convertidas em áudio via *Text-to-Speech* (TTS) utilizando a Web Speech API. Como principal contribuição deste trabalho, a ferramenta foi desenvolvida com foco em acessibilidade, oferecendo navegação por teclado e compatibilidade com leitores de tela.

Para validar a solução, foram realizados dois tipos de avaliação. A primeira, conduzida com especialistas, atestou a qualidade das descrições semânticas geradas pela

LLM, evidenciado pela aprovação unânime dos avaliadores obtida no quesito de benefício educacional. A segunda frente, focada na acessibilidade, comprovou a conformidade da plataforma com diretrizes de acessibilidade web (validadas via Lighthouse, WAVE e AccessMonitor). Os resultados trazem indícios preliminares de viabilidade e potencial da ferramenta na promoção de maior autonomia educacional de estudantes com DV.

2. Fundamentação Teórica

A compreensão de como a autonomia é afetada no contexto digital encontra amparo na Teoria da Distância Transacional (TDT). Essa teoria define a distância na educação não pelo viés geográfico, mas como um espaço cognitivo e comunicacional entre docente e discente, sujeito a potenciais mal-entendidos quanto à interpretação do conteúdo [Moore 2002]. Essa lacuna é moldada por três variáveis interdependentes: o Diálogo (interação professor-aluno), a Estrutura (rigidez e organização do programa) e a Autonomia do estudante. Conforme Moore [Moore 2002], quanto mais rígido o curso e menor o diálogo, maior será a distância transacional, o que exige um grau elevado de independência do aluno para o sucesso da aprendizagem. No caso de estudantes com DV, a falta de acessibilidade em gráficos e elementos visuais atua como uma barreira que amplia esse distanciamento, limitando a capacidade de aprender por conta própria e sem intervenções externas. Ao oferecer uma ferramenta de audiodescrição automática, reduz-se a dependência de intermediários, auxiliando o estudante a compreender elementos visuais de forma autônoma.

No entanto, a automação da audiodescrição, por si só, não garante a redução dessa distância se a interação com a ferramenta for complexa. Sob a ótica da Teoria da Carga Cognitiva (TCC), a memória de trabalho humana possui capacidade limitada, que pode ser saturada pelo excesso de informações irrelevantes [Sweller 1988]. Portanto, a eficácia de uma interface assistiva para estudantes com DV depende da minimização do esforço mental necessário para operar o sistema. Atividades de busca exaustiva, como a navegação manual por menus extensos e seções complexas, geram uma alta carga cognitiva que não contribui para o aprendizado efetivo. Ao priorizar informações essenciais, a interface evita o colapso cognitivo decorrente da coordenação simultânea de múltiplas fontes de dados, liberando recursos mentais para a assimilação do conteúdo educacional da videoaula.

3. Trabalhos Relacionados

Embora a relevância dessa área de pesquisa para o apoio a pessoas com DV, diversos trabalhos têm explorado abordagens distintas. Uma das primeiras soluções foi proposta por [Kennel 1996] ao desenvolver o AudioGraf, um leitor de diagramas projetado para possibilitar que pessoas com DV leiam e interpretem diagramas de forma independente por meio de uma interface áudio-tátil. Apesar de oferecer autonomia ao transformar elementos gráficos em respostas auditivas, o sistema apresentava a limitação de depender da configuração manual de *links* (relações entre elementos) para que a interpretação fosse completa, exigindo a preparação prévia de cada diagrama.

Apesar de surgirem novas soluções para a análise de diagramas, muitas ainda apresentam objetivos distintos e não utilizam IA. Neste contexto [Kour et al. 2025], desenvolveu um Assistente UML baseado no modelo YOLOv8 para a detecção de elementos em

diagramas de classes UML estáticos. Apesar de fornecer uma metodologia para extração semântica e consulta interativa, o que possibilita ao usuário percorrer manualmente a estrutura lógica do diagrama, ele se restringe ao padrão UML e a imagens fixas.

Diferente das abordagens baseadas em interpretação automática de diagramas, o trabalho de [Luque et al. 2016] apresenta o Model2getherr, um protótipo web voltado para a modelagem de diagramas UML em tempo real, suportando o trabalho conjunto entre pessoas com e sem DV. A ferramenta utiliza uma interface dupla: enquanto usuários sem DV interagem com diagramas gráficos, usuários com DV utilizam uma interface textual baseada em uma Linguagem Específica de Domínio (DSL). O diferencial desta abordagem reside nos mecanismos que notificam o usuário cego imediatamente sobre modificações ou destaques feitos pelo instrutor no diagrama, por meio de bipes sonoros ou atualizações textuais em tempo real. Embora o Model2gether resolva a questão da colaboração síncrona em ambientes de EaD, ele ainda depende de uma representação textual estruturada pré-existente ou editada manualmente para funcionar.

Além dos desafios centrados em diagramas UML, alguns trabalhos buscam mitigar as barreiras no ensino de computação com foco no domínio de modelagem de dados como, por exemplo, as ferramentas voltadas aos DER. O sistema ApreNDER, proposto por [Magalhães and Neto 2010], estruturou uma abordagem que utiliza uma árvore de navegação com retorno sonoro para guiar o estudante com DV na criação de tabelas e relacionamentos conceituais. Recentemente, Junior e Pereira [Junior and Pereira 2025] avançaram nessa linha ao desenvolver o ambiente web inclusivo Diagrama Além da Visão (DAV). O DAV também possibilita que o próprio discente construa e manipule um DER de forma dinâmica por meio de uma interface adaptada para navegação sequencial via teclado. Contudo, por se tratarem de ambientes focados na criação manual e na interpretação de comandos estruturados pré-definidos, tanto o ApreNDER quanto o DAV não contemplam o suporte à interpretação de diagramas apresentados em conteúdos educacionais em formato de vídeo, limitando-se à interação direta dentro da própria ferramenta.

Diferentemente de ambientes focados na criação manual de diagramas (como ApreNDER e DAV) ou de soluções que dependem de representações estruturadas e intervenções prévias (como o uso de DSLs no Model2gether ou *links* no AudioGraf), o DIAGRAMLENS inova ao atuar de forma autônoma. Ao utilizar a IA para gerar audiodescrições diretamente das videoaulas, a ferramenta elimina a dependência de um mediador humano para preparar ou traduzir o conteúdo. Para sintetizar as abordagens da literatura e evidenciar as lacunas existentes, o Quadro 1 apresenta uma análise comparativa dessas ferramentas frente à solução proposta neste trabalho.

Quadro 1. Comparativo entre o DIAGRAMLENS e outras ferramentas

Características	AudioGraf	Assistente UML	Model2gether	ApreNDER/DAV	DiagramLens
Acessibilidade para DV	✓	✓	✓	✓	✓
Interpretação de Diagramas	✓	✓	✗	✗	✓
Suporte à Criação/Edição de Diagramas	✗	✗	✓	✓	✗
Sem Preparação Manual	✗	✓	✗	✗	✓
Colaboração Síncrona	✗	✗	✓	✗	✗
Interação via DSL	✗	✓	✓	✗	✗
Suporte a Videoaulas	✗	✗	✗	✗	✓
Audiodescrição Semântica	✗	✗	✗	✗	✓

4. Metodologia de Desenvolvimento do DIAGRAMLENS

Fundamentado nas teorias de Distância Transacional e Carga Cognitiva, e nas limitações identificadas em trabalhos relacionados, o desenvolvimento do DIAGRAMLENS foi conduzido sob a perspectiva de três processos funcionais. O primeiro consistiu em garantir o processamento e a detecção eficiente de diagramas em videoaulas; o segundo focou em fornecer e otimizar *prompts* para que a LLMs interpretem e gerem descrições semânticas e precisas destes diagramas; e o terceiro priorizou a construção de uma interface intuitiva e acessível para o envio dos vídeos e disponibilização das audiodescrições. Para materializar e integrar esses processos, foi implementada uma arquitetura cliente-servidor, justificada pela necessidade de processamento assíncrono do conteúdo em vídeos com durações variáveis.

Como pode ser visto nas Figuras 1a e 1b, o processamento é iniciado a partir do envio da videoaula pelo usuário ao *backend* (servidor), desenvolvido em Python, através da plataforma Web (cliente), criada com Next.js com TypeScript. O *backend* recebe o arquivo via Interface de Programação de Aplicações (API) Rest e orquestra a extração das cenas. Para otimizar a carga computacional e evitar a redundância de analisar todos os *frames* (quadros), o sistema emprega a biblioteca PySceneDetect para segmentar transições visuais e mudanças de tela, isolando apenas os quadros que apresentam conteúdos novos ou alterações estruturais. Logo após essa filtragem, os quadros relevantes são submetidos ao modelo de visão computacional YOLO.

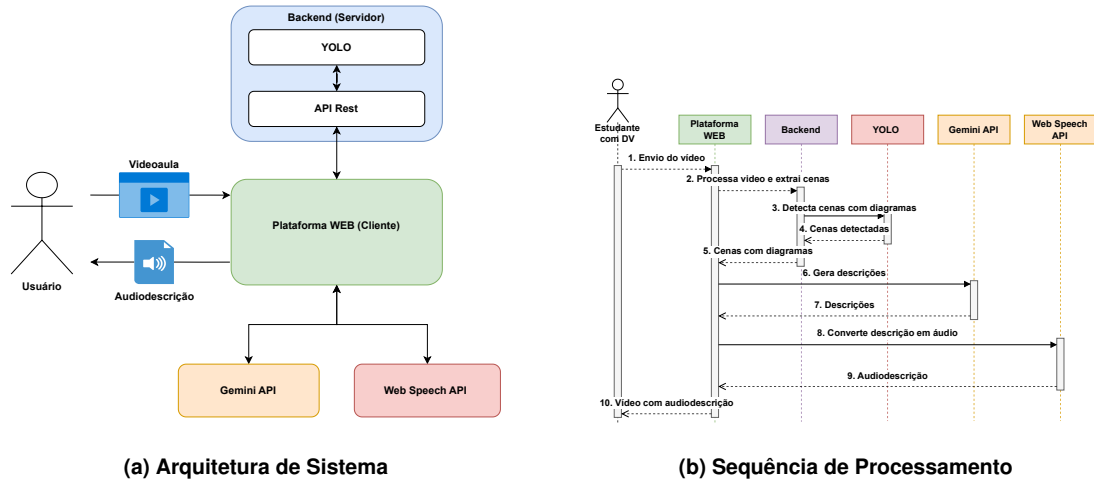


Figura 1. Diagramas de Arquitetura e Sequência do DIAGRAMLENS

Essa submissão atua como uma segunda camada de filtragem. Embora o PySceneDetect seja eficiente em identificar novas cenas, muitas dessas transições podem conter apenas *slides* textuais ou focar exclusivamente no professor. O papel do YOLO neste fluxo é analisar os quadros previamente extraídos para detectar a presença de diagramas.

Para viabilizar a identificação precisa, adotou-se a variação 11n do YOLO, que contém menos parâmetros e foi projetada para ser leve e rápida. Foi realizado um processo prévio de treinamento do modelo utilizando o *Lecture Video Visual Objects* (LVVO) [Biswas et al. 2025], um *dataset* estruturado para vídeos educacionais de disciplinas que utilizam representações gráficas (e.g., biologia e ciência da computação). No entanto, foi

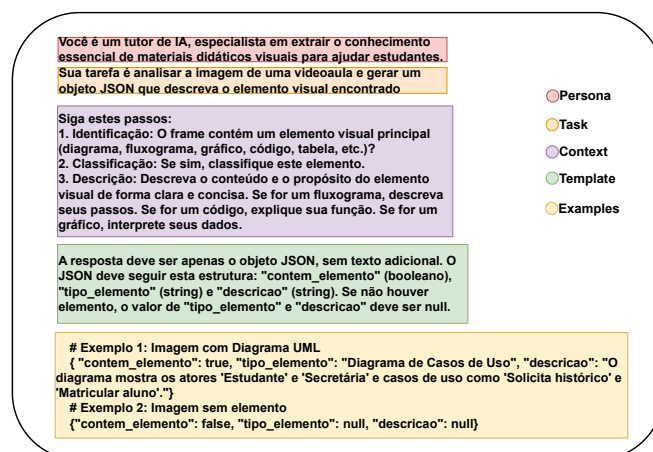


Figura 2. Prompt utilizado pelo Gemini 2.5 Flash-Lite

necessário adequar essa base ao DIAGRAMLENS, já que o *dataset* original não possuía as *bounding boxes* no formato exigido pelo YOLO 11n. Ao final desse refinamento, foi obtido um *dataset* customizado de 1.000 imagens, mapeadas em quatro classes: *table*, *chart-graph*, *photographic-image* e *visual-illustration* (diagramas). O treinamento foi realizado em novembro de 2025 por meio de *fine-tuning* [Gandhi and Gandhi 2025] ao longo de 20 épocas, com o *dataset* particionado em 70% para treino (700 amostras) e 30% para validação (300 amostras). A avaliação do modelo pós-treinamento revelou um desempenho equilibrado, alcançando Precisão de 72,2%, Recall de 71,6% e Mean Average Precision (mAP@50) de 78,2%, com um tempo de inferência de apenas 2,4 milissegundos por quadro.

Uma vez detectados os diagramas, o sistema avança para a interpretação e a geração de descrições semânticas, visando traduzir as representações visuais em narrativas textuais. Para essa tarefa, foi adotado o modelo Gemini 2.5 Flash-Lite, justificado pela disponibilidade de uma cota de uso gratuito de até 500 requisições diárias (limite que se mostrou adequado para a execução dos testes e para a validação do sistema no escopo desta pesquisa), além da facilidade de integração proporcionada por sua API.

A fim de garantir uma maior precisão na descrição, foram aplicadas técnicas de engenharia de *prompt* fundamentadas no catálogo de padrões de *prompt* proposto por White et al. [White et al. 2023]. Diferente do catálogo original, que foca na aplicação de padrões de forma isolada ou combinada para resolver problemas genéricos, esta pesquisa implementou um *prompt* (Figura 2) customizado segmentado em cinco blocos: Persona, Task, Context, Template e Examples. Este *prompt* estende os padrões do catálogo original, integrando-os a uma abordagem de few-shot prompting [Brown et al. 2020] para direcionar especificamente a interpretação de diagramas da área de computação pelo LLM, garantindo precisão e consistência nas descrições geradas. Por fim, com o material gerado pela LLM, o sistema realiza a conversão da descrição em áudio via Text-to-Speech (TTS), utilizando a Web Speech API e disponibilizando a audiodescrição para o usuário.

5. Apresentação da Ferramenta

A ferramenta DIAGRAMLENS consiste em uma aplicação web de audiodescrição automática que integra visão computacional, LLM e TTS. Para garantir a autonomia do

estudante desde o primeiro contato, o *design* da tela inicial (Figura 3) foi concebido de forma minimalista, exibindo apenas as informações essenciais para evitar a sobrecarga cognitiva de usuários que utilizam leitores de tela. A interação começa por meio de um *skip link* acessível por teclado, que direciona o estudante diretamente para o campo de envio da videoaula.

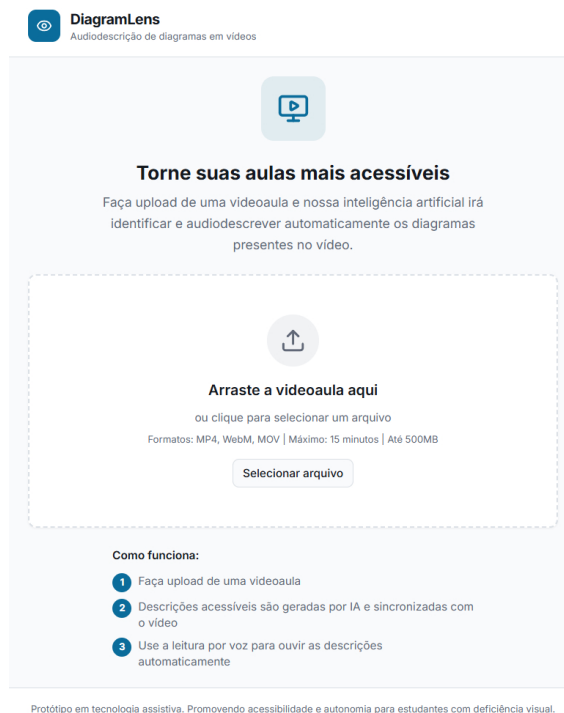


Figura 3. Tela inicial do DIAGRAMLENS

Com o envio do arquivo, inicia-se o fluxo de processamento (Figura 1b): envio do vídeo para o servidor, extração de cenas, identificação de diagramas, geração de descrições textuais por LLM e conversão em audiodescrição. Para mitigar qualquer incerteza e garantir *feedback* em tempo real durante esse período, o sistema fornece atualizações contínuas de status por meio de um componente com o atributo *aria-live*=“*polite*”. Dessa forma, os leitores de tela informam automaticamente sobre o andamento das etapas, como a quantidade de cenas detectadas, analisadas e descrições geradas, além do percentual geral de progresso, eliminando a necessidade de comandos manuais de verificação.

Ao final do processamento, os elementos detectados são armazenados como anotações vinculadas aos seus respectivos *timestamps* (tempos), garantindo a sincronização exata entre o conteúdo visual e a descrição. O estudante passa, então, a interagir com uma interface composta por um *player* de vídeo e uma *timeline* interativa (Figura 4), que atua na prática como um mapa sonoro do vídeo. Nessa *timeline*, cada elemento detectado representa um ponto clicável identificado via *aria-label* (no formato “Anotação em HH:MM: [tipo do elemento]”), o que possibilita saltos diretos e facilita a localização espacial do conteúdo pelo leitor de tela. Paralelamente, os elementos também são organizados em cartões de anotação que consolidam os metadados extraídos pela LLM, agrupando a miniatura, o tempo exato, o tipo de elemento figurado e a sua descrição

textual.

Além disso, para garantir que toda essa experiência ocorra sem dependência visual, a ferramenta foi construída com navegabilidade completa via teclado e compatibilidade com leitores de tela populares, como o NVDA e o JAWS. A interface faz uso frequente de atributos *aria-labels* explícitos em botões de ação, comunicando claramente a função do elemento antes de sua execução. Por fim, a ferramenta oferece modos de audiodescrição flexíveis para se adaptar ao ritmo de estudo do aluno: um modo automático, no qual as descrições acompanham o vídeo, e um modo sob demanda, focado na leitura individual de anotações específicas. Em ambos os modos, um controle de velocidade ajustável entre 0.5x e 2x possibilita que o estudante adapte a voz sintética à sua preferência auditiva de aprendizagem.

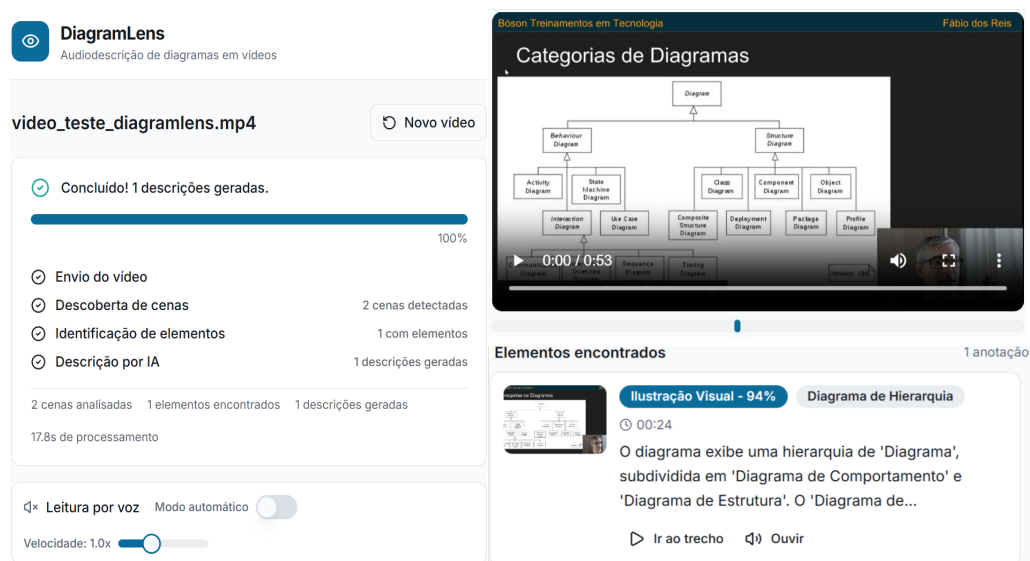


Figura 4. Tela de processamento de vídeo do DIAGRAMLENS

6. Avaliações e Resultados

Embora o desenvolvimento da ferramenta esteja apoiado em uma fundamentação sólida, fez-se necessário assegurar que a solução proposta efetivamente cumprisse seu papel de mediação pedagógica e de inclusão digital. Portanto, a avaliação deste trabalho foi estruturada em duas frentes distintas e complementares, abordando tanto a qualidade do conteúdo gerado quanto a acessibilidade da plataforma.

6.1. Avaliação com especialistas

A primeira frente de validação avaliou a qualidade das descrições narrativas geradas pela LLM, sendo conduzida por meio de um experimento envolvendo 11 especialistas nas áreas de Engenharia de Software, Computação e Educação, e Tecnologias Assistivas, todos com mais de três anos de experiência profissional. O objetivo principal foi avaliar as descrições com base em três dimensões: a **Precisão Técnica** das descrições elaboradas; a **Qualidade da Comunicação** para a viabilidade da construção mental do diagrama a partir do texto; e a **Validação de Propósito**, que verifica a eficácia da solução em promover a autonomia de estudantes com DV no consumo de videoaulas. Nesta etapa, não houve a

participação direta de pessoas com DV. O perfil de cada especialista pode ser consultado nos artefatos complementares desta pesquisa.

Para a coleta de dados foi utilizado um formulário eletrônico no qual os especialistas foram expostos, de forma pareada, à imagem original do diagrama e à respectiva descrição textual gerada pela LLM. Os elementos visuais utilizados no experimento foram extraídos de 5 diagramas distintos (Classes, Casos de Uso, Sequência, Entidade-Relacionamento e Fluxograma) provenientes de videoaulas públicas de Engenharia de Software, assegurando uma reprodução aproximada do contexto real de aplicação da plataforma. Diante de cada cenário, os participantes responderam a 6 questões estruturadas em uma escala *Likert* de 5 pontos (variando de 1: “Discordo Totalmente” a 5: “Concordo Totalmente”), projetadas especificamente para mensurar as três dimensões anteriormente citadas (Quadro 2).

Quadro 2. Questões do formulário para cada diagrama

Dimensão	Questão	Descrição
Precisão Técnica	Q1	A descrição identifica corretamente o tipo do diagrama mostrado na imagem
	Q2	A descrição identifica corretamente todos os elementos e relacionamentos do diagrama
Qualidade da Comunicação	Q3	A ordem dos elementos na descrição permite uma construção mental clara do diagrama sem ver a imagem
	Q4	A densidade de informações é equilibrada, evitando tanto a omissão de dados importantes quanto o excesso de detalhes irrelevantes
Validação de Propósito	Q5	A descrição promove a autonomia de um estudante de computação para compreender o diagrama
	Q6	Uma ferramenta capaz de gerar estas descrições automaticamente em videoaulas beneficia significativamente o aprendizado de estudantes de computação

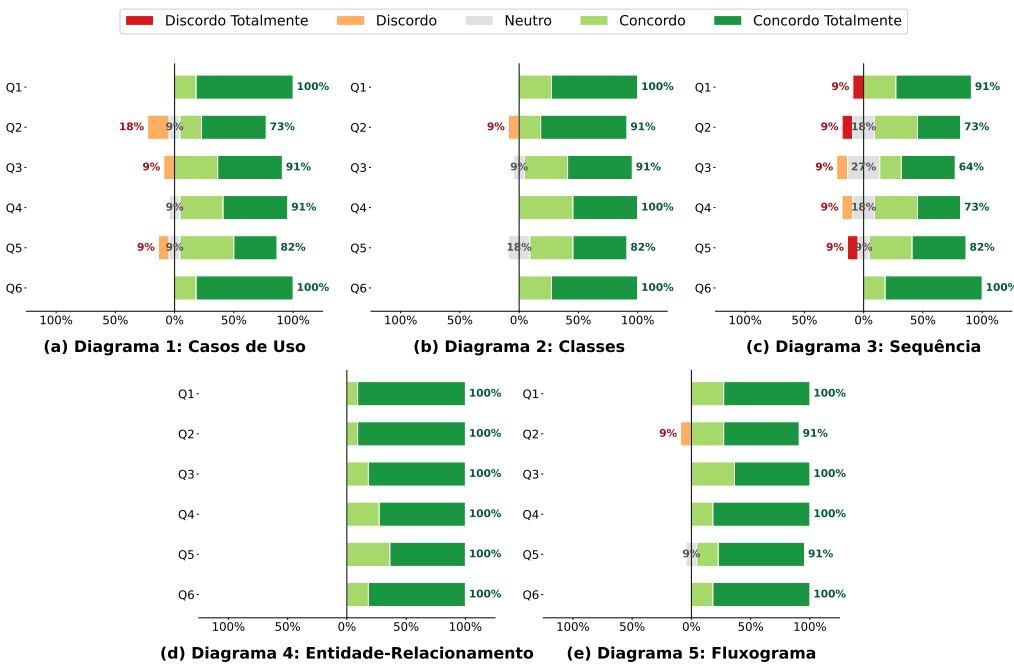


Figura 5. Resultados da avaliação das descrições geradas pela LLM

A análise dos dados coletados, ilustrada na Figura 5 (subdividida em 5a a 5e para representar individualmente os cinco diagramas), revela o comportamento real diante de diferentes elementos visuais. De modo geral, os resultados atestam a eficácia do *Prompt+LLM* utilizado no DIAGRAMLENS em gerar descrições com base em diagramas identificados em videoaulas, mas evidenciam variações importantes dependendo do tipo de diagrama analisado.

Primeiramente foi avaliado o consenso entre os 11 especialistas, foi aplicado o Coeficiente de Concordância W de Kendall [Kendall and Gibbons 1962] sobre a matriz de avaliações. Os resultados apontam um coeficiente geral $W = 0.22$ com alta significância estatística ($\chi^2 = 70.17$; $df = 29$; $p < 0.001$), rejeitando a hipótese de que o alinhamento de opiniões tenha ocorrido ao acaso. O coeficiente W, embora estatisticamente significativo, aponta para uma concordância de fraca a moderada entre os especialistas, o que reflete a variabilidade de opiniões. Esse comportamento é em parte decorrente do ‘efeito teto’, visto que 92.12% de todas as respostas concentraram-se nos níveis ‘Concordo’ (4) e ‘Concordo Totalmente’ (5), gerando um elevado número de empates amostrais que estabilizam as variâncias de postos. Complementarmente, o instrumento mostrou uma alta consistência interna e confiabilidade [Freitas and Rodrigues 2005], com um Alfa de Cronbach de $\alpha = 0.861$.

Avaliando cada dimensão individualmente, para a **Precisão Técnica** (Q1 e Q2), a identificação do tipo de diagrama (Q1) obteve 100% de respostas positivas nas Figuras 5a, 5b, 5d, 5e, sofrendo uma ligeira queda para 91% na Figura 5c. Contudo, a assertividade em identificar elementos e relacionamentos específicos (Q2) oscilou. Enquanto o Diagrama 4 (Figura 5d) alcançou 100% de exatidão, este valor recuou para 73% de aprovação nos Diagramas 1 e 3 (Figuras 5a e 5c), havendo inclusive 18% de discordância no Diagrama 1. Isso se deu por uma falha do LLM ao identificar um relacionamento no Diagrama 1.

Avaliando a **Qualidade da Comunicação** (Q3 e Q4), foi observado que o equilíbrio na clareza na construção mental (Q3) e na densidade textual (Q4) tiveram desempenhos consistentes na maioria dos casos. As Figuras 5d e 5e mostram 100% de concordância de que a descrição possibilitou uma clara construção mental. Em contrapartida, o Diagrama 3 (Figura 1c) revelou ser o cenário mais crítico da avaliação: a concordância sobre a Q3 caiu para 64%, somada a uma considerável taxa de neutralidade (27%) e de discordância (9%). Esse resultado justifica-se pela estruturação linear da descrição gerada pela LLM, que omitiu a direcionalidade das mensagens e os tipos de retorno.

Por fim, os dados referentes à **Validação de Propósito** (Q5 e Q6) indicam a utilidade social do projeto. O impacto no benefício para o aprendizado (Q6) obteve 100% de respostas positivas em todos os 5 diagramas testados. No quesito de fomento à autonomia (Q5), a avaliação também foi positiva, variando de 82% (Figuras 5a, 5b e 5c) a 100% (Figura 5d). Ainda assim, as taxas de neutralidade e discordância observadas nesse quesito sinalizam que, para alguns especialistas, a descrição textual gerada apoia, mas pode não substituir completamente a dependência de suporte externo em determinados tipos de diagramas.

Portanto, os resultados validados pelo grupo de especialistas confirmam que a descrição automatizada de diagramas geradas por IA é uma estratégia viável para a in-

clusão de estudantes com DV. A aprovação unânime em relação ao benefício educacional atesta que a ferramenta supre uma lacuna de acessibilidade. O que leva a concluir que, em virtude das variações de precisão e comunicação observadas nos Diagramas 1 e 3, e embora o ganho de autonomia seja evidente, futuras otimizações na arquitetura de geração de descrições via LLM devem concentrar-se no aprimoramento da identificação de estruturas e seus elementos, bem como na transmissão dos elementos visuais com maior clareza e fidelidade.

6.2. Avaliação de acessibilidade

A segunda frente consistiu em uma auditoria técnica de acessibilidade na interface da plataforma, para garantir que a infraestrutura web não apresentasse barreiras aos usuários com DV que utilizam tecnologias assistivas, como leitores de tela. Este procedimento utilizou três validadores automáticos amplamente reconhecidos no meio acadêmico e técnico. O Lighthouse foi empregado para diagnosticar a semântica estrutural do HTML e os nomes acessíveis, enquanto o WAVE possibilitou a verificação de contrastes e rótulos funcionais. Por fim, para assegurar a conformidade com as diretrizes e padrões nacionais e internacionais, foi utilizado o AccessMonitor. Essa abordagem multidimensional possibilitou correlacionar a qualidade do conteúdo gerado com a operabilidade da interface, garantindo uma experiência de uso inclusiva de ponta a ponta.



Figura 6. Resultados dos testes de acessibilidade para as interfaces do DIAGRAMLENS

A avaliação contemplou todas as interfaces do sistema: a Tela Inicial (responsável pelo primeiro contato e navegação do usuário) e a Tela de Processamento de Vídeo (ambiente central de interação e consumo do conteúdo audiodescrito). As Figuras 6a e 6b ilustram os relatórios gerados pelas ferramentas de auditoria em ambas as telas.

Os dados extraídos dos testes mostram a conformidade da plataforma com as diretrizes de acessibilidade. Na Tela Inicial, o validador Lighthouse registrou a aprovação em todos os 25 critérios avaliados, indicando que as práticas de marcação semântica do HTML e a estruturação do *Document Object Model* (DOM) foram aplicadas adequadamente. Já na Tela de Processamento de Vídeo, a ferramenta atestou a conformidade em 24 dos 25 critérios avaliados, um índice altamente satisfatório que comprova a manutenção do padrão de acessibilidade exigido para a navegação inclusiva.

No validador WAVE, os resultados mantiveram a consistência técnica: não foram detectados erros de estrutura ou de contraste em nenhuma das telas analisadas. A ferramenta confirmou a correta aplicação de dezenas de atributos *Accessible Rich Internet*

Applications (ARIA) e elementos estruturais de navegação, apontando apenas dois alertas na tela de processamento, que não comprometeram a acessibilidade geral.

Corroborando as métricas obtidas, o teste realizado no AccessMonitor confirmou o alto grau de aderência do código às diretrizes de acessibilidade. Em uma escala de 0 a 10, a plataforma alcançou índices satisfatórios. A Tela Inicial obteve a nota 9,4, validando com sucesso 29 critérios técnicos encontrados, enquanto a Tela de Processamento de Vídeo apresentou desempenho ainda superior, atingindo a nota 9,6, com 30 critérios técnicos encontrados. As ressalvas apontadas pelo validador foram mínimas e não inviabilizam a operabilidade do sistema: a interface inicial apresentou duas infrações pontuais relacionadas a um elemento focável e à divergência em um nome acessível, e a interface de processamento registrou apenas um erro semântico referente ao uso de propriedades ARIA não permitidas em determinadas *tags*.

7. Conclusões e Discussões

O DIAGRAMLENS demonstrou ser uma ferramenta promissora e viável para apoiar a inclusão e a autonomia de estudantes com DV. Os resultados obtidos nas duas frentes de avaliação atestam o cumprimento do propósito da ferramenta. A avaliação de acessibilidade confirmou que a plataforma possui uma estrutura em conformidade com diretrizes de acessibilidade, alcançando índices satisfatórios em validadores como Lighthouse, WAVE e AccessMonitor. Paralelamente, a avaliação por especialistas validou aspectos de qualidade e viabilidade das descrições automatizadas de diagramas em cenários reais de ensino, com aprovação unânime quanto ao benefício educacional.

Apesar dos resultados promissores, a solução apresenta algumas limitações. A principal delas reside na carência de validação direta com usuários finais para atestar a usabilidade prática e o impacto real da ferramenta no cotidiano das pessoas com DV. No âmbito técnico, a dependência de uma API de LLM externa introduz restrições relacionadas ao limite de *tokens* e de uso, o que pode restringir a geração de descrições. Além disso, o fluxo operacional é diretamente afetado pela duração das videoaulas submetidas à plataforma. Como a extração de cenas e as inferências ocorrem de forma assíncrona, o processamento torna-se progressivamente mais demorado conforme a extensão do vídeo aumenta, exigindo um tempo de espera maior por parte do estudante. Também reconhece-se que a avaliação de acessibilidade, embora útil, não garante a qualidade da experiência real de uso por pessoas com DV, a qual é subjetiva e depende de fatores como ritmo de leitura de tela e carga cognitiva de navegação.

Diante dessas restrições e das oportunidades de melhoria evidenciadas durante o estudo, trabalhos futuros devem focar em otimizar a infraestrutura tecnológica e expandir o rigor analítico da aplicação. Em primeiro lugar, estão sendo estruturados testes de usabilidade e avaliações diretas com pessoas com DV para colher *feedbacks* reais, garantindo que a interface e as descrições atendam às suas necessidades cognitivas e de navegação. No aspecto estrutural, serão desenvolvidas estratégias para um processamento de vídeo mais eficiente, reduzindo a carga computacional e acelerando a disponibilização final do áudio. O modelo de visão computacional também será expandido e treinado para identificar, categorizar e interpretar diagramas de outros domínios, ampliando a versatilidade da ferramenta. Por fim, também será refinado o *prompt* utilizado pela LLM, visando elevar a precisão em transmitir com maior fidelidade os elementos visuais gerados na descrição.

Declaração sobre uso de Inteligência Artificial

Foram utilizadas ferramentas de Inteligência Artificial (IA) para a revisão do conteúdo.

Agradecimentos

Agradecemos o apoio do Programa Interno de Auxílio Financeiro aos Programas de Pós-Graduação Stricto Sensu (AUXPPG) da UEFS e/ou o apoio do Programa de Apoio à Pós-Graduação (PROAP) da CAPES.

Disponibilidade de Artefatos

O código-fonte da ferramenta DIAGRAMLENS está disponível em <https://github.com/almerindopaixao/diagramlens/tree/v1.0.0>. Nesse repositório, é possível encontrar os códigos do *backend* e do *frontend* da aplicação, bem como as dependências e bibliotecas utilizadas, juntamente com suas respectivas versões.

Além disso, os dados complementares da pesquisa, incluindo o documento com o perfil dos especialistas que participaram do processo de avaliação, estão disponíveis em <https://doi.org/10.5281/zenodo.21461634>.

Referências

- Barrére, E. (2014). Videoaulas: aspectos técnicos, pedagógicos, aplicações e bricolagem. *Jornada de Atualização em Informática na Educação*, 3(1).
- Biswas, D., Shah, S., and Subhlok, J. (2025). Lecture video visual objects (lvvo) dataset: A benchmark for visual object detection in educational videos.
- Brown, T. B. et al. (2020). Language models are few-shot learners.
- Franck, K. M., Pereira, R. F., and Filho, J. V. D. (2021). Diagrama Entidade-Relacionamento: uma ferramenta para modelagem de dados conceituais em Engenharia de Software. *Research, Society and Development*, 10(8).
- Freitas, A. L. and Rodrigues, S. (2005). A avaliação da confiabilidade de questionários: uma análise utilizando o coeficiente alfa de cronbach.
- Gandhi, V. and Gandhi, S. (2025). Fine-tuning without forgetting: Adaptation of yolov8 preserves coco performance.
- Gavric, A., Bork, D., and Proper, H. A. (2024). How Does UML Look and Sound? Using AI to Interpret UML Diagrams Through Multimodal Evidence. In *Advances in Conceptual Modeling: ER 2024 Workshops, AISA, CMLS, EmpER, QUAMES, JUSMOD, LLM4Modeling, Pittsburgh, PA, USA, October 28–31, 2024, Proceedings*, Berlin, Heidelberg. Springer-Verlag.
- Inep (2026). Resumo técnico do censo da educação superior 2024. Technical report, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), Brasília. Acesso em: 11 mai. 2026.
- Junior, R. R. and Pereira, C. (2025). Diagramando para todos os “olhares”: Acessibilidade digital para ensino e prática de diagrama entidade-relacionamento. In *Anais do XXXVII Simpósio Brasileiro de Informática na Educação*, pages 153–167, Porto Alegre, RS, Brasil. SBC.

- Kendall, M. G. and Gibbons, J. D. (1962). *Rank correlation methods*. Griffin London.
- Kennel, A. R. (1996). Audiograf: a diagram-reader for the blind. In *Proceedings of the Second Annual ACM Conference on Assistive Technologies*, Assets '96, page 51–56, New York, NY, USA. Association for Computing Machinery.
- Klein, A., Camargo, H. C., Caetano, B. P., and Paula, M. M. V. d. (2025). VisER: uma Proposta de Modelagem Inclusiva para Descrição do Modelo Entidade-Relacionamento. In *Workshop sobre Educação em Computação (WEI)*, pages 1313–1323. SBC.
- Kour, M., Hains, G., and Saber, M. (2025). Diagram Analysis for Visually Impaired Readers. Technical report, universite paris est creteil val de marne.
- Lima, P. C. and Fonseca, L. P. (2016). Recursos digitais adaptados ou construídos para o ensino a distância de deficientes visuais. In *Anais do XIII Congresso Brasileiro de Ensino Superior a Distância*, volume 1, pages 607–617, São João del-Rei. Congresso Brasileiro de Ensino Superior a Distância.
- Luque, L., Santos, C. L. F., Cruz, D. O., Brandão, L. d. O., and Brandão, A. A. F. (2016). Model2gether: uma ferramenta de apoio ao ensino e a aprendizagem de modelos por cegos.
- Magalhães, R. L. and Neto, M. M. (2010). Aprender: Ferramenta de apoio à construção de diagrama entidade relacionamento para deficientes visuais. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 1.
- Marco, L. R. D., Gasparini, I., Alves, B. N., Sudoski, C. H., Vasconcellos, D. M., da Silveira, E. C., Vossen, L. V., and Santos, M. T. S. (2025). Análise da diversidade e evasão em cursos de computação no brasil. In *Anais do Congresso da Universidade do Estado de Santa Catarina (UDESC)*, pages 1–3, Santa Catarina, Brasil. UDESC.
- Moore, M. (2002). Teoria da distância transacional. *Revista Brasileira de Aprendizagem Aberta e a Distância*, 1.
- Pavão, A. C. O. and Pavão, S. M. d. O. (2020). Audiodescrição na intervenção pedagógica das dificuldades de aprendizagem. *Educação e Fronteiras*, 10(28):34–45.
- Rocha, T. B. and Miranda, T. G. (2009). Acesso e permanência do aluno com deficiência na instituição de ensino superior. *Revista Educação Especial*, 22(34):197–211.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2):257–285.
- Verde, A., Alves, L., Fraga, L., and Soares, L. (2023). Py2uml: a tool for visually impaired students to build uml diagrams from python coding. In *Proceedings of the XXXVII Brazilian Symposium on Software Engineering*, SBES '23, page 491–496, New York, NY, USA. Association for Computing Machinery.
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., and Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with chatgpt.