WebMedia 2025

31° SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB

MINICURSOS DO XXXI SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB

10 A 14 DE NOVEMBRO DE 2025
RIO DE JANEIRO, RJ, BRASIL

ORGANIZADORES

Windson Viana de Carvalho (UFC), Rudinei Goularte (ICMC/USP), Roberto Willrich (UFSC), Eduardo Barrére (UFJF), Sergio Colcher (PUC-Rio), Julio Cesar Duarte (IME), Álvaro da Veiga (PUC-Rio)

REALIZAÇÃO SBC Sociedade Brasileira de Computação

ORGANIZAÇÃO PUC RIO, IME

PATROCÍNIO Fundação Padre Leonel Franca, CNPq, AWS, ifood, ana health, btg pactual, nic.br, cgi.br

XXXI Simpósio Brasileiro de Sistemas Multimídia e Web

De 10 a 14 de novembro de 2025

Rio de Janeiro, Brasil

LIVRO DE MINICURSOS

Organizadores

Windson Viana de Carvalho (UFC)
Rudinei Goularte (USP)
Roberto Willrich (UFSC)
Eduardo Barrére (UFJF)
Sergio Colcher (PUC-Rio)
Julio Cesar Duarte (IME)
Álvaro da Veiga (PUC-Rio)

Realização

Sociedade Brasileira de Computação – SBC

Organização

Pontifícia Universidade Católica do Rio de Janeiro – PUC-Rio
Instituto Militar de Engenharia – IME

Dados Internacionais de Catalogação na Publicação (CIP)

S612 Simpósio Brasileiro de Sistemas Multimídia e Web (31. : 10 – 14 novembro 2025 : Rio de Janeiro)

Minicursos do WebMedia 2025 [recurso eletrônico] / organização: Windson Viana de Carvalho ... [et al.]. – Dados eletrônicos. – Porto Alegre: Sociedade Brasileira de Computação, 2026.

118 p. : il. : PDF ; 11,8 MB

Modo de acesso: World Wide Web.
ISBN 978-85-7669-679-7 (e-book)

1. Computação – Brasil – Evento. 2. Sistemas de informação. 3. Inteligência artificial. I. Carvalho, Windson Viana de. II. Goularte, Rudinei. III. Willrich, Roberto. IV. Barrére, Eduardo. V. Colcher, Sergio. VI. Duarte, Julio Cesar. VII. Veiga, Álvaro da. VIII. Sociedade Brasileira de Computação. IX. Título.

CDU 004(063)

Ficha catalográfica elaborada por Annie Casali – CRB-10/2339

Biblioteca Digital da SBC – SBC OpenLib

Sociedade Brasileira de Computação
Av. Bento Gonçalves, 9500
Setor 4 | Prédio 43.412 | Sala 219 | Bairro
Agronomia Caixa Postal 15012 | CEP 91501-970
Porto Alegre - RS
(51) 99252-6018
sbc@sbc.org.br

Prefácio

Tradicionalmente, o Simpósio Brasileiro de Sistemas Multimídia e Web proporciona aos participantes a oportunidade de explorar e aprender sobre temas relevantes da área na trilha Minicursos e Tutoriais. Em sua 31ª edição, realizada na cidade do Rio de Janeiro, entre 11 e 14 de Novembro de 2025, o evento contou com 3 tutoriais e 5 minicursos, escolhidos em um processo de seleção das propostas enviadas em resposta a uma chamada pública de alcance nacional.

Assim como nas edições anteriores do WebMedia, os ministrantes dos minicursos foram convidados a produzir um texto correspondente ao conteúdo apresentado, para publicação na forma de capítulo de livro. Assim, este volume reúne, a posteriori, o material referente a três minicursos.

O Capítulo 1, intitulado *Grandes Modelos de Linguagem Multimodais (MLLMs): Da Teoria à Prática*, aborda o tema dos atuais MMLs sob a perspectiva da multimodalidade, combinando a capacidade de compreensão e geração de linguagem natural com percepções extraídas de modalidades como imagem e áudio. O capítulo explora técnicas práticas de pré-processamento, engenharia de *prompts* e construção de *pipelines*multimodais, assim como aponta tendências promissoras.

O Capítulo 2, *Pesquisa com Seres Humanos em Computação: Aspectos Éticos, Regulamentação e Cenários Didáticos*, apresenta fundamentos históricos e normativos da ética em pesquisas na área de Computação que envolvam seres humanos. Sob um panorama didático, são analisados cenários hipotéticos que ilustram situações recorrentes de pesquisa em Computação evidenciando como princípios éticos se aplicam e a necessidade de apreciação ética institucional.

O Capítulo 3, *UX Research Envolvendo Pessoas Com Deficiência: Um Caminho Para Uma Webmídia Inclusiva*, aborda a questão da acessibilidade digital no Brasil, discutindo a importância de realizar pesquisas focando na Experiência do Usuário (UX Research) que contem com a participação direta de pessoas com deficiência. Além disso, aponta as diretrizes técnicas e os desafios práticos e éticos envolvidos no design de sistemas de Web e multimídia que sejam democráticos e acessíveis.

Agradecemos aos organizadores do evento pela oportunidade e aos revisores pelo trabalho na avaliação e na seleção das propostas aceitas. Temos certeza de que este livro será útil para todas as pessoas interessadas nos temas abordados.

Rio de Janeiro, novembro de 2025.

Windson Viana de Carvalho (UFC)
Rudinei Goularte (ICMC/USP)
Coordenadores de Minicursos e Tutorais

XXXI Simpósio Brasileiro de Sistemas Multimídia e Web

10 a 14 de Novembro de 2025
Rio de Janeiro, Brasil

Coordenação Geral

Sergio Colcher (PUC-Rio) – *Coordenador Geral*
Julio Cesar Duarte (IME) – *Coordenador Geral*
Álvaro da Veiga (PUC-Rio) – *Coordenador Geral*
Carlos de Salles Soares Neto (UFMA) – *Coordenador do Comitê de Programa*
Alessandra Alaniz Macedo (FFCLRP/USP) – *Coordenadora do Comitê de Programa*

Coordenação de Minicursos e Tutoriais

Windson Viana de Carvalho (UFC)
Rudinei Goularte (USP)

Coordenação de Palestras e Painéis

Jonice de Oliveira Sampaio (UFRJ)
Jussara Marques de Almeida (UFMG)
Alan L.V. Guedes (Reading, UK)

Coordenadores do VII Concurso de Teses e Dissertações (CTD)

Julio Cesar S. Reis (UFV)
Marcelo Manzato (USP)

Coordenadores do V Concurso de Trabalhos de Iniciação Científica (CTIC)

Humberto Marques (PUC Minas)
Daniel de Sousa Moraes (PUC-Rio)

Coordenadores do XXIV Workshop de Ferramentas e Aplicações (WFA)

Carlos André G. Ferraz (UFPE)
Carlos Henrique G. Ferreira (UFOP)

Coordenação do IV WebMedia for Everyone

Kamila Rios da Hora Rodrigues (USP)
Leo Sampaio Ferraz Ribeiro (USP)
Bruna Carolina Rodrigues da Cunha (USP)

Workshop Brasileiro de Sistemas Web3 (BrWeb3)

Antonio José Grandson Busson (BTG Pactual)
Sergio Colcher (PUC-Rio)
Julio Cesar Duarte (IME)

Coordenação do Workshop Futuro da TV Digital Interativa

Joel dos Santos (CEFET/RJ)
Alan L.V. Guedes (Reading, UK)

Coordenadores de Publicação

Roberto Willrich (UFSC)
Eduardo Barrére (UFJF)

Coordenação de Patrocínios e Contatos Institucionais

Adriano Cesar Machado Pereira (UFMG)
Leonardo Chaves Dutra da Rocha (UFSJ)

Coordenação do Prêmio LF

Manoel C. Marques Neto (IFBA)

Coordenação da Comissão Especial de Sistemas Multimídia e Web

Carlos André G. Ferraz (UFPE) – *Coordenador*
Débora C. Muchaluat Saade (UFF) – *Vice-Coordenadora*

Comitê Gestor

Adriano César Machado Pereira (UFMG)
Alessandra Macedo (USP)
Carlos de Salles Soares Neto (UFMA)
Carlos Pernisa (UFJF)
Celso Alberto Saibel Santos (UFES)
Cezar Teixeira (UFSCar)
Diego Roberto (UFSJ)
Eduardo Barrére (UFJF)
Kamila Rios Rodrigues (ICMC/USP)
Leonardo Rocha (UFSJ)
Joel dos Santos (CEFET/RJ)
Jorge Luis Victória Barbosa (UNISINOS)
José Valdeni de Lima (UFRGS)
Maria da Graça Campos Pimentel (USP)
Marcelo Moreno (UFJF)
Paulo Vitor Borges (PUC-Rio)
Roberto Willrich (UFSC)
Rudinei Goularte (USP)
Thiago Henrique Silva (UTFPR)

Sociedade Brasileira de Computação (SBC)

Presidência

Thais Vasconcelos Batista (UFRN) – *Presidente*
Cristiano Maciel (UFMT) – *Vice-Presidente*

Diretoria

Alírio Santos Sá (UFBA)
André Luís de Medeiros Santos (UFPE)
Carlos Eduardo Ferreira (USP)
Claudia Lage Rebello da Motta (UFRJ)
Denis Lima do Rosário (UFPA)
Eunice Pereira dos Santos Nunes (UFMT)
Flávia Maria Santoro (INTELLI)
Francisco Dantas Medeiros Neto (UERN)
José Viterbo Filho (UFF)
Leila Ribeiro (UFRGS)
Lisandro Zambenedetti Granville (UFRGS)
Marcelo Antonio Marotta (UNB)
Michelle Silva Wangham (UNIVALI)
Renata de Matos Galante (UFGRS)
Rodrigo Silva Duran (IFB)
Ronaldo Alves Ferreira (UFMS)

Contato

Av. Bento Gonçalves, 9500
Setor 4 - Prédio 43.412 - Sala 219
Bairro Agronomia
91.509-900 – Porto Alegre RS
CNPJ: 29.532.264/0001-78
http://www.sbc.org.br

Sumário

Grandes Modelos de Linguagem Multimodais (MLLMs): Da Teoria à Prática

Neemias da Silva, Júlio C. W. Scholz, John Harrison, Marina Borges, Paulo Ávila, Frances A Santos, Myriam Delgado, Rodrigo Minetto, Thiago H Silva

Abstract

Multimodal Large Language Models (MLLMs) combine the natural language understanding and generation capabilities of LLMs with perception skills in modalities such as image and audio, representing a key advancement in contemporary AI. This chapter presents the main fundamentals of MLLMs and emblematic models. Practical techniques for preprocessing, prompt engineering, and building multimodal pipelines with LangChain and LangGraph are also explored. For further practical study, supplementary material is publicly available online: `https://github.com/neemiasbsilva/MLLMs-Teoria-e-Pratica`*. Finally, the chapter discusses the challenges and highlights promising trends.*

Resumo

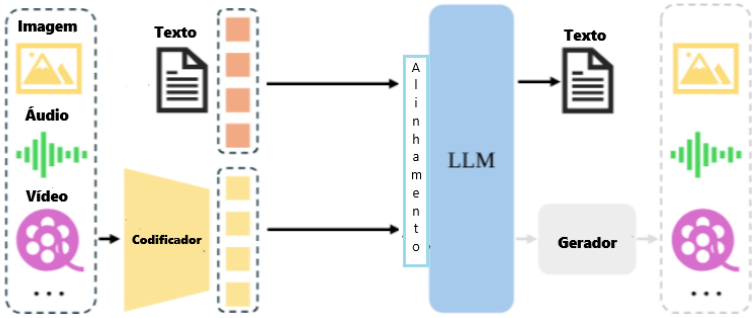
Os Grandes Modelos de Linguagem Multimodais (MLLMs do inglês Multimodal large language models) combinam a capacidade de compreensão e geração de linguagem natural dos LLMs com habilidades de percepção em modalidades como imagem e áudio, constituindo um avanço central na IA contemporânea. O capítulo apresenta os principais fundamentos dos MLLMs e modelos emblemáticos. Também são exploradas técnicas práticas de pré-processamento, engenharia de prompts e construção de pipelines multimodais com LangChain e LangGraph. Para aprofundamento prático, material complementar está disponível publicamente na web: `https://github.com/neemiasbsilva/MLLMs-Teoria-e-Pratica`*. Por fim, o capítulo discute desafios e aponta tendências promissoras.*

1.1. Introdução e Contextualização

Os Grandes Modelos de Linguagem Multimodais (MLLMs do inglês *Multimodal large language models*) representam um dos avanços mais significativos na interseção entre inteligência artificial, processamento de linguagem natural e visão computacional

[Caffagni et al. 2024, Wu et al. 2023, Yin et al. 2024]. Diferentemente de seus predecessores, os Grandes Modelos de Linguagem (LLMs do inglês *Large language models*) que operavam predominantemente com texto - os MLLMs possuem a capacidade notável de processar, interpretar e gerar conteúdo a partir de múltiplas modalidades, como texto, imagem, áudio e, em alguns casos, vídeo.

A Figura 1.1 traz uma visão geral da arquitetura de um MLLM. Os dados de entrada em múltiplas modalidades são primeiro codificados por modelos especializados e, assim como texto, são transformados em representações vetoriais chamadas *embeddings*. Um componente fundamental de um MLLM é o módulo de alinhamento, que adapta os *embeddings* ao formato compreensível pelo LLM (núcleo em azul formado em geral por um *decoder* de *Transformer* - o mesmo componente arquitetural que impulsiona LLMs de texto puro como o GPT-4 e o LLaMA). Este então processa os dados em conjunto, relacionando texto e outras modalidades. Assim, um MLLM consegue descrever, responder e raciocinar sobre diferentes tipos de informação. Na saída, há modelos que geram apenas texto e outros que, a partir de um mesmo processo de raciocínio interno, utilizam um gerador para converter *tokens* gerados pelo LLM em múltiplas modalidades.



Os MLLMs representam, portanto, um aprimoramento dos modelos de linguagem e visão - VLLMs do inglês *Vision-language large models* - que trabalham apenas com texto e imagem. Nesses modelos, tipicamente, um codificador de visão projeta as imagens em um espaço latente e um adaptador de *prompt* alinha essa representação com a entrada do LLM.

1.1.1. Relevância dos MLLMs na Web e Mídias Digitais

A *Web* moderna é um ecossistema essencialmente multimodal. Por exemplo, plataformas de *e-commerce* dependem da combinação de imagens de produtos com descrições textuais, redes sociais como *Instagram* e *TikTok* são construídas sobre a interação entre vídeo, áudio, imagem e texto, e até mesmo a busca na *web* está evoluindo para permitir consultas por imagem. Os MLLMs surgem como a ferramenta ideal para navegação, organização, moderação e criação de conteúdo para este ambiente complexo. Eles permitem diferentes recursos.

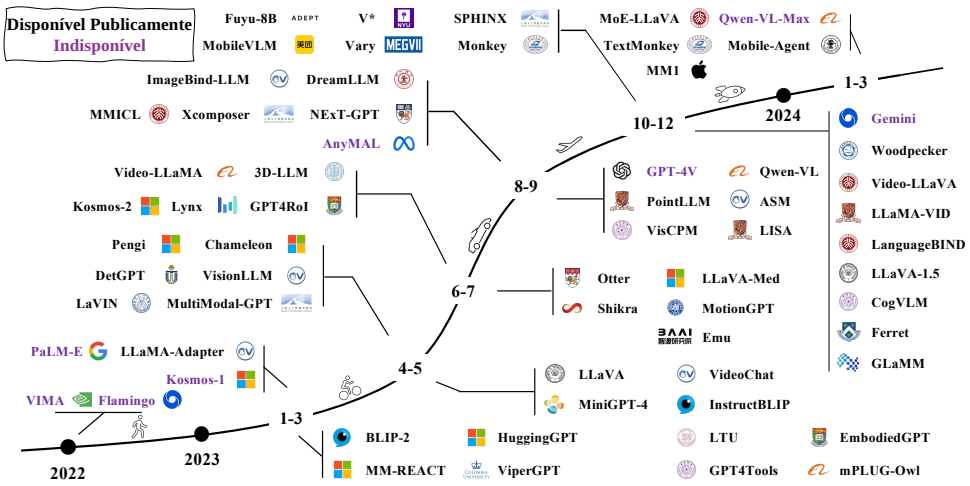
Busca e recuperação semântica: Encontrar produtos, notícias ou conteúdo multimodal com base em consultas complexas que envolvem tanto texto quanto imagem.

Acessibilidade: Gerar descrições textuais automáticas para imagens (*alt text*), tornando a web mais acessível para usuários com limitação visual.

Criação de conteúdo: Auxiliar na criação de conteúdos variados, como gerar legendas para *posts* em redes sociais, criar ilustrações a partir de rascunhos textuais ou mesmo produzir vídeos curtos baseados em um roteiro.

Moderação de conteúdo: Identificar conteúdo inadequado de forma mais contextual, analisando simultaneamente imagens e o texto associado, o que é mais eficaz do que analisar cada modalidade isoladamente.

A Figura 1.2 ilustra a rápida evolução e proliferação de MLLMs nos últimos anos, destacando a aceleração no desenvolvimento de modelos tanto proprietários quanto de código aberto, um testemunho do intenso interesse e investimento nesta área.



1.1.2. Desafios e Oportunidades no Processamento Conjunto das Multimodalidades

A motivação para o desenvolvimento e o estudo dos MLLMs reside na combinação de desafios complexos e oportunidades transformadoras decorrentes do processamento conjunto de diferentes modalidades. Os desafios são significativos [Liu et al. 2025, Kalai et al. 2025]. Diferentes modalidades existem em espaços de representação fundamentalmente distintos, por exemplo, *pixels* para imagens e *tokens* para texto. O principal obstáculo é, portanto, a criação de uma "ponte" semântica entre esses espaços, um processo conhecido como alinhamento de *embeddings*. Outros desafios críticos incluem a dessincronização de modalidades, onde o texto associado a uma imagem pode ser ambíguo, enganoso ou descrever apenas parcialmente o conteúdo visual; as alucinações multimodais, onde os MLLMs podem inventar detalhes visuais ou atributos textuais não presentes nas entradas, gerando informações incorretas com alta confiança; o custo com-

putacional elevado do treinamento e inferência com dados de alta dimensionalidade como imagens, levantando questões sobre eficiência energética e acessibilidade no desenvolvimento e no uso desses modelos; e as questões de viés e justiça, já que os MLLMs podem herdar e até amplificar vieses presentes nos dados de treinamento multimodais, levando a estereótipos prejudiciais em suas saídas.

No entanto, apesar dos desafios, os MLLMs atuais possibilitam várias aplicações em diversas áreas [Kuang et al. 2025, da Silva et al. 2025]. A capacidade de "raciocinar" sobre o mundo de forma integrada, tal como os humanos fazem naturalmente, abre um leque de aplicações anteriormente difíceis (ou não viáveis) de serem desenvolvidas. Por exemplo, MLLMs viabilizam o desenvolvimento de assistentes inteligentes avançados que podem verdadeiramente "ver" o mundo através da câmera de um *smartphone* e interagir com ele de forma contextual, ajudando usuários desde a identificação de um organismo vegetal até a solução de problemas em manuais de instrução. Na área de educação personalizada, permitem a criação de tutores interativos que utilizam diagramas, ilustrações e textos para explicar conceitos complexos de forma adaptativa. Para análise de dados complexos, possibilitam a interpretação automatizada de relatórios médicos que combinam imagens de raio-X com laudos textuais, ou análise de imagens de satélite acopladas a dados geolocalizados para monitoramento ambiental. Na automação robótica, fornecem aos robôs uma compreensão de alto nível de seu ambiente, permitindo que executem tarefas baseadas em instruções naturais, como "pegue a xícara vermelha sobre a mesa" (compreensão exemplificada por modelos como o *PaLM-E* [Driess et al. 2023]).

1.1.3. Visão Geral do Material

Este texto foi concebido para guiar os interessados em uma jornada compreensiva pelos Grandes Modelos de Linguagem Multimodais, abrangendo desde seus fundamentos teóricos até a implementação prática. A estrutura foi planejada para oferecer uma base sólida a iniciantes, ao mesmo tempo que apresenta tópicos avançados e o estado da arte para praticantes experientes. O conteúdo está organizado da seguinte forma:

- **Seção 1.2: Fundamentos de LLMs e Multimodalidade** - Esta seção estabelece as bases, revisitando a arquitetura *Transformer* e a evolução para os LLMs modernos baseados apenas em *decoders*. Em seguida, aborda-se uma parte chave da multimodalidade, explicando o conceito de alinhamento de *embeddings* e das principais arquiteturas de MLLMs, bem como uma análise de modelos emblemáticos como GPT-4V, Gemini e LLaVA.
- **Seção 1.3: Técnicas para Pipeline Multimodal** - Nesta seção são apresentadas, de forma prática, estratégias para o pré-processamento de dados multimodais e técnicas avançadas de engenharia de *prompt*. A seção inclui ainda o uso de arcabouços como LangChain e LangGraph para construir *pipelines* robustos e escaláveis que integram MLLMs a outras ferramentas.
- **Seção 1.4: Aplicações Práticas e Casos de Uso** - Esta seção apresenta exemplos concretos de aplicação de MLLMs, incluindo geração de descrições de imagens, perguntas e respostas multimodais (VQA do inglês *Visual Question Answering*) e uma análise detalhada de um estudo de caso real: o arcabouço MLLMsent para análise de sentimentos em imagens.

- **Seção 1.5: *Hands-on*: Tutorial Prático** - Nesta seção, utilizando modelos acessíveis, *notebooks* práticos mostram desde a classificação direta de imagens com MLLMs até a construção de um *pipeline* completo de classificação via descrições geradas.
- **Seção 1.6: Limitações e Oportunidades** - Esta seção discute, de forma crítica, as principais limitações atuais dos MLLMs, como alucinações, viés e custo computacional; também explora as tendências promissoras para pesquisas futuras, incluindo modelos menores e mais eficientes e o papel dos agentes multimodais.
- **Seção 1.7: Conclusões**.

Este material visa fornecer ao interessado não apenas uma compreensão teórica do funcionamento dos MLLMs, mas também as habilidades práticas necessárias para desenvolver e integrar essas tecnologias transformadoras em seus próprios projetos, contribuindo para a próxima geração de aplicações inteligentes na Web e nas mídias digitais.

1.2. Fundamentos de LLMs e Multimodalidade

Esta seção apresenta uma visão condensada das arquiteturas que sustentam LLMs e MLLMs. A Subseção 1.2.1 revisa a evolução dos *Transformers* aos LLMs modernos; na Subseção 1.2.2 são descritos os mecanismos de alinhamento de *embeddings* que viabilizam a multimodalidade; já a Subseção 1.2.3 detalha como diferentes arquiteturas de MLLMs incorporam múltiplas modalidades; e, por fim, a Subseção 1.2.4 apresenta modelos e arcabouços representativos.

1.2.1. Arquiteturas Base: Dos *Transformers* aos LLMs Modernos

A compreensão dos MLLMs requer uma fundamentação sobre a arquitetura que os sustenta. No centro dessa evolução está o mecanismo de atenção e a arquitetura *Transformer*.

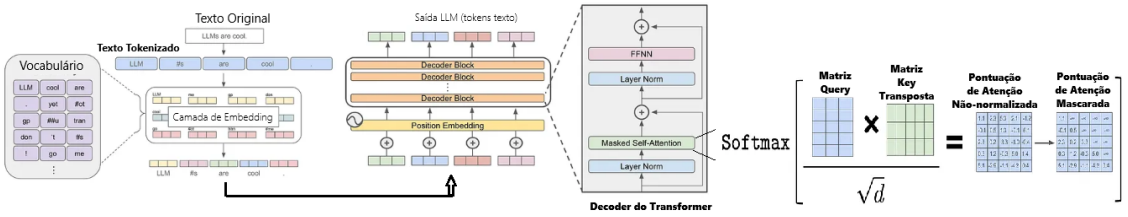
Introduzido pelo artigo seminal *"Attention is all you need"* [Vaswani et al. 2017], o mecanismo de auto-atenção (*self-attention*) permite que um modelo pondere a importância relativa de diferentes *tokens*[1] em uma sequência textual, considerando dependência de longo alcance. Esse mecanismo é a base para codificar contexto de maneira eficiente. O *Transformer* original é composto por uma pilha de *encoders*, que mapeiam uma sequência de entrada para uma representação latente, e uma pilha de *decoders*, que geram uma sequência de saída a partir dessa representação. Essa arquitetura inovadora abriu caminho para especializações em duas direções principais.

Modelos baseados em *encoder*, como o BERT [Devlin et al. 2019], que são otimizados para tarefas de **compreensão de linguagem**. Usam atenção bidirecional, acessando simultaneamente o contexto à esquerda e à direita de cada token, o que os torna adequados para análise de texto, classificação e resposta a perguntas condicionadas a um contexto dado. São considerados modelos não autorregressivos (preveem os *tokens* simultaneamente, sem dependência de *tokens* anteriores). Por outro lado, modelos baseados em *decoder*, como a família GPT [Radford et al. 2018, Radford et al. 2019], são otimizados

para a **geração de linguagem**. Nesse caso, a atenção é causal (ou mascarada), acessando apenas o contexto anterior na sequência para prever o próximo *token*. A geração ocorre de forma autorregressiva, *token* por *token*, o que garante fluidez na produção textual.

No uso contemporâneo, o termo LLMs refere-se predominantemente a *decoders* autorregressivos, como as famílias GPT e LLaMA [Touvron et al. 2023]. Com centenas de bilhões de parâmetros treinados em corpora massivos de texto, esses modelos demonstram capacidades emergentes de "raciocínio", execução de instruções e geração de linguagem natural [Wei et al. 2022a].

A Figura 1.3 ilustra o funcionamento geral de um LLM. Primeiro, a entrada textual é tokenizada, ou seja, dividida em unidades menores (*tokens*). Cada *token* é convertido em um vetor de representação numérica densa (*embedding*), sendo somado ao vetor de representação numérica densa da posição de cada *token* na sequência (*positional encoding embedding*). Esses vetores são enviados a uma pilha de *decoders* do *Transformer*, onde ocorre o processamento autorregressivo gerando um *token* de saída por vez. No *decoder*, a etapa central é a auto-atenção mascarada, onde cada *token* gera três vetores: **q** (*query*), **k** (*key*) e **v** (*value*); os quais, agregados para todos os *tokens* de entrada, formam as respectivas matrizes, Q, K e V. Calcula-se a similaridade entre Q e K, aplica-se *softmax* e multiplica-se pelos valores V, mas com uma máscara de elementos com valores infinitos que impede o modelo de ver *tokens* futuros. O resultado, somado a uma conexão residual, passa por uma camada de normalização, garantindo estabilidade e aprendizado eficiente. Em seguida, os vetores passam por uma rede neural de propagação direta (FFNN do inglês *feed-forward neural network*). Esse processo realizado em cada bloco do *decoder* se repete por todas as camadas do *Transformer* até a última, onde uma camada linear projeta o vetor final para o vocabulário, produzindo uma distribuição de probabilidade para o próximo *token*. Então, o modelo seleciona o referido *token* via amostragem e o realimenta no fluxo, gerando o conteúdo de maneira incremental.



A multimodalidade surge como uma extensão natural desses LLMs. Conforme mostrado na Figura 1.1, a inovação crucial reside na incorporação de *encoders* especializados que convertem entradas não-textuais (como *pixels* de imagem) em representações (*embeddings*) que podem ser projetadas no espaço latente dos LLMs. Através de técnicas de alinhamento multimodal, como ocorre no CLIP [Radford et al. 2021] (ver Figura 1.4), o modelo aprende a associar conceitos de diferentes modalidades, permitindo que este "entenda" uma imagem e converse sobre ela, ou que gere uma imagem a partir de uma descrição textual.

Modelos como o GPT-4V [OpenAI 2023b] e o Gemini [Team et al. 2023] partem da mesma base de *decoders* autorregressivos para textos, mas incorporam informações

adicionais, como visão ou áudio. A estratégia predominante consiste em projetar *embeddings* de todas as modalidades num espaço comum compreensível para o *decoder*. Assim, o núcleo autorregressivo é mantido, mas agora gera respostas para entradas multimodais. Essa herança arquitetural é o que permite aos MLLMs gerar linguagem de forma coerente sobre conteúdos além dos textuais, como visuais e auditivos.

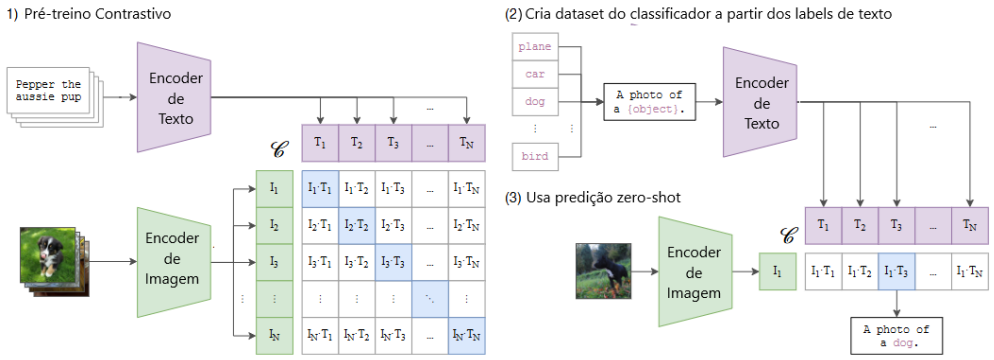
1.2.2. A Ponte para a Multimodalidade: Alinhamento de *Embeddings*

Um ponto-chave dos MLLMs é a capacidade de integrar informações de modalidades distintas, como texto, imagem e áudio, em um modelo coerente. O desafio é que cada modalidade possui uma representação numérica nativa (*tokens*, *pixels*, sinais de áudio, respectivamente), que habita espaços vetoriais distintos e não comparáveis diretamente. A solução usual está no alinhamento de *embeddings*.

Cada modalidade é inicialmente processada por uma arquitetura especializada. Modelos de linguagem baseados em *Transformer*, como o GPT, mapeiam sequências de *tokens* para um espaço vetorial $\mathscr{T}$. *Encoders* visuais, como *Vision Transformers* (ViTs) [Dosovitskiy et al. 2020], projetam *patches* de imagens em um espaço vetorial $\mathscr{V}$. Modelos de áudio, como *wav2vec 2.0* [Baevski et al. 2020], convertem sinais de áudio em um espaço $\mathscr{A}$. Sem um mecanismo de alinhamento, os espaços $\mathscr{T}$, $\mathscr{V}$ e $\mathscr{A}$ são como línguas diferentes: um modelo de linguagem é incapaz de processar nativamente representações vetoriais oriundas do domínio visual sem a devida mediação arquitetural. O *joint embedding* propõe justamente mapear essas modalidades em um espaço semântico latente compartilhado $\mathscr{C}$, ou seja, uma linguagem universal que o modelo entende. Nesse espaço, conceitos semelhantes, como a palavra "cachorro", a foto de um cachorro e o som de um latido, são projetados em regiões vizinhas [Baltrušaitis et al. 2018]. Essa projeção permite que o LLM, operando em $\mathscr{C}$, processe e gere conteúdo condicionado em múltiplas modalidades.

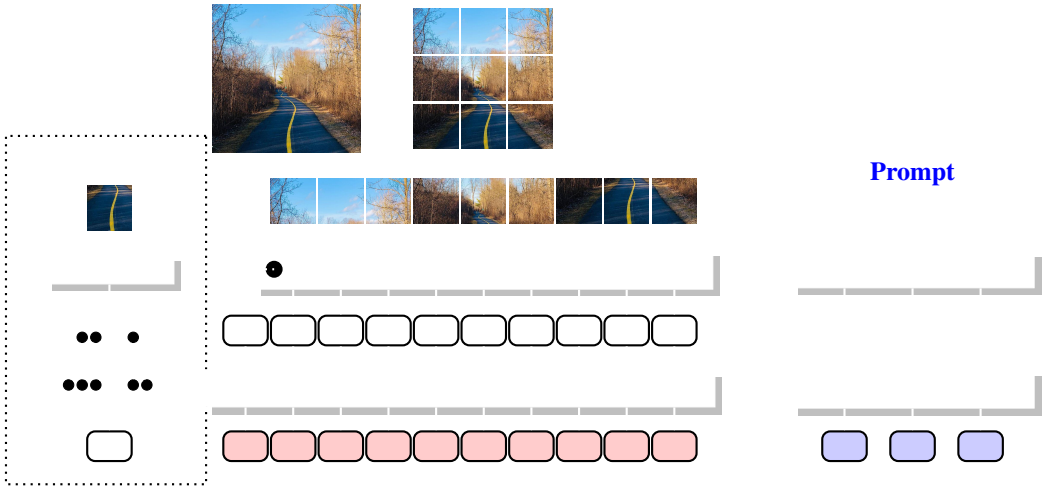
Uma técnica conhecida para produzir esse alinhamento é o aprendizado contrastivo, no qual o modelo aprende representações úteis comparando exemplos relacionados (alinhados) e não relacionados [Chen et al. 2020]. O CLIP (*Contrastive language–image pre-training*) [Radford et al. 2021] é um exemplo emblemático. Ele é treinado com milhões de pares (imagem, legenda) e, conforme ilustrado na Figura 1.4, utiliza dois *encoders*: um *encoder* de texto (ex.: *Transformer*) para mapeamento em $\mathscr{T}$ e um *encoder* de imagem (ex.: ViT) para mapeamento em $\mathscr{V}$. No CLIP, os *encoders* visual e textual são treinados conjuntamente para que ambos produzam *embeddings* em um espaço latente compartilhado $\mathscr{C}$, e o objetivo do aprendizado é maximizar a similaridade entre pares que são relacionados (diagonal principal) e minimizá-la para pares não relacionados (fora da diagonal principal). O resultado é um espaço multimodal compartilhado $\mathscr{C}$ que permite tarefas de recuperação *zero-shot* (inferência em um contexto novo, diferente do treino).

Nos MLLMs, o CLIP atua como uma ponte para a incorporação de imagens, fornecendo um espaço compartilhado $\mathscr{C}$ no qual representações visuais e textuais são alinhadas de forma simétrica [Radford et al. 2021]. Durante o treinamento do CLIP, um par de *encoders*, um visual e outro textual, é otimizado conjuntamente para ter seus *embeddings* de imagens e descrições projetados em um mesmo espaço latente, de modo que pares correspondentes fiquem próximos e pares não correspondentes se afastem. Em modelos multimodais baseados em LLMs, esses *embeddings* visuais produzidos pelo CLIP



podem ser posteriormente projetados para um espaço compreensível pelo modelo de linguagem [Liu et al. 2023], permitindo que o LLM trate a informação visual como uma forma de entrada textual compatível. Assim, o modelo não "traduz" a imagem diretamente para o texto, mas permite que o LLM interprete a informação visual como *tokens* compatíveis com seu espaço latente.

A Figura 1.5 ilustra o processo de geração dos *embeddings* de imagem e texto empregados como entrada em modelos de linguagem multimodais.



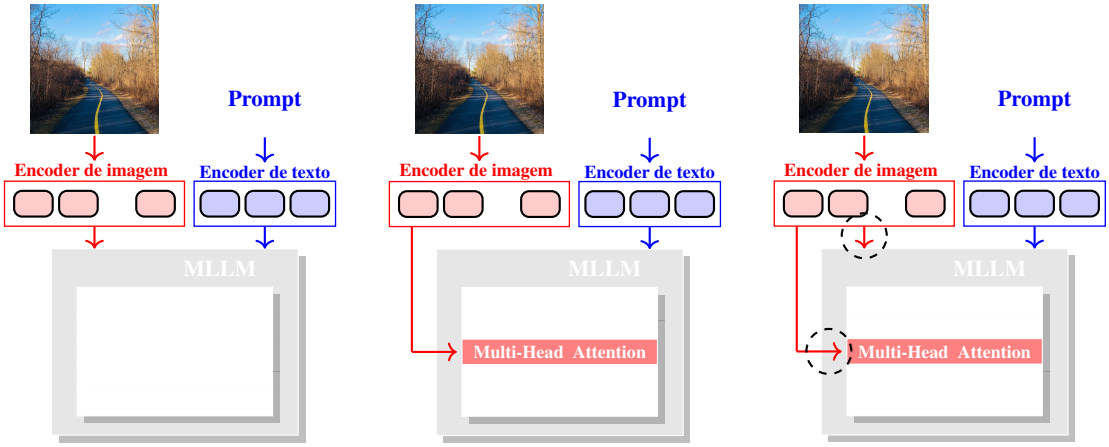
Inicialmente, a imagem fornecida pelo usuário com largura e altura de $W \times H$ *pixels* é dividida em pequenos blocos (ou *patches*) de $P \times P$ *pixels*, que são linearizados e empilhados sequencialmente. Cada *patch* é então projetado em um espaço vetorial de Q dimensões por meio de uma camada totalmente conectada, resultando em uma sequência de vetores que, juntamente com o *token* especial [CLS], formam a entrada para um *Vision Transformer* (ViT). O valor de Q é escolhido de tal forma a se encaixar na entrada de cada

token definido pelo ViT. Após o processamento pelo ViT, são obtidos os *image embeddings* correspondentes, que codificam a informação visual de maneira contextualizada. De forma análoga, a sequência textual, exemplificada pelo *prompt* "Descreva a imagem", é tokenizada, transformando cada palavra em uma unidade discreta. Esses são os *tokens* de texto, cujos *embeddings* são posteriormente projetados, assim como os *embeddings* visuais, em um mesmo espaço latente.

1.2.3. Arquiteturas de MLLMs: Como a Multimodalidade é Incorporada

Com os fundamentos estabelecidos, é possível compreender as principais estratégias de construção de MLLMs. A ideia central é utilizar o LLM como um "sistema cognitivo central", ao qual se conectam *encoders* especializados para processar outras modalidades. A questão é, portanto, de interface: como conectar essas representações ao modelo de linguagem de forma que ele possa "raciocinar" sobre informações multimodais.

Dentro desse panorama, é possível classificar as estratégias de integração multimodal em termos de fusão precoce (também chamada de *input-level fusion*) e fusão tardia (ou *model-level fusion*). Na fusão precoce, as diferentes modalidades são alinhadas e combinadas em um espaço comum logo no início do processamento, de modo que o modelo aprenda representações conjuntas desde as camadas iniciais. Já na fusão tardia, cada modalidade é processada separadamente e apenas em estágios posteriores as representações são combinadas, tipicamente dentro do LLM. A Figura 1.6 ilustra os dois tipos de integração, assim como um terceiro que agrega as duas estratégias, chamado de fusão híbrida. A escolha entre fusão precoce, tardia ou híbrida envolve importantes *trade-offs* práticos em termos de custo computacional, reuso de modelos e capacidade de generalização.



Fusão precoce oferece o maior grau de interação entre modalidades, pois o modelo aprende representações conjuntas desde as camadas iniciais. No entanto, essa abor-

dagem exige grandes volumes de dados multimodais anotados e treinamento ponta a ponta, o que aumenta o custo e o tempo de ajuste fino.

Fusão tardia é mais eficiente do ponto de vista de engenharia, pois permite reaproveitar *encoders* e LLMs já pré-treinados. Ela requer apenas um módulo adicional para combinar as representações - frequentemente por meio de blocos de *cross-attention* - mas pode gerar integrações mais superficiais, com menor sinergia entre modalidades.

Fusão híbrida busca equilibrar as limitações descritas anteriormente, combinando a integração inicial de *embeddings* com mecanismos de fusão em camadas intermediárias. Isso tende a melhorar a flexibilidade e a coerência multimodal, embora introduza complexidade arquitetural e desafios de otimização.

Um componente técnico central nas arquiteturas modernas de fusão tardia e híbrida é o mecanismo de atenção cruzada (*cross-attention*). Diferentemente da autoatenção, na qual cada token interage apenas com outros *tokens* da mesma sequência, a atenção cruzada permite que *tokens* textuais consultem representações visuais ou auditivas em cada etapa de decodificação.

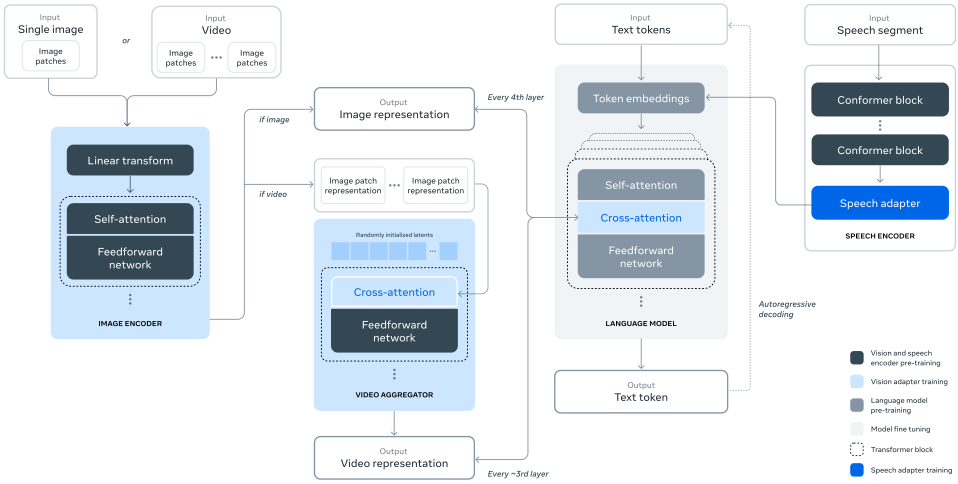
Em modelos como o *Flamingo* [Alayrac et al. 2022], blocos específicos de *cross-attention* são intercalados nas camadas do LLM, funcionando como portas de comunicação entre modalidades. Esses blocos aprendem a pesar dinamicamente quais partes da imagem são mais relevantes para o texto sendo gerado, viabilizando raciocínio multimodal contextual. Essa abordagem preserva o núcleo linguístico do LLM intacto, mas adiciona uma camada de alinhamento adaptativo, na qual o modelo aprende não apenas a compreender a imagem, mas também a consultá-la seletivamente durante a geração textual (um aspecto essencial para tarefas como descrição visual, interpretação de gráficos e raciocínio *visual-textual*).

1.2.4. Modelos Emblemáticos e Arcabouços

A trajetória recente dos MLLMs pode ser ilustrada por alguns modelos emblemáticos que marcaram diferentes fases de sua evolução. Entre eles, destaca-se o GPT-4V(*ision*) [OpenAI 2023b], que estende o GPT-4 para processar imagens. Embora detalhes arquiteturais não sejam públicos, acredita-se que siga o paradigma de encapsulamento de modalidades, integrando um *encoder* visual avançado ao núcleo de linguagem. Sua contribuição reside na demonstração de que modelos comerciais de larga escala podem realizar raciocínio visual sofisticado, incluindo tarefas de OCR e interpretação contextual de imagens do mundo real.

O Gemini [Team et al. 2023] representa outro marco, anunciado como a primeira família de modelos nativamente multimodais em sua concepção. Diferente de arquiteturas adaptadas, foi projetado desde o início para processar texto, imagem, áudio e vídeo de forma integrada. Sua arquitetura baseia-se em *Transformers* modificados e treinados ponta a ponta em dados multimodais, permitindo uma interação mais profunda entre modalidades. A família de modelos Gemini inclui variantes otimizadas para diferentes contextos de uso, desde dispositivos móveis até grandes servidores de processamento.

O PaLM-E [Driess et al. 2023], desenvolvido pelo Google DeepMind, é um modelo multimodal voltado para a robótica, que combina percepção visual e textual com



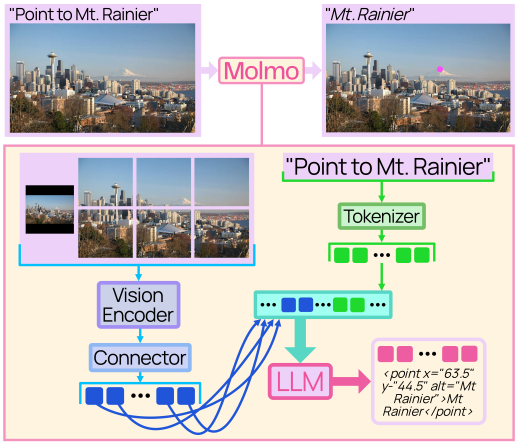
capacidade de planejamento e tomada de decisão em ambientes físicos. Seu diferencial é demonstrar como os MLLMs podem ser incorporados em agentes encarnados, aproximando a inteligência artificial de aplicações práticas no mundo real.

Outro modelo relevante é o Llama 3.2 [Grattafiori et al. 2024], disponibilizado pela empresa Meta. Este modelo é um exemplo ilustrativo da estratégia de integração multimodal por fusão tardia, como mostrado na Figura 1.7. Como se observa, além da modalidade textual, o modelo suporta também entradas visuais e sonoras.

Entre os modelos de código aberto, o LLaVA [Liu et al. 2023] tornou-se um dos mais influentes. Conectando o *encoder* visual do CLIP a um LLM derivado do LLaMA via projeção linear, demonstrou que a qualidade do alinhamento e o ajuste fino (*fine-tuning*) em instruções visuais são mais determinantes que a escala da arquitetura. O BLIP-2 [Li et al. 2023] segue uma linha semelhante, mas introduz um módulo intermediário de projeção treinado para alinhar imagens a modelos de linguagem pré-existentes de forma mais eficiente. Essa abordagem reduziu significativamente o custo de treinamento multimodal, impulsionando a adoção em contextos de pesquisa e aplicações abertas.

Outro modelo aberto é o Molmo [Deitke et al. 2025]. Conforme ilustrado na Figura 1.8, o codificador de imagem utiliza um *Transformer* de visão padrão, especificamente o CLIP. O termo *conector* na figura refere-se a um *projetor* que alinha os *embeddings* da imagem com o modelo de linguagem, implementando uma fusão precoce entre as modalidades. Desenvolvido pela AllenAI, o Molmo foi lançado com pesos e dados abertos, buscando oferecer uma referência transparente e reprodutível para pesquisas em MLLMs.

O modelo NVLM da NVIDIA [Dai et al. 2024] é particularmente interessante porque, em vez de focar em uma única abordagem, explora ambas as estratégias de integração multimodal. Na Figura 1.9, a parte inferior nomeada *Decoder-only* (NVLM-D)



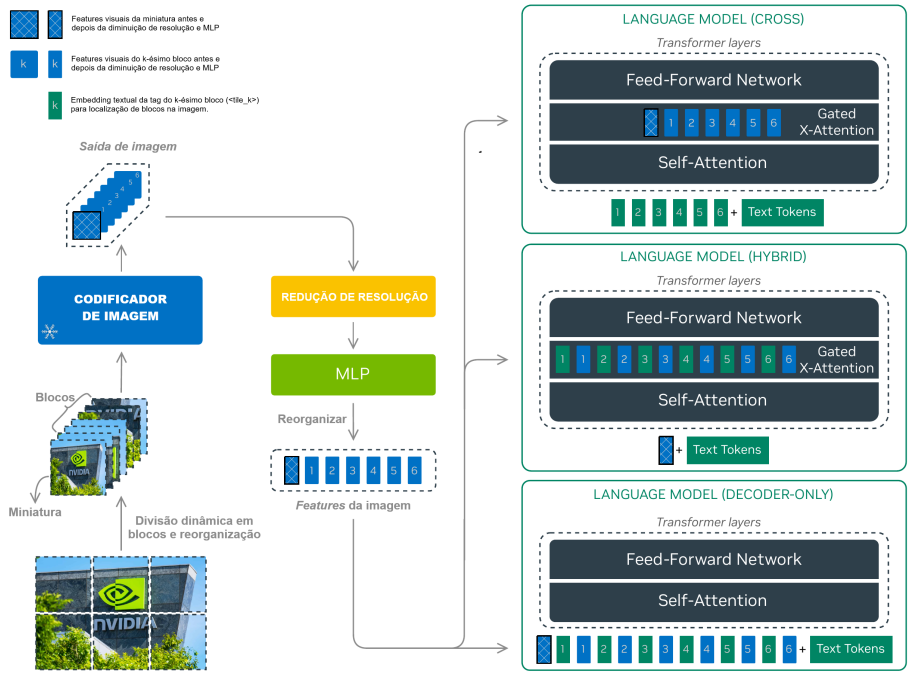
refere-se ao método de fusão precoce. A parte superior, nomeada *Cross* (NVLM-X), refere-se ao método de fusão tardia. Além disso, os autores desenvolvem uma abordagem híbrida, nomeada *Hybrid* (NVLM-H).

Os pesquisadores observam que a estratégia NVLM-X é mais eficiente, computacionalmente, para imagens de alta resolução. Já a estratégia NVLM-D apresenta melhor desempenho em tarefas relacionadas a OCR. Por fim, a estratégia NVLM-H combina os pontos fortes de ambos os métodos, NVLM-D e NVLM-X.

Por fim, ecossistemas como o Hugging Face [Wolf et al. 2020] têm desempenhado um papel crucial na disseminação dessas tecnologias. A biblioteca `transformers` fornece uma API unificada para carregar e utilizar modelos de linguagem e multimodais, além de *pipelines* prontos para tarefas como geração de legendas de imagens, VQA e tradução multimodal. Esse ecossistema facilita o compartilhamento de modelos e dados, acelera a reprodutibilidade e democratiza o acesso aos MLLMs de última geração, reduzindo a barreira de entrada para instituições acadêmicas, pequenas empresas e pesquisas independentes.

1.3. Técnicas para Pipeline Multimodal

Esta seção apresenta os componentes fundamentais para a construção de *pipelines* eficientes com modelos multimodais. O conteúdo é organizado de maneira progressiva, contemplando desde o tratamento inicial dos dados até ferramentas que permitem operacionalizar fluxos mais complexos. (i) Inicialmente, discute-se o pré-processamento multimodal, etapa essencial para garantir a qualidade das entradas e o alinhamento entre modalidades, permitindo que textos e imagens sejam representados em um espaço semântico compartilhado. (ii) Em seguida, são abordadas algumas técnicas de engenharia de *prompt*, incluindo estratégias como o *chain-of-thought* multimodal (por exemplo, descrever o raciocínio sobre uma imagem de forma incremental) e o *few-shot prompting* com pares imagem-texto para orientar o modelo em tarefas específicas. (iii) Por fim, a seção introduz arcabouços que facilitam a construção de *pipelines* avançados, tais como LangChain e LangGraph. Tais arcabouços fornecem abstrações, componentes reutilizáveis,



padronização de código e mecanismos de depuração e monitoramento, simplificando o desenvolvimento e a manutenção de sistemas multimodais.

1.3.1. Pré-processamento de dados

O pré-processamento é uma etapa crucial no desenvolvimento de pipelines multimodais, pois é responsável por converter os dados brutos (imagens - ver exemplo na figura 1.10, áudios, vídeos e textos) em representações vetoriais densas (*embeddings*) compreensíveis pelos MLLMs. Em outras palavras, é nessa fase que se realiza a "tradução" das modalidades para o espaço de linguagem do modelo.

Como discutido anteriormente, os MLLMs não operam diretamente sobre *pixels*, sinais de áudio ou *tokens* crus, mas sobre vetores que representam semanticamente cada conteúdo. A etapa de pré-processamento, portanto, consiste em (i) normalizar e limpar os dados de entrada, (ii) extrair suas representações (via *encoders* especializados), e (iii) projetar essas representações em um espaço semântico compatível com o LLM.

A implementação prática dessa etapa varia conforme a modalidade de entrada. Por exemplo, para mapear imagens e textos em um espaço multimodal compartilhado, há o modelo CLIP – uma das abordagens mais utilizadas [Radford et al. 2021]. O Código 1 mostra como carregar o CLIP e gerar *embeddings* da imagem apresentada na Figura 1.10 e um texto para posterior integração no pipeline.

```
1 from PIL import Image
2 import torch
3 import clip
```

```
from torchvision import transforms

# Carrega o modelo CLIP pré-treinado
device = "cuda" if torch.cuda.is_available() else "cpu"
model, preprocess = clip.load("ViT-B/32", device=device)

# Pré-processa uma imagem de entrada
image = preprocess(Image.open("bird.jpeg")).unsqueeze(0).to(device)

# Pré-processa o texto de entrada
text = clip.tokenize(["uma ave tesourinha em um galho de árvore"]).to(device)

# Extrai embeddings
with torch.no_grad():
    image_features = model.encode_image(image)
    text_features = model.encode_text(text)

# Normaliza e calcula similaridade
image_features /= image_features.norm(dim=-1, keepdim=True)
text_features  /= text_features.norm(dim=-1, keepdim=True)

similarity = (100.0 * image_features @ text_features.T).softmax(dim=-1)
print(f"Similaridade imagem-texto: {similarity.item():.4f}")
```

O BLIP-2 [Li et al. 2023] é uma evolução do CLIP, projetado especificamente para conectar representações visuais a LLMs, usando um módulo intermediário chamado *Q-Former*. Esse módulo aprende a projetar *embeddings* de imagem para o espaço de linguagem. O Código 2 apresenta um código escrito em Python com o BLIP-2, utilizando

a biblioteca *Transformers* da *Hugging Face*. Nota-se que neste exemplo, ao invés de calcular a similaridade, o modelo gera a descrição da imagem (*image captioning*).

```python
from transformers import Blip2Processor, Blip2ForConditionalGeneration
from PIL import Image
import torch

device = "cuda" if torch.cuda.is_available() else "cpu"

# Carrega o modelo BLIP-2
processor = Blip2Processor.from_pretrained("Salesforce/blip2-flan-t5-xl")
model = Blip2ForConditionalGeneration.from_pretrained("Salesforce/blip2-flan-t5-xl",
 ↪  torch_dtype=torch.float16).to(device)

# Pré-processa a imagem
image = Image.open("exemplo_imagem.jpg")
inputs = processor(images=image, return_tensors="pt").to(device, torch.float16)

# Extrai embeddings e gera texto condicional
generated_ids = model.generate(**inputs, max_new_\textit{tokens}=20)
generated_text = processor.batch_decode(generated_ids,
 ↪  skip_special_\textit{tokens}=True)[0]
print("Descrição gerada:", generated_text)
```

Para dados de áudio, o CLAP (*Contrastive Language–Audio Pretraining*) [Elizalde et al. 2023] desempenha um papel semelhante ao do CLIP, mapeando sons e descrições textuais para um espaço compartilhado, conforme mostrado no Código 3.

```python
from transformers import ClapProcessor, ClapModel
import torch
import torchaudio

# Carrega o modelo CLAP
processor = ClapProcessor.from_pretrained("laion/clap-htsat-unfused")
model = ClapModel.from_pretrained("laion/clap-htsat-unfused")

# Carrega um arquivo de áudio
waveform, sr = torchaudio.load("exemplo_audio.wav")

# Pré-processa o áudio e o texto
inputs = processor(audios=waveform, text=["som de pássaros cantando"],
 ↪  return_tensors="pt", padding=True)

# Extrai embeddings multimodais
with torch.no_grad():
    outputs = model(**inputs)
    audio_emb = outputs.audio_embeds
    text_emb = outputs.text_embeds
    similarity = outputs.logits_per_audio  # similaridade áudio-texto
```

```
21
22 print(f"Embedding de áudio: {audio_emb.shape}")
23 print(f"Embedding de texto: {text_emb.shape}")
24 print(f"Similaridade áudio-texto: {similarity.item():.4f}")
```

Além da geração de *embeddings* multimodais, o estágio de pré-processamento compreende procedimentos cruciais para o refinamento da qualidade e o incremento da interpretabilidade dos dados multimodais. Por exemplo, pode-se realizar a normalização, redimensionamento, segmentação de região de interesse (ROI do inglês *Region of Interest*) de imagens; remoção de ruído e normalização de amplitude em áudios; tokenização, remoção de *stopwords* e lematização em textos; geração de metadados contendo informações como formato, resolução, duração, ou anotações semânticas (e.g., entidades nomeadas), que podem auxiliar na recuperação e indexação de exemplos relevantes; e, ainda, o aumento de dados (*data augmentation*) multimodal, como variações visuais, ruído acústico ou reformulação de legendas, para aumentar a robustez do modelo. Tais práticas são especialmente importantes quando o objetivo é construir um conjunto de dados multimodais personalizados.

Desta maneira, o pré-processamento é, portanto, a ponte entre os dados brutos e o raciocínio em múltiplas modalidades do MLLM. Ele transforma *pixels*, sinais de áudio e *tokens* em vetores comparáveis, alinhando múltiplas modalidades em um mesmo espaço de linguagem, permitindo que o modelo "pense" sobre imagens, sons e textos de forma integrada.

1.3.2. Engenharia de *Prompt* para Tarefas Multimodais

Para melhorar os resultados dos MLLMs, um aspecto central é a elaboração de bons *prompts*, processo conhecido como engenharia de *prompt*. Mas o que caracteriza um "bom" *prompt*? De modo geral, é um conjunto de comandos com instruções claras, específicas e contextualizadas, nas quais a tarefa é descrita de maneira objetiva e, quando necessário, acompanhada de exemplos. Além disso, o *prompt* pode incluir instruções sobre o formato da saída (por exemplo, lista, tabela, JSON ou texto corrido) e orientações sobre estilo ou tom da resposta. A estruturação do *prompt* desempenha um papel significativo, especialmente quando ele se torna longo. Nestes casos, recomenda-se organizar o conteúdo em blocos lógicos, como listas numeradas, seções com títulos, ou palavras-chave destacadas. *Prompts* organizados facilitam o processamento interno do modelo e reduzem ambiguidades. Outra estratégia comumente utilizada é direcionar o modelo a assumir uma persona, isto é, uma configuração de comportamento ou papel atribuído ao modelo, influenciando seu estilo de resposta, tom, vocabulário e nível de detalhamento, como se estivesse "interpretando" uma identidade específica.

Com essas práticas, já é possível obter resultados relevantes por meio do *zero-shot prompting*, que consiste em solicitar que o modelo realize uma tarefa sem fornecer exemplos prévios, apenas com instruções diretas. No entanto, em tarefas mais complexas, o *zero-shot* pode não ser suficiente. Assim, utilizam-se técnicas mais avançadas, como o

few-shot prompting, *Chain-of-Thought* (CoT), *self-consistency*, ReAct (*Reason* + *Act*), e o *prompting* multimodal guiado.

O *few-shot prompting* [Brown et al. 2020] consiste em fornecer ao modelo alguns exemplos da tarefa desejada diretamente no *prompt*, antes de solicitar a resposta final. Ao observar esses exemplos, o modelo identifica o padrão de entrada e saída, reproduzindo o estilo, abordagem e nível de detalhamento esperado. Essa técnica é especialmente útil quando a tarefa é complexa, ambígua ou envolve convenções específicas, por exemplo, formato de resumo, estilo narrativo, ou critérios específicos de classificação. Em modelos multimodais, o *few-shot* pode incluir não apenas texto, mas também imagens ou combinações imagem-texto, demonstrando como relacionar as modalidades. É importante escolher exemplos curtos, representativos e consistentes, pois exemplos ruins ou variados demais podem levar o modelo a reproduzir padrões incorretos.

Já o CoT [Wei et al. 2022b] incentiva o modelo a explicitar o raciocínio passo a passo antes de apresentar a resposta final. Esse encadeamento de pensamento ajuda o modelo a organizar as informações de maneira sequencial e lógica, podendo reduzir erros e aumentar a precisão em tarefas que exigem raciocínio complexo, como problemas matemáticos, interpretação de gráficos, inferência causal e análise de imagens com múltiplos elementos. Em MLLMs, o CoT pode envolver a descrição progressiva da cena visual, destacando observações intermediárias antes de chegar à conclusão. A orientação explícita para "pensar em voz alta" tende a melhorar a qualidade da resposta, embora deva ser usada com cuidado em contextos sensíveis ou quando se deseja respostas curtas.

A técnica de *self-consistency* [Wang et al. 2022] parte da ideia de que o modelo pode gerar diferentes cadeias de raciocínio possíveis para uma mesma tarefa, especialmente quando o problema admite mais de um caminho lógico. Em vez de confiar em uma única resposta do CoT, o modelo é instruído a gerar múltiplas explicações ou raciocínios independentes, e posteriormente selecionar (ou combinar) o resultado mais recorrente ou coerente entre eles. Na prática, isso reduz o impacto de cadeias de raciocínio incorretas ou enviesadas que poderiam surgir em uma única tentativa. Apesar de aumentar a qualidade, ela implica maior custo computacional, já que requer múltiplas amostragens.

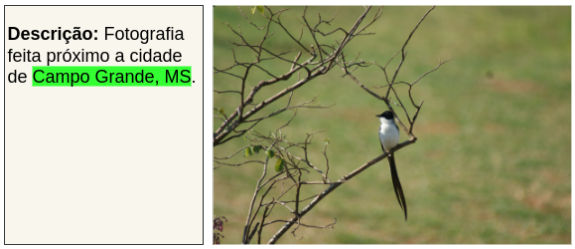
O ReAct [Yao et al. 2022] é uma técnica que combina raciocínio verbal e ações iterativas para resolver problemas de forma interativa. O modelo não apenas descreve o raciocínio, mas também executa ações intermediárias, como consultar documentos, examinar imagens novamente, fazer buscas externas ou solicitar informações adicionais. Esse ciclo "pensar - agir - pensar novamente" torna o modelo mais adequado para funções de agente autônomo, pipelines multimodais e cenários em que é necessário integrar fontes externas de conhecimento. Em MLLMs, isso pode significar interpretar uma imagem, fazer hipóteses, verificar detalhes adicionais na própria imagem ou consultar metadados, e então ajustar a conclusão final. O ReAct reduz erros relacionados a suposições precipitadas, mas depende de um ambiente que permita operações de consulta ou APIs externas.

O *prompting* multimodal guiado [Liu et al. 2023] refere-se a estratégia de orientar explicitamente como o modelo deve integrar e relacionar informações de diferentes modalidades. Por exemplo, pode-se instruir o modelo da seguinte forma: "descreva o que está na imagem e explique como isso se relaciona com o texto fornecido". Em MLLMs, simplesmente fornecer texto e imagem não garante que o modelo interpretará suas rela-

ções de forma alinhada ao objetivo da tarefa. Por isso, o *prompt* pode orientar etapas, como observar a imagem primeiro, identificar elementos-chave, relacionar esses elementos ao texto fornecido, e apenas então formular a resposta final. Essa técnica melhora a coerência entre modalidades e reduz interpretações equivocadas ou superficialmente descritivas. Ela é particularmente útil em tarefas de análise de documentos visuais, gráficos científicos, *captioning*, e explicação de conteúdo visual.

1.3.3. Arcabouços para facilitar a construção de *pipelines* avançados

A depender da complexidade da tarefa, os MLLMs, por si sós, podem não ser suficientes, o que torna necessária a criação de um *pipeline* que combine esses modelos com outras ferramentas para alcançar o objetivo desejado de forma satisfatória. Por exemplo, considerando a plataforma WikiAves[2], uma conhecida comunidade *online* de observadores de aves que mantém uma extensa base de dados sobre aves do Brasil. Supondo que essa plataforma deseje adicionar uma funcionalidade de busca por imagens, permitindo que os usuários façam *upload* de uma foto acompanhada de um breve texto com informações adicionais, como a localidade onde o registro foi feito, conforme ilustrado na Figura 1.11. Dessa forma, mesmo usuários iniciantes poderiam aprender mais sobre as aves, ainda que não conheçam seus nomes, que atualmente constituem o principal critério de busca na plataforma.

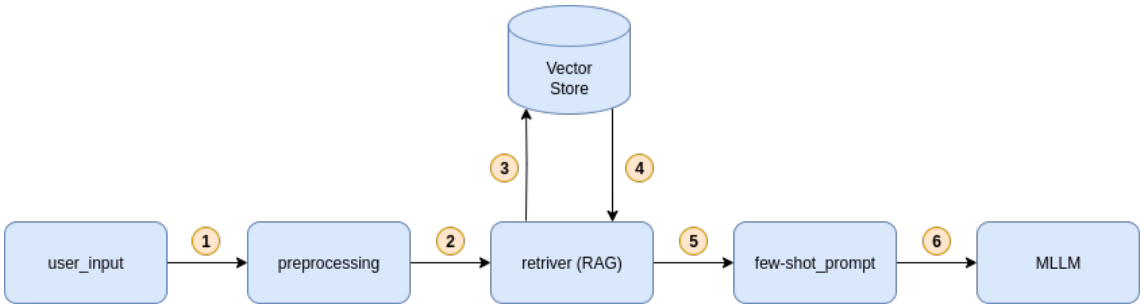


Para isso, a WikiAves pretende utilizar um MLLM capaz de processar as informações fornecidas pelo usuário (isto é, a foto e o breve texto) para identificar a espécie da ave e, em seguida, buscar em sua base de dados o registro correspondente, redirecionando o usuário para a página que contém todas as informações sobre ela. Considerando o exemplo apresentado na Figura 1.11, realizou-se um teste utilizando o modelo GPT-4o-mini-2024-07-18, da OpenAI (ou apenas, GPT-4o-mini, por simplicidade), por meio de uma requisição à API[3] do modelo, representada no Código 4.

```
{
  "model": "gpt-4o-mini",
  "messages": [
    {
      "role": "system",
```

```
      "content": "Você é um assistente especializado em análise visual. Você recebe
      ↪  como entrada uma imagem de uma ave e um breve texto com informações sobre a
      ↪  foto, e deve identificar qual ave se trata e responder o seu nome popular e
      ↪  científico. Responda em português."
    },
    {
      "role": "user",
      "content": [
        {
          "type": "text",
          "text": "Fotografia feita próximo a cidade de Campo Grande, MS."
        },
        {
          "type": "image_url",
          "image_url": {
            "url": "<URL>"
          }
        }
      ]
    }
  ],
  "max_\textit{tokens}": 500
}
```

Obteve-se a seguinte resposta do modelo: "A ave na fotografia é o **Bem-te-vi**. Seu nome científico é **Pitangus sulphuratus**", informações estas claramente erradas. Os motivos que levaram o modelo a cometer esse erro podem ser diversos, incluindo a possível falta de conhecimento sobre a fauna brasileira durante seu pré-treinamento. Para superar essa limitação, a plataforma WikiAves precisaria implementar um *pipeline* mais sofisticado para aprimorar o desempenho do modelo. A Figura 1.12 apresenta um possível *pipeline* para mitigar esse problema.



Como pode-se observar, após obter a entrada fornecida pelo usuário, o primeiro passo do *pipeline* é o pré-processamento, que, conforme discutido na Seção 1.3.1, além de gerar os *embeddings*, pode também ser utilizado para melhorar a qualidade da imagem, identificar e segmentar a região de interesse (isto é, a porção da imagem em que

se encontra a ave) e extrair as entidades nomeadas presentes no texto, neste caso, o local onde a fotografia foi capturada (a cidade de Campo Grande, MS). Com essas informações em mãos, o próximo passo do *pipeline* consiste em recuperar, a partir da base de dados da WikiAves, previamente vetorizada, ou seja, a *vector store*, registros de aves semelhantes que podem ocorrer na região informada, utilizando algum método de RAG (*Retrieve Augmented Generation*) [Lewis et al. 2020]. A partir dos resultados obtidos nessa etapa, constrói-se o *few-shot prompting*, como descrito na Seção 1.3.2, no qual são inseridos no *prompt* os exemplos de aves semelhantes recuperados via RAG. Dessa forma, o modelo pode aprender em tempo de execução com os exemplos fornecidos no contexto, para elaborar uma resposta mais assertiva.

Pode parecer um processo bem mais complexo do que simplesmente realizar uma solicitação à API do modelo, mas atualmente existem diversos arcabouços que facilitam a construção de *pipelines* como esse, tais como LangChain e LangGraph. Esses arcabouços fornecem várias ferramentas já implementadas, o que torna seu uso bastante simples, além de contribuírem para a padronização do código e oferecerem recursos para depuração de eventos.

No restante desta seção, serão fornecidos detalhes da utilização do LangChain e do LangGraph, arcabouços amplamente adotados pela indústria e pela academia, para a construção de *pipelines* com MLLMs.

1.3.3.1. LangChain

Com o LangChain [LangChain Inc. 2025a], é possível modelar um *pipeline* multimodal como uma sequência de componentes encadeados (*chains*), em que cada etapa realiza uma tarefa específica e passa os resultados para a próxima.

Cada etapa pode ser implementada como uma função encapsulada em um *RunnableLambda* ou *LLMChain*, compondo uma cadeia sequencial com o *SequentialChain*. O Apêndice A.1 mostra um exemplo em alto nível, em Python, de como implementar o *pipeline* utilizando o LangChain.

Durante a execução, o estado do *pipeline* é atualizado a cada etapa, e seus campos (como roi, rag_docs e *prompt*) são propagados entre as funções. Tal arquitetura confere transparência e modularidade ao fluxo de execução, viabilizando o desacoplamento de componentes e a substituição granular de etapas sem comprometer a integridade do *pipeline*. Essa abordagem também permite incluir ramos de decisão e verificação de erros, integrar modelos e ferramentas externas (como *embedders* multimodais, APIs e bancos vetoriais). Além disso, o LangChain possui integrações nativas com múltiplos tipos de modelos LLMs, MLLMs, *embeddings*, *retrievers*, ferramentas de visão computacional, e mais, o que simplifica o desenvolvimento de pipelines híbridos [LangChain Inc. 2025a].

A combinação de LangChain com ferramentas como LangSmith [LangChain Inc. 2025c] facilita a instrumentação, rastreamento e depuração de cada etapa do *pipeline*, sendo possível visualizar o fluxo completo de execução, incluindo as entradas e saídas de cada *chain*; medir latência, custos de chamadas e métricas de desempenho; identificar gargalos, falhas e respostas inesperadas em tempo real.

Modelar o *pipeline* com LangChain traz vantagens práticas e conceituais, tais como a **modularidade**, sendo que cada componente é independente, facilitando manutenção e testes; o **reuso**, onde as *chains* podem ser combinadas ou reutilizadas em outros fluxos; a escalabilidade, já que permite adicionar novas etapas à *chain*; e, por fim, a **integração**, tendo interoperabilidade direta com LangSmith, LangServe e LangStudio. Contudo, o LangChain apresenta algumas limitações em cenários que exigem controle mais sofisticado de fluxo e estado.

Por exemplo, cada *chain* é essencialmente uma função isolada, e o estado global do *pipeline* precisa ser propagado manualmente entre as etapas, o que aumenta a complexidade e o risco de inconsistências, especialmente em *pipelines* multimodais nos quais *embeddings* de diferentes tipos de dados (como imagem, texto e áudio) precisam ser sincronizados. Além disso, fluxos condicionais e ramificações lógicas não são tratadas de forma nativa. Implementações que envolvem verificações ou caminhos alternativos, como o caso em que nenhuma ave é detectada na imagem, acabam exigindo código imperativo fora da estrutura de *chains*, reduzindo a clareza e modularidade do projeto. Outra limitação é a ausência de uma representação gráfica explícita do *pipeline*, o que dificulta a visualização e a depuração em fluxos complexos.

1.3.3.2. LangGraph

Em contrapartida às limitações do LangChain, o LangGraph [LangChain Inc. 2025b] permite modelar o *pipeline* como um grafo de estados, no qual cada nó representa uma etapa e cada aresta define as condições de transição. Isso simplifica a manutenção, facilita a instrumentação e oferece suporte nativo a laços (*loops*), subgrafos e fluxos condicionais. Por esse motivo, o LangGraph se mostra mais adequado para *pipelines* multimodais com múltiplas dependências ou ramos de execução, proporcionando maior clareza estrutural, controle de estado compartilhado e escala organizacional do que o modelo sequencial tradicional do LangChain. Com o LangGraph [LangChain Inc. 2025b], o *pipeline* é modelado como um grafo de estados (ou nós), em que cada nó representa uma etapa específica do fluxo de processamento. Caso ocorra uma condição de erro em qualquer etapa, por exemplo, nenhuma ave detectada na imagem durante o pré-processamento, é possível definir um fluxo condicional que encerra a execução do *pipeline* de forma controlada.

Cada nó pode ler e/ou atualizar o estado compartilhado (*State*) do grafo. O LangGraph oferece recursos para definir arestas condicionais, laços, subgrafos, entre outros mecanismos que tornam o *pipeline* mais flexível e modular [LangChain Inc. 2025b].

Considerando o exemplo de *pipeline* ilustrado na Figura 1.12, são apresentados no Apêndice A.2 trechos de código com uma implementação em alto nível, escrita em Python, utilizando o arcabouço LangGraph.

Uma vez configurado o grafo, o LangGraph executa o *pipeline* de forma determinística e observável, permitindo inspecionar o fluxo de dados e as transições entre nós. Durante a execução, o estado é propagado entre os nós, de modo que cada etapa pode consumir os resultados anteriores e produzir novas saídas, enriquecendo progressivamente o contexto compartilhado.

Essa representação em grafo oferece modularidade, facilitando a substituição, atualização ou reuso de componentes individuais, por exemplo, trocar o mecanismo de recuperação. Também simplifica a instrumentação e o monitoramento, permitindo identificar gargalos de execução, métricas de desempenho e resultados intermediários. Ferramentas como o LangSmith são amplamente utilizadas nesse contexto, pois oferecem recursos avançados de rastreamento, depuração e visualização interativa dos fluxos de execução, facilitando o entendimento de como os dados percorrem o grafo e como cada nó contribui para o resultado final.

Por fim, o LangGraph integra-se naturalmente ao ecossistema LangChain, tornando possível combinar agentes, memória e ferramentas adicionais dentro do mesmo fluxo. Isso torna o arcabouço especialmente adequado para *pipelines* multimodais complexos, em que há múltiplas fontes de dados e etapas de raciocínio encadeadas.

1.4. Aplicações Práticas e Casos de Uso

Os MLLMs representam um dos avanços mais expressivos na confluência entre processamento de linguagem natural, visão computacional e aprendizado profundo. Após a consolidação teórica e o amadurecimento das arquiteturas que integram texto, imagem, áudio e vídeo, torna-se essencial compreender como esses modelos se traduzem em aplicações concretas. Esta seção discute exemplos reais de uso dos MLLMs em diferentes contextos, evidenciando sua versatilidade e potencial para transformar a interpretação e interação desses modelos com dados multimodais.

1.4.1. Análise em Conjunto de Dados Públicos

A disponibilidade de conjuntos de dados públicos contribui para o avanço tecnológico, ao habilitar a realização de testes de hipóteses sobre integração semântica entre múltiplas modalidades no contexto dos MLLMs. Essa disponibilidade possibilita o desenvolvimento de modelos em larga escala com maior capacidade de generalização e desempenho cada vez mais robusto. No entanto, a qualidade, o tamanho e a diversidade dos conjuntos de dados exigem atenção especial, uma vez que esses fatores afetam de forma decisiva a habilidade do modelo em compreender os dados e produzir representações consistentes.

Conjuntos de dados como *Visual Genome* estenderam a noção multimodal ao incluir relações semânticas entre objetos, atributos e regiões específicas da imagem. Esse avanço permite aos MLLMs uma evolução da compreensão de contextos visuais de forma relacional e não apenas descritiva, podendo ser denominadas de *visual grounding* e *scene graph generation* [Krishna et al. 2017].

No campo das emoções, conjunto de dados como o *PerceptSent* [Lopes et al. 2023] se destacam por incorporarem anotações afetivas em imagens permitindo análises mais profundas além do conteúdo objetivo. O conjunto de dados explora como as pessoas atribuem emoções a diferentes estímulos visuais consolidando a saída com informações quantitativas, pontuação do sentimento atribuído e perfil do avaliador, e até qualitativa, breves comentários sobre a explicação do sentimento ou motivações para a pontuação do sentimento. Essa abordagem híbrida amplia o potencial dos MLLMs em tarefas como análise de humor em redes sociais e psicologia computacional.

No entanto, é importante notar que, embora a disponibilidade desses conjuntos de dados tenha impulsionado diversas frentes de investigação, o emprego de tais conjuntos de dados requer um escrutínio crítico rigoroso. Questões como representatividade de gênero e cultura, curadoria automatizada e licenciamento ético têm sido foco de debate [Birhane et al. 2021]. A adoção de dados públicos, ou não, deve ser acompanhada de práticas de avaliação e filtragem, garantindo que o aprendizado multimodal não reproduza vieses ou distorções perceptuais.

1.4.2. Geração de descrições de imagens

A geração de descrições a partir de imagens, conhecida pelo termo *image captioning*, é uma das aplicações de MLLMs que vai além da identificação básica de objetos em uma imagem, mas envolve a interpretação do contexto, a inferência de relações e a tradução de percepções visuais em linguagem natural coerente e contextualizada.

Historicamente, os primeiros avanços para a geração de descrições de imagens ocorreram com o uso de arquiteturas *enconder-decoder* baseadas em aprendizado profundo (*deep learning*) nas quais a rede neural convolucional extraía características visuais e uma rede neural recorrente, ou LSTM, era responsável por gerar a sequência textual [Vinyals et al. 2015, Hossain et al. 2019]. Entretanto, arquiteturas como essas apresentavam limitações em capturar nuances semânticas mais complexas.

Com a transição para o uso de arquiteturas baseadas em *Transformers* o resultado na qualidade das descrições foi aprimorado, entregando resultados mais consistentes e dinâmicos. Modelos que incorporam mecanismos de atenção cruzada, permitem correlacionar regiões específicas da imagem com palavras ou conceitos relevantes do texto, por exemplo, o modelo BLIP-2 [Li et al. 2023]. Com o surgimento dos MLLMs, modelos avançados, como o GPT-4V, possibilitam o uso de técnicas de *captioning* contextual, nas quais é possível gerar diversos tipos de versões de *outputs* do modelo, ou seja, descrições mais humorísticas ou mais técnicas sobre a mesma imagem [OpenAI 2023a].

Como exemplo de aplicação prática, destacam-se modelagens como o Flamingo, desenvolvido pela DeepMind, que introduz a capacidade *few-shot* na geração multimodal [Alayrac et al. 2022]. Diferentemente de modelos que demandam retreinamento, o Flamingo é capaz de adaptar-se rapidamente a novos domínios a partir de um número reduzido de referências. Esse recurso viabiliza aplicações práticas em contextos especializados, como descrição automatizada de imagens médicas e anotação assistida em comércio eletrônico, onde o estilo e vocabulário de descrição variam conforme o domínio.

Pesquisas recentes têm explorado o uso de *image captioning* como ferramenta auxiliar para aprimorar os modelos multimodais. Propostas como o *Vision-language reinforcement model* utilizam uma abordagem baseada em reforço para refinar a qualidade das legendas, atribuindo recompensas a descrições semanticamente ricas e penalizando redundâncias [Dzabraev et al. 2024]. Em demonstrações práticas, o modelo produziu legendas capazes de expressar elementos subjetivos e emocionais, como tons de humor, clima ou expressões faciais, um passo relevante para tecnologias de acessibilidade digital, como o aplicativo *Be my eyes*, que usa *captioning multimodal* para descrever imagens a pessoas com deficiência visual [Be My Eyes 2023].

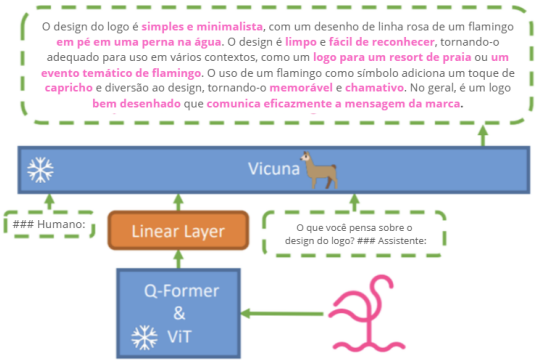
Dessa forma, a tarefa de geração de descrições de imagens não apenas constitui um campo consolidado de aplicações práticas, mas também serve como base metodológica para as demais tarefas multimodais, como análise de sentimentos visuais ou até mesmo VQA, tarefa na qual a etapa de descrição contextual frequentemente antecede a inferência.

1.4.3. Tarefa de VQA

VQA sintetiza o maior objetivo do uso de MLLMs: a compreensão da conexão entre a interpretabilidade da imagem e a geração do questionamento realizado sobre ela, formulado em linguagem natural, integrando o raciocínio visual e textual.

Por exemplo, diante de uma imagem urbana, perguntas como "Quantos carros estão estacionados?" ou "Qual é o estilo arquitetônico predominante?" envolvem tanto percepção visual, na identificação e contagem de elementos, quanto na interpretação semântica para o reconhecimento cultural.

Esses sistemas baseiam-se em *pipelines* compostos por módulos de percepção visual, codificação linguística e fusão semântica, como é possível observar na Figura 1.13. Modelos como o LLaVA [Liu et al. 2023] e MiniGPT-4 [Zhu et al. 2023] exemplificam essa abordagem. Eles utilizam *embeddings* compartilhados que projetam textos e imagens em um espaço vetorial unificado, no qual o raciocínio cruzado ocorre de forma natural.



As aplicações práticas para tarefas de perguntas e respostas são vastas. Na área da saúde, por exemplo, modelos multimodais podem responder perguntas sobre exames de imagem, auxiliando médicos em diagnósticos preliminares [Gong et al. 2025]. Na área de segurança e monitoramento, modelos multimodais podem interpretar cenas em tempo real e identificar comportamentos anômalos [Li et al. 2024]. Já na educação interativa, sistemas de VQA permitem um aprendizado visual dinâmico, no qual estudantes exploram imagens científicas ou históricas por meio de perguntas em linguagem natural [Pal et al. 2025].

Apesar dos avanços no domínio de perguntas e respostas multimodais, ainda persistem desafios relacionados ao viés de raciocínio e à dependência contextual. No que diz respeito ao viés de raciocínio, o modelo demonstra propensão a gerar respostas fun-

damentadas em correlações estatísticas superficiais, preterindo a realização de inferências visuais logicamente estruturadas. Já a dependência contextual refere-se ao fato de que pequenas variações linguísticas na formulação da pergunta podem alterar significativamente a resposta produzida pelo MLLM. Para mitigar essas limitações, estudos recentes têm combinado modelos de raciocínio simbólico com arquiteturas multimodais, buscando integrar a robustez estatística dos MLLMs com capacidades de interpretação lógica. Essa combinação busca conferir ao modelo maior precisão e maior interpretabilidade [Goyal et al. 2017].

1.4.4. Aplicações em Domínios Específicos: Análise de Sentimentos em Imagens (Arcabouço *MLLMsent*)

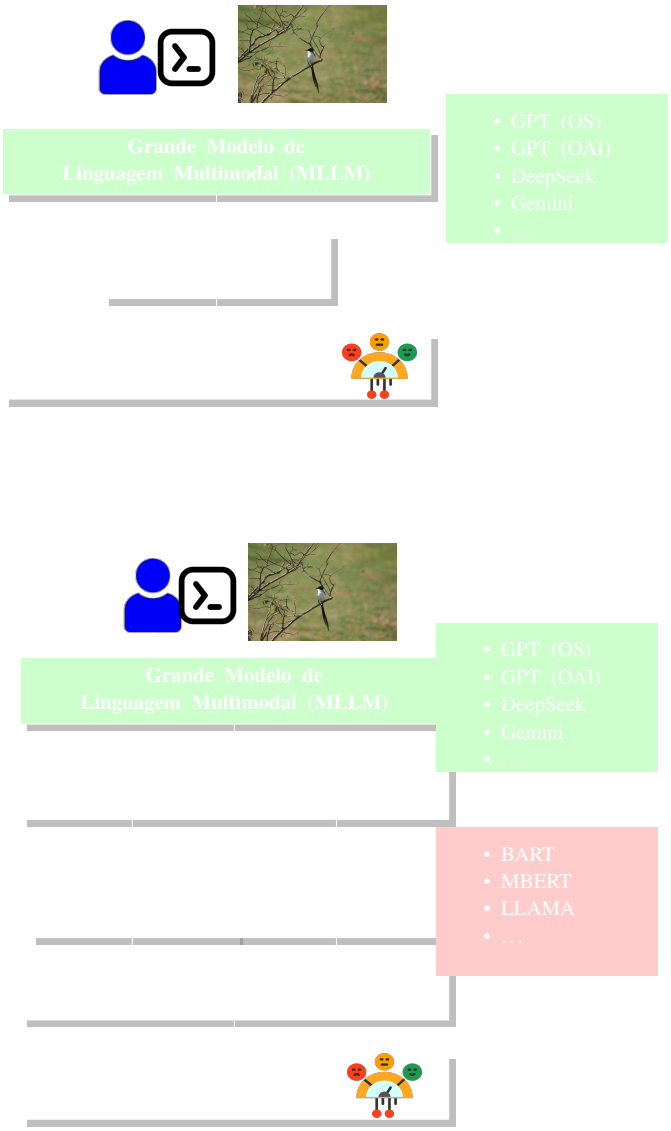
Além das tarefas clássicas, os MLLMs têm demonstrado desempenho promissor em domínios especializados, nos quais a integração entre diferentes modalidades amplia a capacidade de análise de fenômenos complexos. Um desses domínios é a análise de sentimentos em imagens, campo em expansão no contexto da mineração *multimodal* de dados. Nessa abordagem, elementos visuais, como expressões faciais, composição de cores e contexto, são combinados a informações textuais, como legendas ou comentários, para inferir o tom emocional de uma imagem. Estudos indicam, por exemplo, que a correlação entre a paleta cromática de uma imagem e o tipo de emoção expressa no texto associado pode revelar padrões culturais e psicológicos relevantes [Borth et al. 2013].

Diante desse cenário, torna-se relevante compreender como os MLLMs têm aprimorado essa tarefa, substituindo abordagens puramente visuais por métodos capazes de integrar visão e linguagem de forma mais interpretável e contextualizada. A seguir, apresenta-se uma visão geral sobre os principais avanços na análise de sentimentos em imagens e o papel dos MLLMs nesse processo.

Diferente da classificação de objetos, a **análise de sentimentos em imagens** exige que o modelo compreenda o contexto, a emoção e as nuances transmitidas visualmente [Oliveira et al. 2020, Lopes et al. 2023]. Trabalhos recentes, como o arcabouço *MLLMsent* proposto por [da Silva et al. 2025], investigam arquiteturas para solucionar este problema, focando em uma questão central: é mais eficaz classificar o sentimento diretamente da imagem ou a partir de uma descrição textual gerada pelo modelo?

Para responder a isso, o arcabouço explora duas abordagens conceituais distintas, conforme ilustrado na arquitetura da Figura 1.14:

- **Abordagem 1: Classificação Direta (*End-to-End*)**: Nesta arquitetura (diagrama (a) da Figura 1.14), o MLLM é tratado como um classificador direto. O modelo recebe a imagem em um *prompt* instrutivo (por exemplo, "Classifique esta imagem como Positiva, Neutra ou Negativa") e deve retornar apenas o rótulo do sentimento. Esta abordagem testa a capacidade do modelo de realizar um raciocínio multimodal complexo e, crucialmente, de seguir instruções de forma concisa.
- **Abordagem 2: Classificação em Dois Estágios (Via Descrição)**: Esta metodologia (diagrama (b) da Figura 1.14) decompõe o problema, unindo diferentes tipos de modelos e tarefas. Primeiro, o MLLM é usado para sua principal competência de "visão e raciocínio", gerando uma descrição textual que captura os elementos e



o contexto da imagem. Em seguida, para classificar o texto gerado, essa descrição é passada para um modelo de linguagem (e.g. um LLM) focado apenas em texto, mas especializado (via *fine tuning*) em análise de sentimento.

Investigações comparativas [da Silva et al. 2025] demonstram que a abordagem em dois estágios tende a ser mais robusta e a apresentar desempenho superior. A qualidade

da descrição textual gerada pelo MLLM no primeiro estágio constitui um fator crítico para o sucesso do método, como ilustrado na Figura 1.15 por meio da análise qualitativa das descrições produzidas. Além disso, o desempenho global é substancialmente aprimorado quando o classificador de texto é ajustado com dados específicos da tarefa.

Essencialmente, esta aplicação mostra como os MLLMs podem ser usados para transformar um problema complexo de visão computacional (análise de sentimento visual) em um problema tratável de Processamento de Linguagem Natural (NLP - *Natural language processing*), usando a capacidade de raciocínio multimodal do modelo para gerar uma representação textual intermediária e rica em contexto.

1.5. *Hands-on*: Tutorial Prático

Após explorar os conceitos e os resultados obtidos pelo arcabouço ***MLLMsent***, esta seção oferece um guia prático para a implementação de duas abordagens de análise de sentimentos. Por questões didáticas, serão consideradas duas etapas: (1) **Implementando a Abordagem 1: Classificação Direta (*end-to-end*)** e (2) **Implementando a Abordagem 2: Classificação em Dois Estágios (Via Descrição)**. A implementação completa e funcional, incluindo o *notebook* de referência mencionado nestes estudos, está disponível no repositório de origem e pode ser acessada pelo seguinte link: https://github.com/neemiasbsilva/MLLMs-Teoria-e-Pratica/tree/main.

1.5.1. Implementando a Abordagem 1: Classificação Direta

A tarefa de classificação direta de sentimentos em imagens utilizando MLLMs representa uma aplicação prática e poderosa dessa tecnologia. O processo, em sua essência, pode ser composto por duas fases críticas: a **instanciação do modelo**, que envolve carregar o MLLM e seus componentes na memória, e a **inferência**, na qual a imagem e um *prompt* específico são processados para gerar uma classificação. Este ensaio explora duas abordagens centrais: o uso de um modelo local de código aberto (*DeepSeek-VL*) e o acesso a uma API proprietária da OpenAI.

1.5.1.1. Usando MLLMs Open-Source: *DeepSeek-VL*

A primeira metodologia envolve o uso do DeepSeek-VL, uma família robusta de modelos de código aberto. Embora um estudo de referência possa utilizar uma versão específica (como a *V2L-small*), a implementação prática muitas vezes é adaptada às restrições de hardware disponíveis (como o uso da versão *VL* padrão). A implementação local exige uma configuração inicial do ambiente de desenvolvimento. O primeiro passo é obter o código-fonte, isso se dá clonando o repositório oficial do modelo (DeepSeek-VL2).

Após o download do código, é recomendável criar um ambiente virtual isolado (usando ferramentas como *conda*, *venv* ou *uv*), a fim de evitar conflitos de dependências. A etapa seguinte consiste na instalação das dependências. O PyTorch, idealmente com suporte a CUDA, representa a dependência central do ambiente, seguido pelos demais pacotes definidos nos arquivos de requisitos do projeto.

Com o ambiente pronto, o próximo passo é carregar o modelo em memória. Este processo foi abstraído no Código 5 para focar na lógica, omitindo as importações e comandos detalhados que podem ser encontrados no *notebook* de implementação (*Classify Sentiment DeepSeek VL*).

```
# 1. Definir os caminhos
MODEL_PATH = "identificador_do_modelo_no_huggingface"
CACHE_DIR = "diretorio_local_para_salvar_os_pesos"

# 2. Carregar o Processador e o Tokenizer
# O 'Processor' prepara tanto o texto quanto as imagens
vl_chat_processor = DeepseekVLV2Processor.from_pretrained(MODEL_PATH)
tokenizer = vl_chat_processor.tokenizer

# 3. Carregar o Modelo de Linguagem Causal
# 'trust_remote_code=True' é necessário para modelos com arquiteturas customizadas
vl_gpt = AutoModelForCausalLM.from_pretrained(
    MODEL_PATH,
    trust_remote_code=True
)
# 4. Preparar o modelo para inferência
# Mover para a GPU (ex: .cuda()) e ativar o modo de avaliação (.eval())
vl_gpt.to(device).eval()
```

É crucial destacar que modelos de alta performance como o *DeepSeek-V2L-small/medium/large* demandam recursos computacionais expressivos. No caso da versão *small* os recursos computacionais são de 80GB de *VRAM*. A escolha do modelo (*MODEL_PATH*) deve ser compatível com o hardware disponível.

Para realizar a inferência, a entrada é formatada como um diálogo. O *prompt* (a pergunta) deve ser cuidadosamente elaborado, instruindo o modelo a focar exclusivamente na classificação (por exemplo, "Analise esta imagem e classifique-a como Negativa, Neutra ou Positiva.") e incluindo uma *tag* especial `<image>` para indicar onde a imagem deve ser inserida. O fluxo de inferência segue a lógica descrita no Código 6.

```
# 1. Definir o prompt e o caminho da imagem
question = "<image>\nAnalise esta imagem..."
image_path = "caminho/para/imagem.jpg"

# 2. Estruturar a conversa (lista de dicionários)
conversation = [ {"role": "<|User|>", "content": question, "images": [image_path]},
    {"role": "<|Assistant|>", "content": ""} ]

# 3. Carregar as imagens (ex: usando a função load_pil_images)
pil_images = load_pil_images(conversation)

# 4. Preparar os inputs
# O processador converte o diálogo e as imagens em tensores
prepare_inputs = vl_chat_processor(
    conversations=conversation,
    images=pil_images
).to(vl_gpt.device)

# 5. Executar a geração (com torch.no_grad() para otimização)
# Otimizações como 'incremental_prefilling' podem ser usadas
outputs = vl_gpt.generate(
    inputs_embeds=...,
    attention_mask=...,
    max_new_tokens=512,
    do_sample=False )

# 6. Decodificar a resposta
# Converter os tokens de saída de volta para texto
answer = tokenizer.decode(outputs[0])
print(answer)
```

Este processo completo, desde o carregamento até a geração, oferece controle total sobre o *pipeline*, mas depende inteiramente da capacidade do hardware local.

1.5.1.2. Usando MLLMs Proprietários: OpenAI (GPT-4o)

Uma alternativa que abstrai a complexidade de hardware é a utilização de APIs proprietárias, como a oferecida pela *OpenAI* (utilizando modelos como o GPT-4o). Esta aborda-

gem elimina a necessidade de *VRAM* robusta, mas introduz custos operacionais baseados no consumo de *tokens*.

A configuração é significativamente mais simples. Após criar uma conta na plataforma e gerar uma chave de API, ela deve ser configurada (preferencialmente como uma variável de ambiente, por exemplo, `OPEN_API_KEY`). A interação com o modelo é então mediada por um cliente oficial. O Código 7 ilustra a lógica da chamada à API:

```python
# 1. Instanciar o cliente
# O cliente usará automaticamente a variável de ambiente OPEN_API_KEY
client = OpenAI()

# 2. Codificar a imagem
# Define-se uma função que lê a imagem e a converte para base64
base64_image = encode_image_to_base64("caminho/para/imagem.jpg")

# 3. Definir o prompt
prompt = "Analise esta imagem... selecione apenas uma classe."

# 4. Realizar a chamada de 'Chat Completion'
response = client.chat.completions.create(
    model="gpt-4o-mini",
    messages=[
        {
            "role": "user",
            "content": [
                # O prompt de texto
                {"type": "text", "text": prompt},

                # A imagem codificada
                {
                    "type": "image_url",
                    "image_url": {"url": f"{base64_image}"}
                },
            ],
        }
    ],
    max_tokens=50 # Limita o tamanho da resposta
)

# 5. Extrair a resposta
answer = response.choices[0].message.content
print(answer)
```

Esta abordagem é ideal para prototipagem rápida, aplicações *web* ou cenários onde a aquisição de infraestrutura de GPU é inviável. Contudo, é importante levar em consideração o custo de *prompt tokens* ou *completion tokens*, que pode ser consultado na documentação oficial da API.

Estes dois cenários detalharam os fluxos de trabalho conceituais para a classificação direta de sentimentos aplicando MLLMs, contrastando a abordagem local (*DeepSeek-VL)*, que oferece controle e soberania sobre o modelo ao custo de aquisição de hardware

específico, em relação à abordagem via API (*OpenAI*), que oferece simplicidade e escalabilidade ao custo de dependência de serviços de terceiros.

1.5.2. Implementando a Abordagem 2: Classificação em Dois Estágios (Via Descrição da Imagem)

A segunda abordagem, conhecida como Classificação via Descrições, adota uma estratégia de "dividir para conquistar", transformando um problema multimodal complexo em uma tarefa sequencial de Processamento de Linguagem Natural.

Nesta arquitetura, o *pipeline* é composto por duas etapas distintas:

1. **Geração de Descrição**: Um MLLM analisa a imagem de entrada e gera uma descrição textual detalhada.
2. **Classificação Textual:** Um modelo de linguagem focado exclusivamente em texto, que foi treinado especificamente para análise de sentimento, recebe a descrição gerada e a classifica com um rótulo (Negativo, Neutro ou Positivo).

Esta subseção foca inteiramente na segunda etapa: o processo de ajuste fino do classificador textual. O *ModernBERT* será o modelo de NLP e, para treiná-lo, será utilizado um *pipeline* robusto. Este *pipeline* inclui configuração de ambiente, carregamento de dados (descrições e rótulos), definição da arquitetura do modelo e a execução de um treinamento validado por validação cruzada K-*fold*, com monitoramento de métricas via *TensorBoard*.

Para destacar a lógica e os conceitos principais, apresentam-se a seguir apenas os trechos de código mais relevantes. O código-fonte completo e executável encontra-se disponível no repositório oficial do arcabouço: MLLMsent-framework.

Antes de qualquer treinamento, um *pipeline* de aprendizado de máquina requer uma configuração estruturada. O processo começa pela definição de todos os hiperparâmetros e caminhos do experimento.

Isso é comumente gerenciado por um arquivo de configuração (`config.yaml`), que armazena informações como a taxa de aprendizado, o tamanho do lote, o número de épocas e os caminhos para o modelo pré-treinado (neste caso, `answerdotai/ModernBERT-large`. O fluxo de preparação de dados segue a lógica apresentada no Código 8.

```
# 1. Criar a estrutura de diretórios necessária
mkdir "experiments-finetuning/logs"
mkdir "checkpoints"

# 2. Definir e salvar o arquivo 'config.yaml'
CONFIG = {
     "experiment_name": "Experiment Using ModernBERT",
     "learning_rate": 1e-5,
     "batch_size": 4,
     "epochs": 5,
     "model_path": "answerdotai/ModernBERT-large",
     "log_dir": "..." }
save_config_to_yaml(CONFIG)
```

```

# 3. Carregar a configuração para o script Python
config = load_config("path/to/config.yaml")

# 4. Preparar o diretório de dados
mkdir "data/gpt4-openai-classify"

# 5. Baixar o conjunto de dados das descrições (Ex: Usar gdown para baixar de um ID do Google Drive)
gdown.download(id=GDRIVE_ID, output=DATA_FILE_PATH)

# 6. Carregar os dados em um DataFrame
df = load_experiment_data(config)
```

O conjunto de dados, um arquivo `.csv`, contém essencialmente duas colunas: uma com as **descrições textuais** geradas pelo MLLM e outra com os **rótulos de sentimento** correspondentes.

Para que o modelo *ModernBERT* possa processar texto, são necessárias duas classes principais do *PyTorch*: `Dataset` e `DataLoader`, conforme pode ser visto no Código 9. A classe `CustomSentimentDataset` é o coração da preparação de dados. Sua responsabilidade é pegar uma linha do DataFrame (um texto e um rótulo) e convertê-la em tensores numéricos que o *ModernBERT* entende.

```

class CustomSentimentDataset(Dataset):
    def __init__(self, dataframe, tokenizer, max_len):
        self.texts = dataframe.text.values
        self.targets = dataframe.sentiment.values
        self.tokenizer = tokenizer
        self.max_len = max_len

def __getitem__(self, index):
    text = self.texts[index]
    # O Tokenizador converte o texto em IDs, máscara de atenção, etc.
    inputs = self.tokenizer.encode_plus(
        text,
        max_length=self.max_len,
        padding='max_length',  # Garante que todas as sentenças tenham o mesmo tamanho
        truncation=True        # Trunca sentenças longas
    )
    # Retorna um dicionário de tensores
    return { 'ids': tensor(inputs['input_ids']),
        'mask': tensor(inputs['attention_mask']),
        'targets': tensor(self.targets[index]) }
```

O `DataLoader` é então usado para agrupar esses itens em lotes de forma eficiente, embaralhando os dados de treinamento a cada época.

Após a definição e preparação do conjunto de dados, é possível instanciar o *ModernBERT* e ir para a etapa de treinamento, conforme ilustrado no Código 10. Em vez de um treinamento completo, utiliza-se o aprendizado por transferência (*transfer learning*). Para isso, carrega-se o *ModernBERT* pré-treinado e adiciona-se uma "cabeça" de classificação customizada no topo.

```
class ModernBERTModel(nn.Module):
def __init__(self, model_path, num_classes):
    # 1. Carrega o "corpo" (body) pré-treinado do BERT
    self.model = AutoModel.from_pretrained(model_path)
    # 2. Define uma "cabeça" (head) de classificação
    self.classifier = nn.Sequential(
        nn.Linear(self.model.config.hidden_size, 1024),
        nn.ReLU(),
        nn.Linear(1024, num_classes) # Projeta para o número de classes (3)
    )
def forward(self, ids, mask):
    # 1. Passa os dados pelo corpo do BERT
    output = self.model(ids, attention_mask=mask)
    # 2. Extrai a representação do token [CLS] (primeiro token)
    # Este token é projetado para conter o significado agregado da sentença
    CLS_token_state = output.last_hidden_state[:, 0, :]
    # 3. Passa o [CLS] pela cabeça de classificação
    logits = self.classifier(CLS_token_state)
    return logits
```

O processo de treinamento é encapsulado em várias funções auxiliares:

- `train_one_epoch`: Itera sobre os lotes de treinamento. Para cada lote, ela move os dados para a GPU, zera os gradientes, executa o *forward pass* (propagação direta), calcula a *loss* (perda), executa o *backward pass* (calcula os gradientes) e atualiza os pesos do modelo (`opmizer.step()`).
- `validate_one_epoch`: Faz o mesmo processo, mas no conjunto de validação e dentro de um contexto `torch.no_grad()`, o que desabilita o cálculo de gradientes para economizar memória e garantir que o modelo não "aprenda" com os dados de validação.
- `fit`: função que orquestra o treinamento de um único *fold*. Ela inicializa o `SummaryWriter` do *TensorBoard* (para a visualização das métricas), define a função de perda (como `CrossEntropyLoss`) e implementa a lógica de parada antecipada (*early stopping*). A cada época, ela chama `train_one_epoch` e `validation_one_epoch`, registra as métricas e verifica se o *F1-score* de validação melhorou. Se não houver melhoria por um número definido de épocas ("paciência"), o treinamento é interrompido.

- `val`: Função chamada após o término do `fit` para registrar os resultados finais daquela *fold* específica.

A função mais importante é a `train`, que gerencia todo o processo de validação cruzada K-*fold* (neste caso, com K=5), conforme ilustrado no Código 11. Esta técnica é crucial para obter uma estimativa robusta da performance do modelo.

```
def train(config):
# 1. Carregar configuração e definir o dispositivo (GPU/CPU)
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")

# 2. Carregar o conjunto de dados completo
df = load_experiment_data(...)

# 3. Inicializar o KFold (ex: 5 splits)
kfold = KFold(n_splits=5, shuffle=True, random_state=42)
df_metrics = pd.DataFrame([]) # Para salvar os resultados de cada fold
best_f1_global = 0

# 4. Loop principal do K-Fold
for fold, (train_idx, val_idx) in enumerate(kfold.split(df)):
    print(f"========== FOLD {fold + 1} / 5 ==========")

    # 5. Instanciar um *novo* modelo, tokenizador e otimizador
    # É crucial que cada fold treine um modelo do zero
    model = ModernBERTModel(...).to(device)
    tokenizer = AutoTokenizer.from_pretrained(...)
    optimizer = AdamW(model.parameters(), lr=config["learning_rate"])

    # 6. Dividir os dados para este fold
    train_df = df.iloc[train_idx]
    val_df = df.iloc[val_idx]

    # 7. (Importante) Calcular pesos das classes
    # Para lidar com desbalanceamento (ex: mais rótulos 'Neutro' que 'Negativo'),
    # calculam-se os pesos para a função de perda.
    class_weights = calculate_class_weights(train_df).to(device)

    # 8. Criar DataLoaders para este fold
    train_dl = data_loader(train_df, ...)
    val_dl = data_loader(val_df, ...)

    # 9. Chamar a função 'fit' para treinar o modelo deste fold
    model, loss_fn = fit(model, class_weights, config["epochs"], optimizer, ...)

    # 10. Chamar a função 'val' para avaliar o modelo deste fold
    df_metrics, f1_val = val(model, val_dl, loss_fn, fold, ...)

    # 11. Salvar o checkpoint se este for o melhor modelo *global*
    if f1_val > best_f1_global:
        best_f1_global = f1_val
        save_checkpoint(model, ...)

    # 12. Limpar a memória da GPU para o próximo fold
    del model
    torch.cuda.empty_cache()
```

```

# 13. Após todos os folds, calcular e imprimir as métricas finais
mean_f1 = df_metrics["f1_score"].mean()
std_f1 = df_metrics["f1_score"].std()
print(f"F1-Score Médio: {mean_f1} ± {std_f1}")
```

Em um ambiente de *notebook*, o TensorBoard é carregado, conforme pode ser visto no Código 12. Isso permite o monitoramento em tempo real das curvas de aprendizado (perda, acurácia, F1-score) para cada *fold* do treinamento, fornecendo *insights* valiosos sobre a estabilidade e convergência do modelo.

```
# Carregar o arquivo de configuração
config_path =
↪  "experiments-finetuning/openai-modernbert-experiment-p3-alpha3/config.yaml"
config = load_config(config_path)

# Iniciar o processo completo de treinamento K-Fold
train(config, config_path)

# Carregar a extensão do TensorBoard
%load_ext tensorboard

# Iniciar o TensorBoard
# Ele vai monitorar o diretório de logs definidos no config.yaml
%tensorboard --logdir
↪  experiments-finetuning/openai-modernbert-experiment-p3-alpha3/logs
```

1.5.3. Implementações completas de exemplos práticos

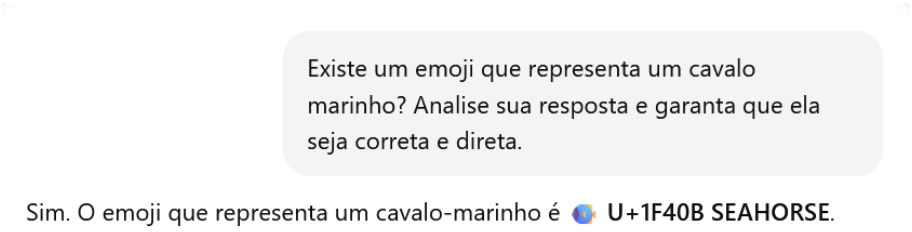
O código para o *MLLMSent* e outros exemplos práticos está disponível no repositório do GitHub deste trabalho.. No momento da escrita, ele contém exemplos para *fine tuning* com *ModernBERT*, identificação de objetos e RAG com LangGraph.

1.6. Limitações e Oportunidades

Apesar de os MLLMs demonstrarem grande capacidade de atuar em diferentes tarefas, como geração e interpretação de linguagens e imagens, estes ainda possuem limitações inerentes que restringem sua confiabilidade e capacidade de generalização. A limitação mais recorrente desses modelos é a tendência à alucinação, isto é, à geração de respostas aparentemente plausíveis, porém factualmente incorretas, frequentemente apresentadas com excesso de confiança [Kalai et al. 2025].

A Figura 1.16 contém um exemplo do fenômeno de alucinação em um modelo de linguagem; o modelo foi questionado sobre a existência de um *emoji* para "cavalo-marinho". Embora tal *emoji* não exista[4], o modelo gerou uma resposta incorreta com alta

confiança, apresentando o *emoji* de um peixe junto ao código *Unicode* também incorreto. Esse tipo de erro evidencia a necessidade de cautela e verificação humana ao interpretar suas respostas, especialmente em contextos que exigem precisão factual.



Pesquisadores da OpenAI, empresa responsável pela criação dos modelos GPT, explicam no artigo *"Why language models hallucinate"* [Kalai et al. 2025] que o fenômeno de alucinação ocorre porque os procedimentos de treinamento e avaliação dos modelos de linguagem recompensam a produção de respostas plausíveis, mesmo sob incerteza, em vez de incentivar o reconhecimento explícito dessa incerteza. Tecnicamente, isso se relaciona a um problema de calibração, no qual a confiança expressa pelo modelo não corresponde de forma adequada à probabilidade real de acerto. Como analogia didática, pode-se comparar esse comportamento ao de um estudante que, diante de uma questão difícil, opta por arriscar uma resposta que parece correta em vez de admitir que não sabe. Além disso, os métodos atuais de avaliação de LLMs geralmente penalizam respostas que expressam incerteza, o que incentiva a geração de respostas confiantes, ainda que incorretas. Portanto, a persistência das alucinações em MLLMs pode ser compreendida como resultado de um desalinhamento entre os incentivos de treinamento e a necessidade de respostas calibradas, precisas e honestas.

As alucinações também podem ocorrer em tarefas que envolvem o processamento de imagens. Mesmo em situações em que os MLLMs compreendem corretamente o conteúdo de uma imagem, eles frequentemente geram respostas incorretas. Segundo [Liu et al. 2025], esse problema ocorre porque os modelos apresentam baixa atenção aos *tokens* visuais e são mais suscetíveis a erros quando confrontados com perguntas indiretas ou enganosas. Para mitigar esse tipo de alucinação, são propostas duas estratégias: o Refinamento Guiado por Conteúdo (*Content guided refinement*), em que o modelo é orientado a realizar uma análise preliminar do conteúdo visual antes de formular a resposta, e o Refinamento de Atenção Visual (*Visual attention refinement*), que destaca áreas da imagem mais relevantes para a pergunta e oculta parcialmente ou totalmente as áreas irrelevantes.

Embora tenham ocorrido avanços em modelos multilíngues, há evidências de que LLMs frequentemente perdem desempenho ou consistência quando operam em idiomas que não o inglês [Mondshine et al. 2025], o que motivou o desenvolvimento de um questionário (ECLeKTic [Goldman et al. 2025]) para avaliar a capacidade de transferência de conhecimento interlíngua dos LLMs. Os resultados revelam que modelos de estado da arte têm dificuldade de transferir conhecimento factual entre idiomas, ou seja, eles acertam bem em perguntas no idioma em que aprenderam o fato, mas falham quando a mesma

pergunta é traduzida para outro idioma.

À primeira vista, a enorme quantidade de dados disponíveis na internet poderia sugerir que um conjunto de dados obtido de múltiplas fontes representa adequadamente a diversidade de pensamentos e visões de mundo. No entanto, segundo [Bender et al. 2021], diversos fatores restringem a participação e a visibilidade de determinados grupos nesse espaço virtual, afetando a heterogeneidade dos dados disponíveis na internet. O acesso à internet é desigual, sendo mais comum entre jovens de países desenvolvidos[5], o que leva à super-representação desses perfis. Além disso, plataformas populares como *Wikipedia*, *Reddit* e X (*Twitter*), embora aparentem ser espaços abertos, possuem dinâmicas e práticas de moderação que frequentemente marginalizam certos grupos sociais. Isso cria um ciclo de exclusão em que apenas determinadas vozes continuam a ser amplificadas, enquanto comunidades menos representadas têm sua produção textual negligenciada ou excluída dos grandes conjuntos de dados usados no treinamento de modelos de linguagem.

Segundo [Jegham et al. 2025], o treinamento do GPT-3 consumiu aproximadamente 1.287 MWh de eletricidade, emitiu 550 toneladas de CO_2 e utilizou 700 mil litros de água apenas para resfriamento. Ainda mais relevante, a fase de inferência, que ocorre continuamente após o treinamento, pode representar até 90% do consumo energético total de um modelo. É estimado que cada consulta realizada na versão GPT-4o, consuma 0,42 Wh, e, ao ser escalado para 700 milhões de consultas diárias, equivale a um gasto anual de 391 a 463 mil MWh, emissões de 138 a 163 mil toneladas de CO_2 e o uso de 1,3 a 1,6 milhão de litros de água. Essas quantidades correspondem aproximadamente ao consumo elétrico de 35 mil residências e à água potável anual de 1,2 milhão de pessoas.

Entre as abordagens que visam contornar o alto custo de treinamento e inferência dos modelos de linguagem de larga escala, destaca-se o método *LoRA* (*Low-rank adaptation of large language models*), proposto por [Hu et al. 2022]. Essa técnica consiste em congelar os pesos do modelo-base e introduzir matrizes de baixa dimensão nas camadas lineares, permitindo o ajuste fino do modelo sem a necessidade de modificar os parâmetros originais. Dessa forma, o *LoRA* possibilita uma redução significativa no número de parâmetros treináveis, resultando em menor consumo de memória e tempo de treinamento. Além disso, o método favorece a modularidade do processo de adaptação, permitindo combinar múltiplos ajustes específicos de tarefas distintas sem a necessidade de treinar novamente o modelo principal.

Esforços têm sido dirigidos para reduzir a dependência das grandes empresas que dominam o desenvolvimento e a infraestrutura dos modelos multimodais de linguagem. Resultando na popularização de MLLMs de código aberto, impulsionada pela melhoria de desempenho, pela transparência e pela possibilidade de execução local. Modelos como LLaVA, MiniCPM-V e Qwen2-VL-7B, oferecem capacidades multimodais competitivas frente a modelos proprietários de grande escala. Paralelamente, há um movimento concentrado na criação de modelos menores e mais eficientes, projetados para operar em computadores pessoais e dispositivos embarcados, sem necessidade de infraestrutura em nuvem.

Essas iniciativas são sustentadas por avanços em técnicas de compressão e eficiência, como quantização [Frantar et al. 2022, Xiao et al. 2023], que reduz a precisão dos

pesos e ativações (por exemplo, de 32 para 8 ou 4 bits), diminuindo o tamanho do modelo e acelerando a inferência; e destilação de conhecimento [Hinton et al. 2015], processo em que um modelo grande transfere suas representações e comportamento para um modelo menor, mantendo parte significativa do desempenho original. Combinados à crescente disponibilidade de hardware otimizado para redes neurais e *Transformers*, como unidades de processamento neural (NPUs do inglês *neural processing units*) e aceleradores integrados, esses recursos impulsionam a descentralização dos MLLMs, ampliando sua acessibilidade, privacidade e eficiência.

Nos avanços mais recentes no âmbito do processamento de vídeos por MLLMs, observa-se uma tendência de expansão da compreensão multimodal, que ultrapassa a análise isolada de vídeo ou áudio e passa a integrar ambas as modalidades em um único arcabouço. O *Video-LLaMA* [Zhang et al. 2023], integra *encoders* pré-treinados e congelados de imagem e som com módulos de alinhamento (*Q-Formers*), possibilitando uma interpretação conjunta e temporalmente coerente de sinais visuais e auditivos. Já o *Video-of-Thought* [Fei et al. 2024] introduz um modelo de raciocínio hierárquico e progressivo, inspirado no *Chain-of-Thought* [Wei et al. 2022b], que estrutura a compreensão de vídeos em etapas sucessivas: percepção em nível de pixel, análise semântica e inferência fundamentada em conhecimento de senso comum.

Complementarmente, o *VideoEspresso* [Han et al. 2025] introduz um conjunto de dados em larga escala com anotações multimodais de raciocínio, permitindo treinar modelos que relacionam coerentemente quadros, objetos e linguagem ao longo do tempo. Já o *MetaMind* [Zhang et al. 2025] amplia o horizonte para a metacognição e interação social, utilizando múltiplos agentes (Teoria da Mente, Moral e Resposta) para inferir intenções e emoções humanas a partir de vídeos. Em conjunto, esses trabalhos indicam que as próximas gerações de MLLMs tenderão a incorporar raciocínio multimodal, cognitivo e social, aproximando-se de representações mais humanas de percepção, contexto e comportamento.

À medida que os MLLMs se difundem e encontram aplicação em uma variedade de domínios científicos, tecnológicos e sociais, o avanço de sua capacidade de resolver problemas cada vez mais complexos tem tornado os algoritmos, os conjuntos de dados e as metodologias empregados em sua construção cada vez mais sofisticados. Como resultado, há modelos em que a entrada e saída de dados e informações são conhecidas, mas o processo de decisão dos MLLMs não é transparente nem facilmente interpretável. Em outras palavras, esses modelos são análogos a sistemas de caixa-preta e, diante disso, assegurar transparência, responsabilidade e justiça em seu funcionamento tornou-se um desafio significativo. Nesse contexto, uma tendência importante é o desenvolvimento da Inteligência Artificial Explicável (xAI do inglês *explainable AI*), que propõe técnicas para tornar os modelos mais interpretáveis e explicáveis [Mersha et al. 2024, Longo et al. 2024].

1.7. Conclusões

Este trabalho apresentou os fundamentos dos MLLMs, destacando suas arquiteturas, o funcionamento de *pipelines* completos e os principais *trade-offs* envolvidos em aplicações reais. Foram ainda exploradas técnicas práticas de pré-processamento, engenharia de *prompt* e o uso de arcabouços como LangChain e LangGraph para construir fluxos

multimodais eficientes.

Os exemplos práticos mostraram como MLLMs podem ser aplicados a desafios do mundo real, ao mesmo tempo em que evidenciam suas limitações atuais e tendências futuras, como o avanço de modelos mais eficientes e de agentes multimodais.

Para aprofundar o aprendizado, recomenda-se consultar o repositório do minicurso no GitHub (`https://github.com/neemiasbsilva/MLLMs-Teoria-e-Pratica`), que contém *notebooks* com implementações guiadas e conjuntos de dados para experimentação.

Agradecimentos

Este trabalho foi parcialmente financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (processos 314603/2023-9, 441444/2023-7, 313122/2023-7 e 444724/2024-9), pelo INCT-TILD-IAR e pelo projeto SocialNet (FAPESP processo 2023/00148-0).

Referências

[Alayrac et al. 2022] Alayrac, J.-B., Donahue, J., Luc, P., Miech, A., Barr, I., Hasson, Y., Lenc, K., Mensch, A., Millican, K., Reynolds, M., et al. (2022). Flamingo: a visual language model for few-shot learning. *Advances in neural information processing systems*, 35:23716–23736.

[Baevski et al. 2020] Baevski, A., Zhou, Y., Mohamed, A., and Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in neural information processing systems*, 33:12449–12460.

[Baltrušaitis et al. 2018] Baltrušaitis, T., Ahuja, C., and Morency, L.-P. (2018). Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence*, 41(2):423–443.

[Be My Eyes 2023] Be My Eyes (2023). Be My Eyes – Be My AI: Accessible visual assistance powered by GPT-4. `https://www.bemyeyes.com/bme-ai`. Acesso em: out. 2025.

[Bender et al. 2021] Bender, E., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? pages 610–623.

[Birhane et al. 2021] Birhane, A., Prabhu, V. U., and Kahembwe, E. (2021). Multimodal datasets: misogyny, pornography, and malignant stereotypes.

[Borth et al. 2013] Borth, D., Ji, R., Chen, T., Breuel, T., and Chang, S.-F. (2013). Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *Proceedings of the 21st ACM International Conference on Multimedia*, MM '13, page 223–232, New York, NY, USA. Association for Computing Machinery.

[Brown et al. 2020] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language

models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

[Caffagni et al. 2024] Caffagni, D., Cocchi, F., Barsellotti, L., Moratelli, N., Sarto, S., Baraldi, L., Cornia, M., and Cucchiara, R. (2024). The revolution of multimodal large language models: a survey. *arXiv preprint arXiv:2402.12451*.

[Chen et al. 2020] Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PmLR.

[da Silva et al. 2025] da Silva, N. B., Harrison, J., Minetto, R., Delgado, M. R., Nassu, B. T., and Silva, T. H. (2025). Multimodal LLMs see sentiment. *arXiv preprint arXiv:2508.16873*.

[Dai et al. 2024] Dai, W., Lee, N., Wang, B., Yang, Z., Liu, Z., Barker, J., Rintamaki, T., Shoeybi, M., Catanzaro, B., and Ping, W. (2024). NVLM: Open frontier-class multimodal LLMs. *arXiv preprint arXiv:2409.11402*.

[Deitke et al. 2025] Deitke, M., Clark, C., Lee, S., Tripathi, R., Yang, Y., Park, J. S., Salehi, et al. (2025). Molmo and PixMo: Open weights and open data for state-of-the-art vision-language models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 91–104.

[Devlin et al. 2019] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186.

[Dosovitskiy et al. 2020] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

[Driess et al. 2023] Driess, D., Xia, F., Sajjadi, M. S. M., Lynch, C., Chowdhery, A., Ichter, B., Wahid, A., Tompson, J., Vuong, Q., Yu, T., Huang, W., Chebotar, Y., Sermanet, P., Duckworth, D., Levine, S., Vanhoucke, V., Hausman, K., Toussaint, M., Greff, K., Zeng, A., Mordatch, I., and Florence, P. (2023). PaLM-E: An embodied multimodal language model.

[Dzabraev et al. 2024] Dzabraev, M., Kunitsyn, A., and Ivaniuta, A. (2024). VLRM: Vision-language models act as reward models for image captioning. *arXiv preprint arXiv:2404.01911*.

[Elizalde et al. 2023] Elizalde, B., Deshmukh, S., Al Ismail, M., and Wang, H. (2023). CLAP learning audio concepts from natural language supervision. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE.

[Fei et al. 2024] Fei, H., Wu, S., Ji, W., Zhang, H., Zhang, M., Lee, M.-L., and Hsu, W. (2024). Video-of-thought: Step-by-step video reasoning from perception to cognition. *arXiv preprint arXiv:2501.03230*.

[Frantar et al. 2022] Frantar, E., Ashkboos, S., Hoefler, T., and Alistarh, D. (2022). GPTQ: Accurate post-training quantization for generative pre-trained transformers. *arXiv preprint arXiv:2210.17323*.

[Goldman et al. 2025] Goldman, O., Shaham, U., Malkin, D., Eiger, S., Hassidim, A., Matias, Y., Maynez, J., Gilady, A. M., Riesa, J., Rijhwani, S., et al. (2025). ECLeKTic: a novel challenge set for evaluation of cross-lingual knowledge transfer. *arXiv preprint arXiv:2502.21228*.

[Gong et al. 2025] Gong, L., Yang, J., Han, S., and Ji, Y. (2025). MedBLIP: A multimodal method of medical question-answering based on fine-tuning large language model. *Computers in Medical Imaging and Graphics*, 124:102581.

[Goyal et al. 2017] Goyal, Y., Khot, T., Summers-Stay, D., Batra, D., and Parikh, D. (2017). Making the V in VQA matter: Elevating the role of image understanding in visual question answering. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

[Grattafiori et al. 2024] Grattafiori, A., Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Vaughan, A., et al. (2024). The LLaMA 3 herd of models. *arXiv preprint arXiv:2407.21783*.

[Han et al. 2025] Han, S., Huang, W., Shi, H., Zhuo, L., Su, X., Zhang, S., Zhou, X., Qi, X., Liao, Y., and Liu, S. (2025). Videoespresso: A large-scale chain-of-thought dataset for fine-grained video reasoning via core frame selection. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 26181–26191.

[Hinton et al. 2015] Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.

[Hossain et al. 2019] Hossain, M. D., Sohel, F., Shiratuddin, M. F., and Laga, H. (2019). A comprehensive survey of deep learning for image captioning. *ACM Computing Surveys (CSUR)*, 51(6):1–36.

[Hu et al. 2022] Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al. (2022). Lora: Low-rank adaptation of large language models. *ICLR*, 1(2):3.

[Jegham et al. 2025] Jegham, N., Abdelatti, M., Elmoubarki, L., and Hendawi, A. (2025). How hungry is AI? benchmarking energy, water, and carbon footprint of LLM inference. *arXiv preprint arXiv:2505.09598*.

[Kalai et al. 2025] Kalai, A. T., Nachum, O., Vempala, S. S., and Zhang, E. (2025). Why language models hallucinate. *arXiv preprint arXiv:2509.04664*.

[Krishna et al. 2017] Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., Chen, S., Kalantidis, Y., Li, L.-J., Shamma, D. A., Bernstein, M. S., and Fei-Fei, L. (2017). Visual genome: Connecting language and vision using crowdsourced dense image annotations. *International Journal of Computer Vision (IJCV)*, 123(1):32–73.

[Kuang et al. 2025] Kuang, J., Shen, Y., Xie, J., Luo, H., Xu, Z., Li, R., Li, Y., Cheng, X., Lin, X., and Han, Y. (2025). Natural language understanding and inference with MLLM in visual question answering: A survey. *ACM Computing Surveys*, 57(8):1–36.

[LangChain Inc. 2025a] LangChain Inc. (2025a). LangChain. `https://www.langchain.com/`. Acesso em: 16 out. 2025.

[LangChain Inc. 2025b] LangChain Inc. (2025b). LangGraph. `https://www.langchain.com/langgraph`. Acesso em: 16 out. 2025.

[LangChain Inc. 2025c] LangChain Inc. (2025c). LangSmith. `https://www.langchain.com/langsmith/observability`. Acesso em: 16 out. 2025.

[Lewis et al. 2020] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in neural information processing systems*, 33:9459–9474.

[Li et al. 2024] Li, J., Bercea, C. I., Müller, P., Felsner, L., Kim, S. H., Rueckert, D., Wiestler, B., and Schnabel, J. A. (2024). Multi-image visual question answering for unsupervised anomaly detection. In *CoRR*, volume abs/3sJvMF5m5D. Preprint on OpenReview.

[Li et al. 2023] Li, J., Li, D., Savarese, S., and Hoi, S. (2023). BLIP-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *International conference on machine learning*, pages 19730–19742. PMLR.

[Liu et al. 2023] Liu, H., Li, C., Wu, Q., and Lee, Y. J. (2023). Visual instruction tuning. *Advances in neural information processing systems*, 36:34892–34916.

[Liu et al. 2025] Liu, Y., Liang, Z., Wang, Y., Wu, X., Tang, F., He, M., Li, J., Liu, Z., Yang, H., Lim, S., et al. (2025). Unveiling the ignorance of mLLMs: Seeing clearly, answering incorrectly. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 9087–9097.

[Longo et al. 2024] Longo, L., Brcic, M., Cabitza, F., Choi, J., Confalonieri, R., Ser, J. D., Guidotti, R., Hayashi, Y., Herrera, F., Holzinger, A., Jiang, R., Khosravi, H., Lecue, F., Malgieri, G., Páez, A., Samek, W., Schneider, J., Speith, T., and Stumpf, S. (2024). Explainable artificial intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. *Information Fusion*, 106:102301.

[Lopes et al. 2023] Lopes, C. R., Minetto, R., Delgado, M. R., and Silva, T. H. (2023). PerceptSent - exploring subjectivity in a novel dataset for visual sentiment analysis. *IEEE Transactions on Affective Computing*, 14(3).

[Mersha et al. 2024] Mersha, M., Lam, K., Wood, J., AlShami, A. K., and Kalita, J. (2024). Explainable artificial intelligence: A survey of needs, techniques, applications, and future direction. *Neurocomputing*, 599:128111.

[Mondshine et al. 2025] Mondshine, I., Paz-Argaman, T., and Tsarfaty, R. (2025). Beyond english: The impact of prompt translation strategies across languages and tasks in multilingual LLMs. *arXiv preprint arXiv:2502.09331*.

[Oliveira et al. 2020] Oliveira, W. B. d., Dorini, L. B., Minetto, R., and Silva, T. H. (2020). OutdoorSent: Sentiment analysis of urban outdoor images by using semantic and deep features. *ACM Trans. on Information Systems*, 38(3).

[OpenAI 2023a] OpenAI (2023a). GPT-4 technical report. *arXiv preprint arXiv:2303.08774*.

[OpenAI 2023b] OpenAI (2023b). GPT-4V(ision) system card. `https://cdn.openai.com/papers/GPTV_System_Card.pdf`. Acesso em: 3 dez. 2025.

[Pal et al. 2025] Pal, R., Kar, S., Prasad, D. K., and Sekh, A. A. (2025). Multilingual visual question answering for visually impaired people. *Discover Artificial Intelligence*, 5.

[Radford et al. 2021] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. (2021). Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR.

[Radford et al. 2018] Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., et al. (2018). Improving language understanding by generative pre-training. *OpenAI blog*.

[Radford et al. 2019] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

[Raschka 2024] Raschka, S. (2024). Understanding Multimodal LLMs. https://magazine.sebastianraschka.com/p/understanding-multimodal-llms. Acesso em: 16 out. 2025.

[Team et al. 2023] Team, G., Anil, R., Borgeaud, S., Alayrac, J.-B., Yu, J., Soricut, R., Schalkwyk, J., Dai, A. M., Hauth, A., Millican, K., et al. (2023). Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.

[Touvron et al. 2023] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al. (2023). LLaMA: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

[Vaswani et al. 2017] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

[Vinyals et al. 2015] Vinyals, O., Toshev, A., Bengio, S., and Erhan, D. (2015). Show and tell: A neural image caption generator. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3156–3164.

[Wang et al. 2022] Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdhery, A., and Zhou, D. (2022). Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171*.

[Wei et al. 2022a] Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., et al. (2022a). Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*.

[Wei et al. 2022b] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., Zhou, D., et al. (2022b). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

[Wolf et al. 2020] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., et al. (2020). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, pages 38–45.

[Wolfe 2024] Wolfe, C. R. (2024). Decoder-only transformers: The workhorse of generative LLMs. `https://cameronrwolfe.substack.com/p/decoder-only-transformers-the-workhorse`. Acesso em: out. 2025.

[Wu et al. 2023] Wu, J., Gan, W., Chen, Z., Wan, S., and Yu, P. S. (2023). Multimodal large language models: A survey. In *2023 IEEE International Conference on Big Data (BigData)*, pages 2247–2256. IEEE.

[Xiao et al. 2023] Xiao, G., Lin, J., Seznec, M., Wu, H., Demouth, J., and Han, S. (2023). SmoothQuant: Accurate and efficient post-training quantization for large language models. In *International conference on machine learning*, pages 38087–38099. PMLR.

[Yao et al. 2022] Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K. R., and Cao, Y. (2022). ReAct: Synergizing reasoning and acting in language models. In *The eleventh international conference on learning representations*.

[Yin et al. 2024] Yin, S., Fu, C., Zhao, S., Li, K., Sun, X., Xu, T., and Chen, E. (2024). A survey on multimodal large language models. *National Science Review*, 11(12):nwae403.

[Zhang et al. 2023] Zhang, H., Li, X., and Bing, L. (2023). Video-LLaMA: An instruction-tuned audio-visual language model for video understanding. *arXiv preprint arXiv:2306.02858*.

[Zhang et al. 2025] Zhang, X., Chen, Y., Yeh, S., and Li, S. (2025). MetaMind: Modeling human social thoughts with metacognitive multi-agent systems. *arXiv preprint arXiv:2505.18943*.

[Zhu et al. 2023] Zhu, D., Chen, J., Shen, X., Li, X., and Wang, X. (2023). MiniGPT-4: Enhancing vision-language understanding with advanced large language models. *arXiv preprint arXiv:2304.10592*.

A. Exemplos de *pipelines* avançados

A.1. *Pipeline* com o arcabouço LangChain

Essa seção apresenta uma possível implementação das etapas funcionais da *chain* e a criação da cadeia sequencial no LangChain.

```python
from typing import TypedDict, Optional, List, Any
from langchain.schema.runnable import RunnableLambda
from langchain.chains import SequentialChain

class PipelineState(TypedDict, total=False):
    image: Any
    text: str
    roi: Optional[Any]
    ner_entities: dict
    rag_docs: List[dict]
    prompt: str
    model_output: str
    error: Optional[str]
```

```python
def preprocess_node(state: PipelineState) -> PipelineState:
    image = enhance_image(state["image"])
    roi = detect_and_crop_bird(image)
    if roi is None:
        state["error"] = "Nenhuma ave detectada na imagem"
        return state
    state["roi"] = roi
    return state
```

```python
def ner_node(state: PipelineState) -> PipelineState:
    ents = run_ner(state["text"])
    state["ner_entities"] = ents
    return state
```

```python
def rag_node(state: PipelineState) -> PipelineState:
    roi, ents = state["roi"], state["ner_entities"]
    query_embedding = embed_multimodal(roi, ents)
    docs = vector_store_search(query_embedding, filters=ents)
    state["rag_docs"] = docs
    return state
```

```python
def prompt_prep_node(state: PipelineState) -> PipelineState:
    examples = format_as_examples(state["rag_docs"])
    prompt = assemble_prompt(examples, state["ner_entities"])
    state["prompt"] = prompt
    return state
```

```python
def mlmmodel_node(state: PipelineState) -> PipelineState:
    result = call_multimodal_model(
        image=state["roi"],
        prompt=state["prompt"]
    )
    state["model_output"] = result
    return state
```

```python
def error_handler(state: PipelineState) -> PipelineState:
    print("Erro no pipeline:", state.get("error"))
    return state
```

```python
preprocess_chain = RunnableLambda(preprocess_node)
ner_chain = RunnableLambda(ner_node)
rag_chain = RunnableLambda(rag_node)
prompt_chain = RunnableLambda(prompt_prep_node)
model_chain = RunnableLambda(mlmmodel_node)
error_chain = RunnableLambda(error_handler)
```

```
main_chain = SequentialChain(
    chains=[
        preprocess_chain,
        ner_chain,
        rag_chain,
        prompt_chain,
        model_chain
    ],
    input_variables=["image", "text"],
    output_variables=["model_output"]
)
```

```
def run_pipeline(image, text):
    state = {"image": image, "text": text}
    state = preprocess_node(state)
    if state.get("error"):
        return error_handler(state)
    for node in [ner_node, rag_node, prompt_prep_node, mlmmodel_node]:
        state = node(state)
        if state.get("error"):
            return error_handler(state)
    return state
```

```
result = run_pipeline(input_image, input_text)
print("Resultado final:", result.get("model_output"))
```

A.2. *Pipeline* com o arcabouço LangGraph

Nesta seção, apresenta-se uma possível implementação das funcionalidades de cada nó do grafo de estados, a instanciação do grafo, definição de suas conexões, e, por fim, sua execução.

```
from typing import TypedDict, Optional, List, Any
from langgraph.graph import StateGraph, Node
from langgraph.graph import send, conditional_edge  # helpers para fluxo condicional
# (dependendo da versão do LangGraph usada, os nomes das APIs podem variar)

class PipelineState(TypedDict, total=False):
    image: Any                # imagem bruta (e.g. PIL Image, numpy array, etc)
    text: str                 # descrição textual da imagem
    roi: Optional[Any]        # imagem da região de interesse (ave detectada)
```

```
    ner_entities: dict         # entidades extraídas do texto
    rag_docs: List[dict]       # documentos recuperados pelo RAG
    prompt: str                # prompt construído para o MLLM
    model_output: str          # saída final do modelo
    error: Optional[str]       # mensagem de erro, se houver
```

```
def preprocess_node(state: PipelineState) -> PipelineState:
    image = state["image"]
    # Aplica realce de contraste, nitidez, normalização etc
    image = enhance_image(image)
    # Processa a imagem, detecta bounding box da ave, recorta região
    roi = detect_and_crop_bird(image)  # retorna None se não detectar
    if roi is None:
        state["error"] = "Nenhuma ave detectada na imagem"
        return state
    state["roi"] = roi
    return state
```

```
def ner_node(state: PipelineState) -> PipelineState:
    text = state["text"]
    ents = run_ner(text)
    state["ner_entities"] = ents
    return state
```

```
def rag_node(state: PipelineState) -> PipelineState:
    roi = state["roi"]
    ents = state["ner_entities"]
    # gera embedding multimodal (imagem + texto) ou combinação
    query_embedding = embed_multimodal(roi, ents)
    docs = vector_store_search(query_embedding, filters=ents)
    state["rag_docs"] = docs
    return state
```

```python
def prompt_prep_node(state: PipelineState) -> PipelineState:
    docs = state["rag_docs"]
    # transforma os docs em exemplos (texto + imagens associadas)
    examples = format_as_examples(docs)
    prompt = assemble_prompt(examples, state["ner_entities"])
    state["prompt"] = prompt
    return state
```

```python
def mlmmodel_node(state: PipelineState) -> PipelineState:
    image = state["roi"]
    prompt = state["prompt"]
    resp = call_multimodal_model(image=image, prompt=prompt)
    state["model_output"] = resp
    return state
```

```python
def error_node(state: PipelineState) -> PipelineState:
    # opcional: levantar exceção ou retornar um estado final com mensagem
    return state
```

```python
graph = StateGraph[PipelineState]()

n_pre = graph.add_node(Node(preprocess_node, name="Preprocess"))
n_ner = graph.add_node(Node(ner_node, name="NER"))
n_rag = graph.add_node(Node(rag_node, name="RAG"))
n_prompt = graph.add_node(Node(prompt_prep_node, name="PromptPrep"))
n_model = graph.add_node(Node(mlmmodel_node, name="MLLM"))
n_err = graph.add_node(Node(error_node, name="Error"))
```

```python
graph.add_edge(n_pre, n_ner)
graph.add_edge(n_ner, n_rag)
graph.add_edge(n_rag, n_prompt)
graph.add_edge(n_prompt, n_model)
```

```
# condição de erro: se no preprocess estado.error está definido → pula para error_node
graph.add_conditional_edge(
    from_node=n_pre,
    to_node=n_err,
    condition=lambda state: state.get("error") is not None
)
# caso contrário, segue para n_ner
graph.add_edge(n_pre, n_ner)
```

```
graph.set_start(n_pre)
```

```
initial_state: PipelineState = {"image": input_image, "text": input_text}
final_state = graph.run(initial_state)
final_state.model_output # contém a predição ou final_state.error a mensagem de falha
```

Capítulo 2

Pesquisa com Seres Humanos em Computação: Aspectos Éticos, Regulamentação e Cenários Didáticos

Maria da Graça Campos Pimentel e Kamila Rios da Hora Rodrigues

Resumo

A pesquisa em Computação tem incorporado, de forma crescente, métodos empíricos que envolvem pessoas e dados pessoais, tornando a reflexão ética um componente central do processo científico. Este texto discute fundamentos históricos e normativos da ética em pesquisa, apresenta os princípios do Relatório Belmont e sintetiza o arcabouço regulatório brasileiro aplicável à pesquisa com seres humanos, incluindo as resoluções do Conselho Nacional de Saúde e a Lei Geral de Proteção de Dados Pessoais. Como abordagem didática, são analisados cenários hipotéticos que ilustram situações recorrentes de pesquisa em Computação com participação humana direta, evidenciando como princípios éticos e exigências regulatórias orientam decisões metodológicas, a proteção de participantes e a necessidade de apreciação ética institucional.

2.1. Introdução

Em grande parte da pesquisa contemporânea em Computação, pessoas, comportamentos humanos e dados pessoais integram diretamente o objeto de estudo, seja como participantes, seja como fontes de dados ou alvos de intervenções. Ainda assim, é recorrente que pesquisas na área sejam avaliadas a partir de referenciais derivados da pesquisa biomédica, nos quais a sensibilidade ética costuma ser associada principalmente à presença de intervenções clínicas ou de riscos físicos diretos.

O uso desses referenciais, quando aplicado sem considerar as especificidades da pesquisa em Computação, tende a reduzir a visibilidade de implicações éticas relevantes em determinados tipos de estudo, em especial aqueles que envolvem interação com participantes, coleta sistemática de dados pessoais ou inferências sobre comportamento humano. Em áreas como Sistemas Multimídia e Web, Interação Humano-Computador, Análise de Redes Sociais e Aprendizado de Máquina aplicado a dados humanos, tais

implicações emergem não do domínio em si, mas das escolhas metodológicas adotadas, incluindo o desenho do estudo, a forma de coleta e o tratamento dos dados.

Nesse contexto, a ética em pesquisa ocupa um papel estruturante no processo científico, orientando escolhas metodológicas, formas de governança institucional e o atendimento a marcos normativos e legais aplicáveis.

Este texto tem por objetivo apresentar, de forma integrada, (a) um panorama sintético da gênese da ética em pesquisa e de seus marcos normativos históricos; (b) um enquadramento conceitual baseado nos princípios do Relatório Belmont; (c) uma visão operacional das estruturas brasileiras de regulação ética e de proteção de dados aplicáveis à pesquisa em Computação; e (d) um componente didático baseado na análise de cenários hipotéticos, utilizados para exercitar a análise ética de situações recorrentes de pesquisa com *participação humana direta* e coleta de dados junto a participantes. A discussão de cenários baseados em dados pessoais obtidos indiretamente (*p.ex.*, registros digitais e dados de plataformas) constitui um desdobramento natural, não abordado neste texto.

Uma versão anterior deste tutorial apresenta uma discussão mais extensa das definições e dos princípios da legislação brasileira aplicável à pesquisa com seres humanos, bem como o detalhamento do processo de submissão de protocolos de projetos de pesquisa para avaliação por Comitês de Ética em Pesquisa (CEPs) por meio da Plataforma Brasil [Pimentel et al. 2023]. No presente texto, a discussão mantém o foco na pesquisa em Computação e situa os principais marcos legais e normativos aplicáveis à pesquisa com seres humanos no Brasil, incluindo a Lei Geral de Proteção de Dados Pessoais (LGPD) e a Lei nº 14.874/2024, em articulação com as resoluções do Conselho Nacional de Saúde. O eixo conceitual adotado é o Relatório Belmont, utilizado como base para a análise dos cenários hipotéticos apresentados em trabalho anterior das autoras e colaboradores [Rodrigues et al. 2024a].

A abordagem adotada neste tutorial é informada por experiência empírica acumulada das autoras em pesquisas previamente submetidas à avaliação ética institucional, conforme detalhado na Seção 2.6.

A organização do texto é como segue. A Seção 2.2 discute a gênese da ética em pesquisa e seus marcos normativos internacionais. A Seção 2.3 situa a discussão no contexto da pesquisa em Computação e apresenta os princípios do Relatório Belmont. A Seção 2.4 sintetiza as estruturas nacionais de regulação ética no Brasil, incluindo resoluções do Conselho Nacional de Saúde e a LGPD. A Seção 2.5 discute cenários hipotéticos com participação humana direta como exercício pedagógico de análise ética. A Seção 2.6 contextualiza a experiência prévia das autoras no tema. A Seção 2.7 apresenta as considerações finais. A Seção 2.8 descreve os cuidados éticos e de proteção de dados do texto, incluindo a explicitação de que não houve pesquisa empírica com seres humanos nem coleta ou tratamento de dados pessoais, e a Seção 2.9 apresenta os agradecimentos.

2.2. Violação, controvérsia e risco humano: a gênese da ética em pesquisa

A consolidação dos princípios éticos que orientam pesquisas envolvendo seres humanos resulta de um processo histórico marcado por violações, controvérsias morais e debates públicos que expuseram a necessidade de limites à atuação científica. A ética em pesquisa

não surge como abstração normativa, mas como resposta institucional a práticas que causaram sofrimento, danos irreversíveis e violações sistemáticas de direitos humanos.

Durante a Segunda Guerra Mundial, experimentos conduzidos por médicos nazistas em prisioneiros de campos de concentração tornaram-se amplamente conhecidos por seu caráter desumano e frequentemente letal. Essas práticas envolveram exposição deliberada a condições extremas, testes médicos invasivos e ausência completa de consentimento, configurando um dos marcos mais graves da história da ciência e da medicina. O Quadro 2.2.1 sintetiza as principais categorias desses experimentos. A reflexão ética posterior a esses eventos influenciou diretamente a formulação de códigos e diretrizes internacionais, incluindo o Código de Nuremberg e debates sobre moralidade no contexto do regime nazista [Mellanby 1947, Kühne 2018].

Quadro 2.2.1: Categorias de experimentos médicos conduzidos pelo regime nazista (1942–1945) [Museum 2025]

Categoria	**Descrição**
Sobrevivência militar	Experimentos voltados a simular condições extremas enfrentadas por militares, incluindo testes de alta altitude em câmaras de baixa pressão, experimentos de congelamento para estudo da hipotermia e testes para tornar água marinha potável.
Testes de medicamentos e tratamentos	Inoculação deliberada de doenças como malária, tifo, tuberculose e hepatite para testar soros e vacinas; exposição a gases tóxicos como fosgênio e mostarda; experimentos com novos fármacos.
Experimentos raciais e ideológicos	Pesquisas destinadas a sustentar a ideologia racial nazista, incluindo estudos com gêmeos conduzidos por Josef Mengele em Auschwitz, experimentos sorológicos com populações Romani e testes de esterilização em massa de grupos considerados “indesejáveis”.

Já no início do século XX, Marie Curie e integrantes de sua equipe foram expostos de forma prolongada a substâncias radioativas durante suas próprias pesquisas, em um período em que os riscos da radiação eram pouco compreendidos e inexistiam medidas adequadas de proteção. No caso de Marie Curie, essa exposição teve consequências graves para sua saúde, culminando em sua morte por leucemia, associada à exposição prolongada à radiação ionizante [Gaśińska 2015].

Em diferentes momentos históricos, tanto antes como depois da Segunda Guerra Mundial, a ausência de distinção clara entre pesquisador e participante também produziu situações eticamente problemáticas como as sumarizadas no Quadro 2.2.3.

No final do século XVIII, Edward Jenner conduziu experimentos relacionados à vacinação contra a varíola envolvendo James Phipps, um menino de oito anos pertencente a uma família em posição socialmente subordinada ao pesquisador, em um contexto no qual não existiam critérios formais de consentimento informado nem procedimentos sistemáticos de avaliação de riscos [Riedel 2005].

Já no início do século XX, Marie Curie e integrantes de sua equipe foram expostos de forma prolongada a substâncias radioativas durante suas próprias pesquisas, em um período em que os riscos da radiação eram pouco compreendidos e não havia protocolos de proteção adequados. No caso de Marie Curie, essa exposição teve consequências

Quadro 2.2.2: Experimentos com exposição a agentes químicos conduzidos pelos EUA (década de 1940) [Dickerson 2015b, Dickerson 2015a]

Aspecto	**Descrição**
Objetivo declarado	Avaliar efeitos fisiológicos e militares da exposição a gás mostarda e outros agentes químicos em cenários simulados de combate.
População participante	Aproximadamente 60 mil soldados; em diversos casos, agrupamentos raciais foram utilizados para investigar supostas diferenças biológicas de sensibilidade.
Procedimentos adotados	Aplicação direta de substâncias com efeito vesicante sobre a pele; confinamento em câmaras saturadas de gás; exposições em campo aberto simulando combate.
Condições éticas	Ausência de consentimento informado adequado; registros médicos incompletos; imposição de sigilo sob ameaça de sanções disciplinares.
Consequências posteriores	Dificuldades no reconhecimento e na compensação de danos após a desclassificação oficial do programa nos anos 1990.

Quadro 2.2.3: Exemplos históricos de experimentação e tensões éticas pré-normativas

Caso histórico	**Contexto**	**Tensão ética evidenciada**
Edward Jenner (1796)	Inoculação experimental contra varíola	Ausência de consentimento formal e risco imposto
Marie Curie (início do séc. XX)	Pesquisa com radioatividade	Exposição a risco desconhecido
Stanley Milgram (1960s)	Experimentos de obediência	Engano, sofrimento psicológico e autoridade

graves para sua saúde, culminando em sua morte por leucemia, associada à exposição prolongada à radiação ionizante [Gaśińska 2015].

Na segunda metade do século XX, estudos em Psicologia passaram a levantar novas questões éticas. Os experimentos de obediência conduzidos por Stanley Milgram, entre 1961 e 1963, investigaram como indivíduos comuns reagiam a ordens de autoridade que envolviam causar sofrimento a terceiros. Para isso, os participantes foram deliberadamente enganados, pressionados a continuar e expostos a intenso estresse emocional, acreditando estar causando danos reais a outras pessoas [Milgram 1963]. Esses estudos integraram um conjunto mais amplo de controvérsias éticas na pesquisa com seres humanos que, nas décadas seguintes, motivaram a consolidação de princípios voltados à proteção de participantes de pesquisa, formalizados no Relatório Belmont [Belmont 1979].

Mais recentemente, propostas metodológicas alternativas têm buscado investigar fenômenos semelhantes aos estudados nos experimentos de Milgram, em particular a obediência à autoridade, sem recorrer ao engano sistemático ou à indução de sofrimento psicológico intenso. Essas abordagens demonstram que é possível estudar processos complexos de tomada de decisão e (des)obediência preservando a autonomia e o bem-estar das pessoas participantes, evidenciando que avanços científicos podem ser compatíveis com os princípios éticos que orientam pesquisas com participação humana [Caspar 2021].

Como resposta institucional aos episódios que levantaram questões éticas, diferentes documentos normativos foram elaborados ao longo do século XX com o objetivo

de estabelecer limites éticos à pesquisa científica. O Código de Nuremberg, formulado em 1947 e incorporado à sentença do julgamento de médicos nazistas conduzido entre 1946 e 1947, consagrou o consentimento voluntário como requisito essencial para pesquisas com seres humanos. Com isso, a ética da pesquisa deixou de se apoiar apenas na beneficência do pesquisador e passou a reconhecer os direitos do participante como limite normativo da atuação científica [Shuster 1997].

A ruptura promovida por esse código deve ser compreendida não apenas como reação a práticas médicas abusivas, mas como reafirmação de uma tradição ética que reconhece proteção moral independentemente de pertencimento racial ou comunitário. Como analisa Kühne, o regime nazista redefiniu a própria noção de dever moral, limitando-a à comunidade racial e qualificando compaixão por "outros" como fraqueza ou traição [Kühne 2018].

Posteriormente, a Declaração de Helsinque, publicada originalmente em 1964 e periodicamente revisada – atualmente em sua décima edição, de 2024 – ampliou esse marco ao oferecer diretrizes para a ética da pesquisa médica em âmbito internacional. O documento distinguiu entre pesquisa associada ao cuidado clínico e pesquisa não terapêutica, reforçou a centralidade da autonomia e da dignidade das pessoas participantes e estabeleceu parâmetros para avaliação independente, proporcionalidade entre riscos e benefícios e responsabilidade profissional [WMA 2024].

A repercussão pública das violações éticas associadas tanto aos experimentos conduzidos por médicos nazistas quanto ao Estudo da Sífilis de Tuskegee (1932–1972) – que envolveu a privação deliberada de tratamento eficaz a participantes em condição de vulnerabilidade – contribuiu decisivamente para a aprovação, nos Estados Unidos, do National Research Act de 1974. Essa lei instituiu a comissão responsável pela elaboração do Relatório Belmont, que sistematizou princípios destinados a orientar a pesquisa com seres humanos [Belmont 1979].

Esses instrumentos, sumarizados no Quadro 2.2.4, contribuíram para consolidar um arcabouço conceitual e normativo que viria a ser sistematizado, nos Estados Unidos, pelo Relatório Belmont [Belmont 1979]. Observa-se, ao longo desses marcos, a mudança de respostas pontuais a violações para a consolidação de princípios normativos que orientam a pesquisa em diferentes áreas do conhecimento

Quadro 2.2.4: Evolução conceitual da ética em pesquisa no século XX

Marco	**Problema histórico**	**Resposta normativa central**	**Avanço conceitual**
Código de Nuremberg (1947)	Experimentos coercitivos e ausência de consentimento	Consentimento voluntário como requisito absoluto	Autonomia como limite ético inegociável
Declaração de Helsinque (1964–2024)	Expansão global da pesquisa médica	Revisão independente e avaliação risco–benefício	Proporcionalidade e responsabilidade profissional
Relatório Belmont (1979)	Necessidade de sistematização normativa	Três princípios operacionais (respeito, beneficência, justiça)	Estrutura ética aplicável a diferentes áreas científicas

2.3. Ética e pesquisa em Computação

No contexto científico, é fundamental distinguir entre moral e ética. A moral refere-se a valores, crenças e julgamentos compartilhados por indivíduos ou grupos sociais, frequentemente associados a intuições sobre "fazer o certo". A ética em pesquisa, por sua vez, diz respeito à reflexão normativa e à formalização institucional dessas práticas no espaço público, por meio de princípios, regras e procedimentos que orientam e regulam a condução da pesquisa científica [Mandal et al. 2011].

O Relatório Belmont reconhece a existência de obrigações morais inerentes à pesquisa científica, mas não propõe uma definição filosófica de moralidade. Em vez disso, parte dessas obrigações para formular princípios éticos operacionais, voltados à orientação institucional e à tomada de decisão responsável em contextos concretos de pesquisa [Belmont 1979].

A distinção entre moral e ética é central porque decisões moralmente bem-intencionadas não substituem procedimentos éticos formais. A ética em pesquisa se materializa em protocolos, documentos, avaliações institucionais e mecanismos de responsabilização [CNS 2012, CNS 2016]. Em particular, ela estabelece critérios para avaliar riscos, benefícios, consentimento, privacidade e proteção de participantes.

Na Computação, essa discussão ganhou relevância à medida que dados digitais, registros audiovisuais, interações mediadas por sistemas e análises automatizadas passaram a representar, direta ou indiretamente, pessoas reais. Mesmo quando não há interação direta com participantes, decisões técnicas podem gerar impactos sociais, psicológicos, reputacionais ou legais, o que torna a reflexão ética indispensável [LGPD 2018].

2.3.1. Princípios de Belmont

O Relatório Belmont sistematizou três princípios éticos fundamentais que orientam pesquisas envolvendo participantes humanos: **respeito às pessoas**, **beneficência** e **justiça** [Belmont 1979]. Esses três princípios constituem referência central para a estruturação normativa da ética em pesquisa em diversos contextos regulatórios, e fornecem um arcabouço conceitual para a análise ética de diferentes contextos de investigação científica [Mandal et al. 2011].

É relevante observar que, na literatura bioética contemporânea: (a) beneficência e não maleficência são distinguidas como deveres correlatos de promover benefícios e evitar danos, e (b) o "Respeito às Pessoas" do Relatório Belmont é frequentemente formulado como princípio da autonomia, com fundamento na tradição kantiana [Varkey 2021].

2.3.1.1. Respeito às pessoas

O princípio do respeito às pessoas estabelece que indivíduos devem ser tratados com dignidade e reconhecidos como agentes autônomos, capazes de tomar decisões sobre sua própria participação em pesquisas. Pessoas com autonomia diminuída – seja por razões etárias, cognitivas, sociais ou institucionais – têm direito à proteção especial, de modo a evitar exploração ou coerção.

A principal aplicação do princípio do respeito ocorre por meio do consentimento informado. As pessoas participantes devem, na medida de sua capacidade, ter a oportunidade de decidir livremente sobre o que deve ou não acontecer com elas no contexto da pesquisa. O processo de consentimento envolve três elementos fundamentais: a disponibilização de informações adequadas sobre a pesquisa, a compreensão dessas informações por parte da pessoa participante e a voluntariedade da decisão, livre de pressões ou influências indevidas. No contexto brasileiro, esse processo é formalizado por meio do Termo de Consentimento Livre e Esclarecido (TCLE).

Em situações em que as pessoas participantes não possuem plena capacidade de decisão – como no caso de crianças, adolescentes ou pessoas legalmente incapazes – o respeito às pessoas exige mecanismos adicionais de proteção. Nesses casos, o consentimento deve ser obtido junto ao responsável legal por meio do TCLE, e, sempre que houver compreensão possível, deve ser complementado pelo Termo de Assentimento Livre e Esclarecido (TALE) obtido junto à pessoa participante.

O TALE expressa a anuência da própria pessoa participante, apresentada em linguagem adequada à sua idade, condição e contexto, e tem como objetivo assegurar que a participação não ocorra de forma coercitiva ou meramente autorizada por terceiros. O uso do TALE não substitui o TCLE do responsável legal, mas o complementa, reforçando o reconhecimento da pessoa participante como sujeito moral, mesmo em contextos de autonomia reduzida.

Exemplos de adaptação do TALE a formatos adequados ao público infantil incluem o uso de recursos multimodais, como histórias em quadrinhos, que permitem apresentar objetivos, procedimentos e direitos de forma compatível com o desenvolvimento cognitivo e linguístico das crianças [Albres and Sousa 2019].

2.3.1.2. Beneficência

O princípio da beneficência estabelece que a pesquisa deve buscar maximizar os benefícios possíveis e, simultaneamente, evitar causar danos às pessoas participantes. Esse princípio envolve tanto a promoção de benefícios quanto a identificação, redução ou eliminação de riscos potenciais.

A aplicação da beneficência exige a avaliação sistemática dos riscos e benefícios envolvidos na pesquisa. A natureza, a magnitude e a probabilidade de ocorrência de riscos devem ser analisadas de forma criteriosa, assim como os benefícios esperados, sejam eles diretos ou indiretos. Essa avaliação fundamenta decisões sobre a viabilidade ética da pesquisa e sobre a adoção de medidas de proteção, mitigação ou monitoramento ao longo de sua execução.

Do ponto de vista procedimental, os riscos identificados – ainda que considerados mínimos – devem ser explicitados de forma clara, proporcional e compreensível nos instrumentos de consentimento, em especial no Termo de Consentimento Livre e Esclarecido (TCLE) e, quando aplicável, no Termo de Assentimento Livre e Esclarecido (TALE). A omissão ou minimização indevida de riscos compromete tanto o princípio da beneficência quanto o respeito à autonomia das pessoas participantes, uma vez que inviabiliza uma decisão verdadeiramente informada.

2.3.1.3. Justiça

O princípio da justiça diz respeito à distribuição equitativa dos riscos e benefícios da pesquisa. Ele exige que nenhum grupo social seja sistematicamente sobrecarregado com os riscos da pesquisa, nem excluído injustificadamente dos potenciais benefícios gerados pelo conhecimento científico.

A aplicação desse princípio está diretamente relacionada aos critérios de seleção de participantes. Procedimentos de recrutamento devem ser justos e eticamente justificáveis, evitando tanto a exclusão arbitrária quanto a seleção conveniente de grupos em situação de maior vulnerabilidade. A justiça envolve, portanto, a análise das razões pelas quais determinados grupos são incluídos ou excluídos da pesquisa e das implicações dessas escolhas para a equidade na produção e na aplicação do conhecimento.

Do ponto de vista procedimental, o princípio da justiça se reflete sobretudo nas decisões de desenho do estudo e nos critérios de inclusão e exclusão de participantes, que devem ser eticamente justificáveis e transparentes. Esses critérios devem também ser comunicados de forma clara nos instrumentos de consentimento – TCLE e, quando aplicável, TALE – assegurando que a participação não implique ônus indevidos, nem restrinja direitos, benefícios institucionais ou oportunidades das pessoas envolvidas. A clareza quanto à voluntariedade da participação e à ausência de prejuízos em caso de recusa ou desistência constitui um elemento essencial para a realização prática da justiça.

2.4. Estruturas nacionais de regulação ética no Brasil

A presente seção apresenta uma visão sintética das estruturas nacionais de regulação ética. Uma discussão mais detalhada sobre a legislação brasileira aplicável à pesquisa com seres humanos (não atualizada em relação à Lei nº 14.874 [Lei14874 2024]) e sobre a submissão de projetos de pesquisa para avaliação por Comitês de Ética em Pesquisa (CEPs) por meio da Plataforma Brasil encontra-se em [Pimentel et al. 2023].

No Brasil, a regulação ética das pesquisas envolvendo seres humanos é estruturada por um conjunto de instâncias institucionais e normativas. Historicamente, esse sistema era coordenado pelo Conselho Nacional de Saúde (CNS), pela Comissão Nacional de Ética em Pesquisa (CONEP) e por uma rede de Comitês de Ética em Pesquisa (CEP), distribuídos nas instituições de ensino e pesquisa. Esse arcabouço era sustentado por resoluções do CNS, por orientações normativas complementares da CONEP e por legislação federal correlata, incluindo normas específicas voltadas à proteção de dados pessoais.

A Lei nº 14.874 institui o Sistema Nacional de Ética em Pesquisa com Seres Humanos (SINEP), estruturado em uma Instância Nacional de Ética em Pesquisa (INEP) e em instâncias de análise ética em pesquisa, representadas pelos CEPs [Lei14874 2024]. Conforme regulamentado pelo Decreto nº 12.651, os CEPs passam a ser definidos de acordo com a complexidade da análise ética e o grau de risco envolvido na pesquisa, distinguindo-se entre CEPs credenciados, responsáveis pela análise de pesquisas de risco baixo e moderado, e CEPs acreditados, habilitados também à análise de pesquisas de risco elevado [Decreto12651 2025]. Para efeitos da análise ética de projetos de pesquisa, os CEPs permanecem como as instâncias responsáveis pela apreciação inicial e pelo acompanhamento dos estudos, e as resoluções normativas do CNS continuam vigentes.

No caso das pesquisas em Ciências Humanas e Sociais (e áreas da Computação), a regulamentação vigente admite procedimentos diferenciados conforme o grau de risco ético envolvido. Um decreto recente estabelece que pesquisas classificadas como de risco baixo podem ser submetidas a procedimento de notificação ou de análise ética simplificada, conforme norma a ser editada pela Instância Nacional de Ética em Pesquisa, sem prejuízo do cumprimento das diretrizes éticas aplicáveis. Em qualquer hipótese, o processo de pesquisa permanece sujeito à avaliação ética completa pelos Comitês de Ética em Pesquisa, podendo resultar em suspensão, cancelamento do estudo ou outras medidas previstas na legislação [Decreto12651 2025].

Esses procedimentos diferenciados no âmbito do sistema de ética em pesquisa não afastam a incidência da legislação de proteção de dados pessoais. O tratamento de dados pessoais em pesquisas, inclusive nas ciências humanas e sociais, deve ser avaliado de forma autônoma à luz da Lei Geral de Proteção de Dados – LGPD (ver 2.4.5), observando-se as bases legais, os princípios e os direitos dos titulares, independentemente da classificação de risco ético atribuída no âmbito do SINEP.

Nesse contexto institucional e normativo, o sistema estabelece princípios, responsabilidades e procedimentos que tornam obrigatória a avaliação ética prévia de pesquisas com seres humanos, independentemente da área do conhecimento, sempre que haja envolvimento direto ou indireto de pessoas ou de dados a elas associados.

2.4.1. Resolução CNS nº 196/1996: marco inicial da governança ética

A Resolução CNS nº 196/1996 constitui o marco fundador da governança ética nacional em pesquisas envolvendo seres humanos. Alinhada a documentos internacionais como o Código de Nuremberg e a Declaração de Helsinque, essa resolução estabeleceu diretrizes éticas gerais aplicáveis à pesquisa científica no país [CNS 1996].

Entre seus principais desdobramentos, a resolução instituiu a obrigatoriedade da criação de Comitês de Ética em Pesquisa nas instituições e criou a Comissão Nacional de Ética em Pesquisa (CONEP) como instância coordenadora do sistema. A partir desse marco, a avaliação ética tornou-se requisito obrigatório antes do início da coleta de dados em pesquisas com seres humanos. Posteriormente, a Resolução 196/1996 foi estendida e substituída pela Resolução CNS nº 466/2012.

2.4.2. Resolução CNS nº 466/2012: diretrizes éticas para pesquisa com pessoas

A Resolução CNS nº 466/2012 revoga e amplia as disposições da Resolução 196/1996, consolidando os princípios éticos fundamentais que regem pesquisas envolvendo seres humanos no Brasil. Entre esses princípios estão o respeito à dignidade e à autonomia das pessoas participantes, a beneficência, a não maleficência e a justiça [CNS 2012].

A norma define o sujeito da pesquisa como participante individual ou coletivo, cuja participação deve ocorrer de forma voluntária e informada. É vedada qualquer forma de remuneração pela participação, sendo admitido apenas o ressarcimento de despesas ou prejuízos decorrentes do envolvimento na pesquisa. A resolução estabelece a exigência de avaliação ética prévia pelo Sistema CEP/CONEP (atualmente INEP/CEP) e institucionaliza a Plataforma Brasil como meio oficial para submissão, acompanhamento e tramitação dos protocolos de pesquisa.

A Resolução 466/2012 enfatiza a proteção do sigilo e da privacidade das pessoas participantes e de suas informações pessoais, bem como a necessidade de atenção especial a grupos considerados vulneráveis. A responsabilidade ética é compartilhada entre pesquisador e instituição, cabendo ao CEP o papel de instância avaliadora e orientadora. Os resultados das pesquisas devem preservar a privacidade e a anonimização, sendo relatados de forma precisa e transparente.

No que se refere ao Termo de Consentimento Livre e Esclarecido (TCLE), a resolução estabelece requisitos como clareza, linguagem acessível e voluntariedade, assegurando o direito de recusa e de desistência a qualquer momento, inclusive após o início da participação. O TCLE deve informar riscos e benefícios, procedimentos, cronograma, garantias de confidencialidade, bem como os contatos da equipe de pesquisa e do CEP. O consentimento, por sua vez, deve ser compreendido como um processo contínuo, que pode ser reafirmado, revisto ou interrompido ao longo da pesquisa, e não apenas como a assinatura inicial de um documento.

2.4.2.1. Grupos vulneráveis e termo de assentimento

A Resolução CNS nº 466/2012 estabelece proteção especial a grupos considerados vulneráveis, como crianças, adolescentes e pessoas legalmente incapazes. Nesses casos, além do TCLE obtido junto ao responsável legal, recomenda-se o uso do Termo de Assentimento Livre e Esclarecido (TALE), como instrumento de anuência livre e sem coerção, apresentado em linguagem adequada à idade e à condição da pessoa participante.

A inclusão de crianças, adolescentes, pessoas com transtornos mentais ou com redução significativa da capacidade decisória deve ser devidamente justificada. A norma também ressalta a necessidade de proteger a autonomia em contextos de dependência ou subordinação – como estudantes, militares, empregados, pessoas privadas de liberdade ou institucionalizadas – de modo a garantir liberdade real de consentimento, sem pressão ou risco de represália.

2.4.3. Resolução CNS nº 510/2016: enquadramento de pesquisas não biomédicas

A Resolução CNS nº 510/2016 estabelece normas éticas específicas para pesquisas envolvendo seres humanos que não se enquadram no modelo biomédico experimental, contemplando abordagens qualitativas, observacionais e interpretativas, com métodos e instrumentos característicos das Ciências Humanas e Sociais e de outras áreas do conhecimento que adotam desenhos empíricos não clínicos [CNS 2016].

A resolução introduz o princípio da proporcionalidade na avaliação ética, reconhecendo que pesquisas diferem quanto ao grau de intervenção, exposição e risco. Nesse contexto, a Resolução 510/2016 prevê hipóteses específicas de dispensa de apreciação ética, desde que não haja identificação direta ou indireta das pessoas participantes, nem riscos à dignidade, imagem, segurança, direitos ou privacidade, e que não envolva grupos ou temas sensíveis. Em todos os demais casos, a submissão ao Comitê de Ética em Pesquisa permanece obrigatória (ver Seção 2.4.3.1).

A Resolução CNS nº 510/2016 tem relevância particular para pesquisas em Computação, em áreas como Web, Análise de Dados e Interação Humano-Computador, uma

vez que reconhece explicitamente abordagens metodológicas não experimentais, estudos observacionais, análises de conteúdos digitais, pesquisas com registros públicos e investigações baseadas em interações mediadas por tecnologia. Ao deslocar o foco exclusivo de intervenções biomédicas, a resolução amplia o enquadramento ético para práticas comuns nessas áreas, como análise de plataformas digitais, estudos com usuários em contextos naturais de uso de sistemas computacionais, e coleta de dados online.

Embora publicada antes da Lei Geral de Proteção de Dados Pessoais (LGPD – Lei nº 13.709/2018), a Resolução CNS nº 510/2016 passou, com a vigência da LGPD, a demandar dupla conformidade normativa. Isso significa que a eventual dispensa de apreciação ética prevista na Resolução 510/2016 não afasta, por si só, as obrigações legais relativas ao tratamento de dados pessoais, incluindo a definição de bases legais, a observância dos princípios da finalidade e da necessidade e o respeito aos direitos das pessoas titulares dos dados (ver Seção 2.4.5).

Nessa perspectiva, as orientações posteriores da Comissão Nacional de Ética em Pesquisa devem ser interpretadas à luz do arcabouço legal vigente, especialmente no que se refere à articulação entre avaliação ética e proteção de dados pessoais.

2.4.3.1. Ofício Circular nº 17/2022: orientações sobre dispensas

O Ofício Circular nº 17/2022 da CONEP não cria novas hipóteses normativas, mas operacionaliza e esclarece dispositivos já previstos na Resolução CNS nº 510/2016, com o objetivo de orientar interpretações e reduzir decisões inconsistentes na aplicação das dispensas de apreciação ética [CONEP 2022].

O documento detalha as hipóteses de dispensa previstas no artigo 1º da Resolução 510/2016, incluindo pesquisas de opinião pública sem identificação desde o recrutamento e sem intervenção, uso de informações de acesso público ou de domínio público, pesquisas censitárias oficiais, uso de bancos de dados agregados sem identificação individual, revisões de literatura, aprofundamentos teóricos sem dados identificáveis e atividades exclusivamente de ensino ou treinamento, desde que não haja intenção de pesquisa.

O Ofício Circular reforça, contudo, que a dispensa de apreciação ética não deve ser inferida de forma automática a partir do tipo de dado, do instrumento de coleta ou do ambiente em que a informação se encontra disponível. Em particular, o fato de um dado ser publicamente acessível não elimina, por si só, a possibilidade de identificação direta ou indireta, nem afasta riscos à privacidade, à dignidade, à imagem ou a outros direitos das pessoas envolvidas.

Da mesma forma, pesquisas de opinião podem demandar apreciação ética mesmo na ausência de identificação explícita, sempre que houver finalidade científica e risco além do cotidiano, seja em função da natureza das perguntas, da população envolvida, do contexto de aplicação ou do potencial de repercussões pessoais, sociais, reputacionais ou institucionais.

O Ofício Circular explicita ainda que, quando houver interesse posterior de pesquisa a partir de atividades inicialmente não caracterizadas como tal – como ações de extensão, atividades didáticas ou treinamentos – a submissão do projeto ao Comitê de Ética em Pesquisa torna-se obrigatória antes da utilização dos dados em pesquisas ou pu-

blicações científicas. Essa orientação reafirma o princípio de que a avaliação ética deve anteceder a conversão de registros, respostas ou observações em evidência científica, inclusive em estudos autodenominados análises ou avaliações retrospectivas.

2.4.4. Resolução CNS nº 674/2022: tipificação, tramitação e prazos

A Resolução CNS nº 674/2022 estabelece a tipificação das pesquisas envolvendo seres humanos e define os fluxos de tramitação dos protocolos no Sistema Nacional de Ética em Pesquisa (SINEP). Seu anexo organiza os projetos em categorias gerais – tipos A, B e C – com subdivisões internas, associando cada tipificação a prazos e procedimentos específicos de análise [CNS 2022].

A tipificação tem como objetivo ajustar o nível de rigor da avaliação ética à natureza da pesquisa, ao grau de intervenção sobre as pessoas participantes e ao nível de risco envolvido. De modo geral, quanto maior a complexidade ou o potencial de risco da pesquisa, mais extensa e criteriosa tende a ser a tramitação no sistema.

Considerando o arranjo institucional vigente a partir da Lei 14874/2024, a tipificação das pesquisas articula-se também à competência dos Comitês de Ética em Pesquisa conforme o grau de risco envolvido: pesquisas classificadas como de risco baixo ou moderado são apreciadas por CEPs credenciados, enquanto pesquisas de risco elevado são analisadas por CEPs acreditados, nos termos da regulamentação do SINEP [Decreto12651 2025].

2.4.4.1. Pesquisas do tipo A

As pesquisas do tipo A caracterizam-se, em geral, pela ausência de intervenções diretas sobre as pessoas participantes ou pela utilização de dados previamente existentes. Esse tipo inclui estudos observacionais, análises de dados históricos devidamente anonimizados, pesquisas de opinião não invasivas e investigações baseadas em material previamente coletado, desde que respeitados os critérios éticos aplicáveis.

O tipo A é subdivido (A1, A2, A3 e A4) para diferenciar pesquisas quanto à origem dos dados, à possibilidade de identificação direta ou indireta e à sensibilidade das informações envolvidas. Pesquisas do tipo A exigem apreciação ética prévia sempre que envolverem seres humanos ou dados a eles associados, cabendo ao CEP avaliar o enquadramento correto e as medidas de proteção adotadas.

2.4.4.2. Pesquisas do tipo B

As pesquisas do tipo B envolvem algum grau de intervenção sobre as pessoas participantes, ainda que não clínica. Incluem, por exemplo, estudos que utilizam dispositivos de monitoramento externo, pesquisas com intervenções ambientais leves, experimentos não clínicos e coletas que possam introduzir alterações temporárias no comportamento, no ambiente ou na rotina dos participantes.

O tipo B é subdividido (B1 e B2) de acordo com o nível de intervenção e o risco associado. A avaliação ética dessas pesquisas exige atenção especial aos princípios da beneficência e do respeito às pessoas, particularmente no que se refere à identificação e

mitigação de riscos, à voluntariedade da participação e à clareza dos procedimentos de consentimento.

2.4.4.3. Pesquisas do tipo C

As pesquisas classificadas como tipo C correspondem àquelas de maior complexidade e maior potencial de risco ético, geralmente associadas a intervenções diretas sobre as pessoas participantes, ao uso de material biológico humano ou a procedimentos que demandam avaliação ética em instâncias especializadas. Na legislação vigente, esse tipo de pesquisa tramita em CEPs acreditados, habilitados à análise ética de pesquisas de risco elevado.

2.4.4.4. Tipificação de pesquisas em Computação

Pesquisas nas áreas de Web, Multimídia, Interação Humano-Computador e Análise de Redes Sociais enquadram-se, predominantemente, nos tipos A ou B, uma vez que, em geral, envolvem estudos observacionais, coleta de dados digitais, experimentos não clínicos ou intervenções leves mediadas por tecnologia. A tipificação correta, contudo, depende das características específicas de cada projeto, como o tipo de dado coletado, a possibilidade de identificação dos participantes, o uso de dispositivos de monitoramento e o grau de intervenção introduzido.

O Quadro 2.4.1, adaptado de trabalho anterior, ilustra exemplos de estudos nessas áreas e seus enquadramentos típicos nos tipos A e B, evidenciando a diversidade de situações encontradas na pesquisa em Computação e reforçando que a tipificação não substitui a análise ética situada e caso a caso [Pimentel et al. 2023].

Além da tipificação, a Resolução CNS nº 674/2022 define prazos máximos de tramitação para cada tipo de pesquisa. Pesquisas enquadradas como tipo A seguem fluxos mais céleres, enquanto pesquisas dos tipos B e C estão sujeitas a prazos mais extensos e, quando aplicável, à apreciação em instâncias adicionais. Essa padronização busca reduzir a imprevisibilidade do processo, alinhar expectativas de pesquisadores e comitês e conferir maior transparência à governança ética nacional.

2.4.5. Lei Geral de Proteção de Dados Pessoais (LGPD) – (Lei nº 13.709/2018)

A Lei Geral de Proteção de Dados Pessoais estabelece o marco legal para o tratamento de dados pessoais no Brasil, incluindo dados coletados, armazenados e analisados em formato digital. No contexto da pesquisa científica, a LGPD introduz, além de uma exigência legal, um reforço ético que se articula diretamente com a avaliação ética no âmbito do Sistema Nacional de Ética em Pesquisa com Seres Humanos, ampliando a responsabilidade de pesquisadores e instituições quanto ao uso dessas informações [LGPD 2018].

Na LGPD, o caráter pessoal de um dado é definido por sua relação com uma pessoa natural identificada ou identificável, *independentemente de o dado estar disponível publicamente*. A publicidade do dado não elimina sua natureza pessoal, nem afasta a aplicação dos princípios legais e éticos que regulam seu tratamento.

Quadro 2.4.1: Exemplos de estudos e sua tipificação, adaptados de [Pimentel et al. 2023]

Tipo A	
Estudos com Dados Pré-existentes	Análise de padrões de acesso a páginas de um portal institucional utilizando logs históricos agregados e anonimizados, sem possibilidade de identificação direta ou indireta de pessoas naturais.
Pesquisas Observacionais	Estudo sobre como diferentes designs de interface de usuário afetam a navegação dos usuários em um site de notícias.
Coleta Dirigida de Dados Não-Invasiva	Pesquisa de opinião online para avaliar a usabilidade de um novo aplicativo móvel.
Estudos com Material Biológico Armazenado	Uso de dados genéticos armazenados para desenvolver algoritmos de bioinformática.
Tipo B	
Coleta de Amostras Biológicas	Estudo que coleta saliva para análise de hormônios relacionados ao estresse em um jogo de realidade virtual.
Uso de Dispositivos de Monitoramento Externo	Uso de pulseiras com sensores de acelerômetro para monitorar atividade física em uma aplicação de *fitness*.
Estudos Intervencionais Não-Clínicos	Teste dos efeitos de diferentes tipos de música em plataformas de streaming sobre a concentração de estudantes.
Pesquisas Observacionais com Intervenções Leves	Alteração da configuração de iluminação em um ambiente de trabalho para avaliar efeitos sobre a produtividade.

A lei distingue dados pessoais comuns e dados pessoais sensíveis – como informações sobre saúde, origem racial ou étnica, convicções religiosas, opiniões políticas, dados genéticos e biométricos –, exigindo justificativa explícita, minimização do tratamento e adoção de medidas técnicas e organizacionais adequadas de proteção. No âmbito da pesquisa, o princípio da necessidade impõe que apenas os dados estritamente relevantes aos objetivos científicos sejam coletados, vedando práticas de coleta excessiva ou indiscriminada, mesmo quando tecnicamente viáveis.

Nos projetos submetidos à apreciação ética, a aderência à LGPD é avaliada de forma integrada aos princípios éticos fundamentais. Os protocolos devem descrever, de maneira clara e verificável, a finalidade do tratamento de dados, as bases legais invocadas, os procedimentos de coleta, armazenamento, retenção, descarte e eventual compartilhamento das informações, bem como os responsáveis pelo tratamento e as medidas de segurança adotadas. A ausência dessas informações configura falha ética relevante.

Nesse contexto, a LGPD reforça a centralidade de técnicas de anonimização e pseudonimização, especialmente em pesquisas que envolvem grandes volumes de dados digitais, registros de plataformas online, logs de sistemas, dados de redes sociais ou informações provenientes de dispositivos de monitoramento. A mera disponibilidade pública desses dados não elimina seu caráter pessoal, tampouco afasta os riscos de reidentificação quando combinados com outras fontes, o que exige avaliação criteriosa e estratégias efetivas de mitigação.

Os instrumentos de consentimento – como o Termo de Consentimento Livre e Esclarecido (TCLE) e, quando aplicável, o Termo de Assentimento Livre e Esclarecido (TALE) – devem incorporar informações compatíveis com a LGPD, incluindo riscos as-

sociados ao tratamento de dados, direitos das pessoas participantes, possibilidades de acesso, correção, limitação ou eliminação dos dados, bem como a possibilidade de revogação do consentimento, nos limites previstos em lei. O consentimento, contudo, não exime o pesquisador do dever de proteção dos dados nem legitima práticas incompatíveis com os princípios éticos ou legais.

Em conjunto, a LGPD e as resoluções éticas nacionais consolidam a proteção da privacidade e dos dados pessoais como dimensão constitutiva dos princípios do respeito às pessoas e da não maleficência. No contexto contemporâneo de pesquisas mediadas por tecnologias digitais e sistemas computacionais, o tratamento ético de dados não pode ser reduzido a um requisito formal, mas deve ser compreendido como parte indissociável da responsabilidade científica, especialmente em um cenário de crescente instrumentalização de dados pessoais para fins acadêmicos, comerciais e institucionais.

2.4.6. Lei nº 14.874, de 28 de maio de 2024

A Lei nº 14.874/2024 consolida, em nível legal, o arranjo institucional da regulação ética da pesquisa com seres humanos no Brasil, formalizando o Sistema Nacional de Ética em Pesquisa com Seres Humanos (SINEP) e conferindo base legal às estruturas e procedimentos anteriormente estabelecidos por resoluções do Conselho Nacional de Saúde.

A lei preserva a obrigatoriedade da avaliação ética prévia de pesquisas envolvendo seres humanos em todas as áreas do conhecimento, ao mesmo tempo em que explicita diretrizes voltadas à racionalização dos fluxos de análise, à proporcionalidade conforme o grau de risco e à articulação com a legislação de proteção de dados pessoais. Nesse sentido, a Lei nº 14.874 pode ser compreendida como consolidação normativa do arranjo previamente estruturado por resoluções do sistema nacional de ética em pesquisa.

2.5. Discussão de cenários hipotéticos

A avaliação ética de pesquisas envolvendo seres humanos exige a análise de situações concretas nas quais princípios éticos abstratos precisam ser traduzidos em decisões metodológicas e institucionais. Esse processo envolve a identificação de riscos potenciais, a consideração de diferentes alternativas de condução da pesquisa e a avaliação da necessidade de apreciação ética institucional.

Cenários hipotéticos constituem instrumentos pedagógicos adequados para esse fim, pois permitem explicitar como os princípios do Relatório Belmont – respeito às pessoas, beneficência e justiça – orientam a reflexão ética em contextos variados. Em cada cenário, esses princípios podem assumir pesos e relevâncias distintas, conforme a natureza da população envolvida, o tipo de dado coletado, o contexto de realização da pesquisa e os riscos associados.

Assim, a análise dos cenários não pressupõe aplicação mecânica ou simétrica dos três princípios, mas busca evidenciar como eles são mobilizados de maneira situada para fundamentar decisões éticas e justificar a necessidade de apreciação prévia por Comitês de Ética em Pesquisa.

Nesse contexto, destaca-se o princípio da justiça, frequentemente reduzido à ideia de voluntariedade, mas dotado de alcance normativo mais amplo. Ele exige que a seleção da população participante seja cientificamente justificada e que a distribuição dos ônus e benefícios da pesquisa seja equitativa, de modo que a escolha da população seja explicitamente fundamentada no desenho metodológico.

A literatura interpretativa do Relatório Belmont permite compreender o princípio da justiça em duas dimensões complementares: uma dimensão interna ao estudo, relacionada à voluntariedade, à ausência de coerção e à proteção dos participantes; e uma dimensão distributiva externa, relacionada à justificativa da escolha da população e à distribuição social de riscos e benefícios. Ambas são contempladas na análise dos cenários, ainda que não se manifestem com igual intensidade em todas as situações.

Os cenários apresentados nesta seção são hipotéticos, extraídos de trabalho anterior [Rodrigues et al. 2024a], inspirados em situações recorrentes na pesquisa em Computação, e têm finalidade exclusivamente didática. Em todos os casos, trata-se de situações que configuram pesquisa envolvendo seres humanos ou dados a eles associados, o que torna obrigatória a submissão prévia a CEP, ainda que algumas atividades possam, à primeira vista, parecer dispensar apreciação ética.

2.5.1. Cenário hipotético 1: Questionário com estudantes universitários

> ***Cenário hipotético 1: Questionário com estudantes universitários***
>
> *Uma equipe de pesquisa desenvolveu um questionário online para consultar estudantes universitários sobre dificuldades de gerenciamento de tempo. O convite para participação foi enviado por e-mail a estudantes de cursos de Computação e continha o título do questionário e um link para resposta. O questionário não coletava nomes ou endereços de e-mail e incluía perguntas sobre experiências, conhecimentos e rotinas dos participantes, incluindo dificuldades de controle do tempo e suas consequências.*

A configuração apresentada caracteriza pesquisa envolvendo seres humanos, pois há coleta sistemática de informações fornecidas por participantes para fins de produção de conhecimento. Ainda que o instrumento não registre identificadores diretos, como nome ou endereço de e-mail, as respostas podem permitir reidentificação indireta em contextos de turmas pequenas ou subgrupos específicos, especialmente quando combinadas variáveis relativas a rotina, experiências acadêmicas e dificuldades relatadas.

Os riscos envolvidos tendem a ser classificados como mínimos, mas não inexistentes. Entre eles destacam-se possível desconforto psicológico decorrente da autorreflexão sobre desempenho acadêmico, exposição de dificuldades pessoais e risco reputacional em caso de tratamento inadequado ou divulgação indevida dos dados.

Conforme os princípios do Relatório Belmont, o respeito às pessoas exige que a participação seja precedida de consentimento livre e esclarecido, com informação adequada sobre objetivos, procedimentos, riscos, voluntariedade e possibilidade de desistência. O cenário não explicita a apresentação de Termo de Consentimento Livre e Esclarecido (TCLE), o que constitui lacuna procedimental relevante. A beneficência demanda

análise prévia dos riscos e definição de medidas concretas de proteção, incluindo estratégias adequadas de anonimização e armazenamento seguro. A justiça impõe cuidado no recrutamento, de modo a evitar qualquer forma de coerção implícita decorrente do uso de canais institucionais ou eventual associação com docentes.

Nos termos da regulamentação brasileira, trata-se de pesquisa com seres humanos e, ainda que classificada como risco mínimo, requer submissão prévia a Comitê de Ética em Pesquisa.

Do ponto de vista da LGPD, configura-se tratamento de dados pessoais, em princípio não sensíveis, exigindo definição explícita de finalidade, base legal adequada, prazo de retenção e procedimentos de descarte. Caso as perguntas envolvam aspectos de saúde ou vulnerabilidade psicológica, o enquadramento poderá envolver dados pessoais sensíveis, demandando cautela adicional.

Um redesenho metodológico eticamente adequado incluiria a explicitação inequívoca do caráter voluntário da pesquisa no convite, a apresentação de TCLE antes do acesso ao formulário, a descrição objetiva de riscos potenciais, a definição clara de prazo de retenção dos dados e a adoção de estratégias que reduzam o risco de identificação indireta. Os principais ajustes são sintetizados no Quadro 2.5.1, que orientou a elaboração da versão com configuração eticamente estruturada correspondente.

Quadro 2.5.1: Cenário 1 – Síntese comparativa

Elemento	**Versão original**	**Redesenho adequado**	**Princípio**
Seleção da população	Não justificada	Pertinência científica explicitada	Justiça
Recrutamento	Convite por e-mail institucional	Separação explícita entre pesquisa e avaliação	Justiça
Consentimento	Não explicitado	TCLE detalhado com riscos e direitos	Respeito
Riscos	Não descritos	Informação clara sobre riscos potenciais	Beneficência
Retenção	Não descrita	Prazo definido e descarte seguro	Beneficência / LGPD
Anonimização	Ausência de identificadores diretos	Estratégias para evitar identificação indireta	Beneficência

Cenário hipotético 1 (configuração metodológica eticamente estruturada): Questionário com estudantes universitários

Uma equipe de pesquisa desenvolveu um questionário online sobre dificuldades de gerenciamento de tempo e enviou convite aos estudantes por meio de mensagem institucional que explicitava tratar-se de pesquisa voluntária, desvinculada de qualquer avaliação acadêmica. A escolha da população foi justificada pela pertinência científica do fenômeno investigado ao contexto universitário, e não apenas por facilidade de acesso institucional. Antes do acesso ao formulário, foi apresentado Termo de Consentimento Livre e Esclarecido contendo objetivos detalhados, descrição de riscos potenciais,

informações sobre confidencialidade, prazo de retenção e procedimentos de descarte dos dados. O questionário foi estruturado de modo a evitar combinações de respostas que permitissem identificação indireta, e os resultados seriam divulgados exclusivamente de forma agregada. Foi assegurado que a recusa ou desistência não implicaria qualquer consequência acadêmica.

2.5.2. Cenário hipotético 2: Uso de realidade virtual em contexto de aula

Cenário hipotético 2: Uso de realidade virtual em contexto de aula

Uma equipe de pesquisa desenvolveu um jogo que utilizava recursos de realidade virtual com o objetivo de avaliar uma nova técnica de interação. Após acordo prévio com a pessoa docente responsável por uma disciplina ministrada em laboratório de Computação, a equipe realizou a atividade durante o horário regular da aula, disponibilizando o jogo aos estudantes. A pessoa docente optou por não permanecer no ambiente durante a sessão e assegurou a presença em aula a todas as pessoas estudantes que participaram da atividade com o jogo.

O cenário caracteriza pesquisa envolvendo seres humanos, uma vez que há aplicação de intervenção experimental com estudantes em ambiente institucional. A atividade ocorre durante o horário regular da aula, o que introduz assimetria estrutural relevante, ainda que a docente não permaneça fisicamente no ambiente.

A principal questão ética refere-se à voluntariedade da participação. A realização da atividade no tempo formal da disciplina pode gerar percepção de obrigatoriedade ou expectativa implícita de adesão. Mesmo na ausência da docente, a associação com a disciplina e a contabilização de presença podem configurar coerção indireta.

O princípio do respeito às pessoas exige garantia efetiva de participação voluntária, com possibilidade real de recusa ou retirada sem qualquer impacto acadêmico. O princípio da beneficência demanda avaliação prévia dos riscos físicos e psicológicos associados ao uso de realidade virtual, tais como tontura, náusea, desconforto visual ou sobrecarga sensorial, bem como definição de protocolos de interrupção imediata da atividade. O princípio da justiça requer que a participação não seja vinculada à frequência ou avaliação da disciplina.

Nos termos da regulamentação brasileira, trata-se de pesquisa com intervenção e requer submissão prévia ao Comitê de Ética em Pesquisa, ainda que classificada como risco mínimo ou moderado, dependendo da intensidade da imersão tecnológica.

Do ponto de vista da LGPD, caso haja registro de desempenho, comportamento, métricas biométricas ou logs de interação individualizados, configura-se tratamento de dados pessoais, exigindo definição de finalidade, base legal, prazo de retenção e medidas de segurança adequadas.

Uma configuração metodológica eticamente estruturada incluiria a realização da atividade fora do horário regular da aula ou a oferta de atividade alternativa equivalente para quem optasse por não participar, a apresentação prévia de TCLE com descrição de

riscos físicos e psicológicos, a explicitação inequívoca da ausência de impacto na frequência ou avaliação acadêmica e a presença de equipe capacitada para interromper imediatamente a atividade em caso de desconforto. Os principais ajustes são sintetizados no Quadro 2.5.2, que orientou a elaboração da versão com configuração eticamente estruturada correspondente.

Quadro 2.5.2: Cenário 2 – Síntese comparativa

Elemento	**Versão original**	**Redesenho adequado**	**Princípio**
Seleção da população	Não justificada	Pertinência científica explicitada	Justiça
Contexto	Durante horário regular de aula	Fora do horário ou com atividade alternativa	Justiça
Voluntariedade	Potencial coerção implícita	Participação explicitamente voluntária	Respeito
Riscos físicos	Não descritos	Descrição prévia e protocolo de interrupção	Beneficência
Impacto acadêmico	Presença vinculada à atividade	Garantia formal de ausência de impacto	Justiça
Proteção de dados	Não especificada	Anonimização e retenção definida	Beneficência / LGPD

Cenário hipotético 2 (configuração metodológica eticamente estruturada): Uso de realidade virtual em contexto de aula

Uma equipe de pesquisa desenvolveu um jogo em realidade virtual para avaliar técnica de interação e convidou estudantes a participar voluntariamente em horário distinto da aula regular, sem qualquer vínculo com frequência ou avaliação acadêmica. A escolha da população foi fundamentada na relação direta entre o contexto universitário e o fenômeno investigado, assegurando que a seleção não decorresse apenas de proximidade institucional. Antes da participação, foi apresentado Termo de Consentimento Livre e Esclarecido contendo objetivos do estudo, descrição detalhada dos possíveis riscos físicos e psicológicos, garantia de retirada a qualquer momento sem prejuízo e informações sobre confidencialidade, forma de armazenamento e destino dos dados após a conclusão da pesquisa. Durante a sessão, esteve presente equipe treinada para monitoramento e interrupção imediata em caso de desconforto. Eventuais dados de interação foram anonimizados antes da análise e utilizados exclusivamente para as finalidades informadas no consentimento.

2.5.3. Cenário hipotético 3: Jogo digital com crianças e uso de imagem

Cenário hipotético 3: Jogo digital com crianças e uso de imagem

Uma equipe de pesquisa desenvolveu um aplicativo que consistia em um jogo de memória destinado a crianças. Com o consentimento da pessoa docente, a equipe visitou a creche da universidade levando tablets com o jogo instalado.

Durante o período da aula, o jogo foi disponibilizado às crianças para uso. Para participar da atividade, as crianças utilizaram a câmera do tablet para tirar uma fotografia, que foi empregada na criação de um avatar no jogo. Ao final da aula, a equipe recolheu os tablets, que armazenavam os dados de interação das crianças, e, no momento da despedida, entregou um bombom a cada criança.

O cenário envolve pesquisa com crianças, grupo que demanda proteção ética reforçada em razão da capacidade limitada de autodeterminação. A autorização da docente não substitui o consentimento dos responsáveis legais nem supre a necessidade de assentimento das próprias crianças, quando compatível com sua faixa etária.

O uso da imagem para criação de avatar configura coleta de dado pessoal, com potencial sensibilidade ampliada em razão da idade das participantes. Além disso, o armazenamento de dados de interação individualizados caracteriza tratamento de dados pessoais. A LGPD, em seu art. 14, estabelece que o tratamento de dados de crianças deve ocorrer em seu melhor interesse, mediante consentimento específico e em destaque de pelo menos um dos responsáveis legais.

Do ponto de vista ético, o princípio do respeito às pessoas exige consentimento formal dos responsáveis e assentimento da criança sempre que possível. O princípio da beneficência demanda avaliação dos riscos associados à coleta e armazenamento de imagens, bem como à eventual possibilidade de identificação futura. A oferta de bombom ao final da atividade exige análise cuidadosa, pois pode configurar indução indevida à participação, ainda que de baixo valor econômico. O princípio da justiça requer atenção ao contexto institucional, evitando que a atividade seja percebida como obrigatória em razão do ambiente escolar.

Nos termos da regulamentação brasileira, trata-se de pesquisa com grupo vulnerável e com coleta de imagem, o que torna obrigatória a submissão prévia ao Comitê de Ética em Pesquisa, com análise criteriosa do protocolo.

Uma configuração metodológica eticamente estruturada incluiria a obtenção prévia de consentimento formal dos responsáveis legais, assentimento das crianças de forma compatível com sua compreensão, substituição da fotografia real por avatar préconfigurado ou imagem gerada internamente pelo sistema, definição clara de política de armazenamento e descarte dos dados e eliminação de qualquer incentivo material vinculado à participação. Os principais ajustes são sintetizados no Quadro 2.5.3, que orientou a elaboração da versão com configuração eticamente estruturada correspondente.

Cenário hipotético 3 (configuração metodológica eticamente estruturada): Jogo digital com crianças e uso de imagem

Uma equipe de pesquisa desenvolveu um jogo digital para crianças e realizou a atividade na creche universitária apenas após obtenção de autorização formal da instituição, consentimento dos responsáveis legais e assentimento das próprias crianças, em formato compatível com sua idade. A seleção da creche foi justificada pela adequação etária ao objetivo do jogo, com análise

Quadro 2.5.3: Cenário 3 – Síntese comparativa

Elemento	**Versão original**	**Redesenho adequado**	**Princípio**
Consentimento	Autorização da docente	Consentimento dos responsáveis + assentimento	Respeito
Uso de imagem	Fotografia real para avatar	Avatar genérico sem captura de imagem	Beneficência / LGPD
Grupo vulnerável	Crianças em ambiente escolar	Proteção reforçada e voluntariedade explícita	Justiça
Incentivo	Bombom ao final da atividade	Ausência de incentivo material	Beneficência
Armazenamento	Dados mantidos nos tablets	Anonimização e descarte definido	Beneficência / LGPD

prévia da distribuição de riscos e benefícios. O jogo foi estruturado sem utilização de fotografia real, empregando avatares genéricos previamente configurados. Os dados de interação foram coletados de forma anonimizada e armazenados em ambiente seguro pelo período necessário à análise da pesquisa, com descarte posterior. Não houve oferta de qualquer incentivo material vinculado à participação. A participação foi explicitamente voluntária, com possibilidade de recusa sem qualquer consequência.

2.5.4. Cenário hipotético 4: Aplicativo para pessoas idosas e saúde mental

Cenário hipotético 4: Aplicativo para pessoas idosas e saúde mental

Uma equipe de pesquisa desenvolveu um aplicativo destinado ao monitoramento da saúde mental de pessoas idosas, com foco na avaliação de aspectos como humor e ansiedade. A equipe convidou pessoas idosas que participavam de atividades de extensão da universidade, realizadas em um bairro de classe média, para uma sessão de teste do aplicativo. Antes do início da atividade, foi fornecida uma explicação verbal sobre o funcionamento do aplicativo. Também antes da sessão, a equipe registrou imagens das pessoas participantes para o mural digital do programa de pós-graduação. Ao final de cada sessão, as pessoas idosas receberam um vale para consumo de café na cantina da universidade.

O cenário envolve tratamento de dados pessoais sensíveis, uma vez que há coleta de informações relacionadas à saúde mental, nos termos do art. 5º, II, da LGPD. Além disso, embora pessoas idosas não constituam automaticamente grupo vulnerável, podem encontrar-se em condição de vulnerabilidade situacional, especialmente quando vinculadas a programas institucionais ou quando há assimetria informacional relevante.

Do ponto de vista ético, o princípio do respeito às pessoas exige consentimento livre, esclarecido e específico, inclusive para o registro e divulgação de imagens. A mera explicação verbal sobre o funcionamento do aplicativo não supre, por si só, a necessidade de consentimento formal documentado, especialmente quando há coleta de dados sensíveis e uso de imagem para fins institucionais.

O princípio da beneficência demanda avaliação cuidadosa dos riscos de estigmatização, exposição indevida de informações relacionadas à saúde mental e eventuais impactos psicológicos decorrentes da autoavaliação de humor e ansiedade. Também requer definição de protocolos para encaminhamento, caso sejam identificados indícios de sofrimento psíquico relevante.

O princípio da justiça orienta a análise dos critérios de recrutamento e da adequação do benefício oferecido. O vale-café pode ser compreendido como forma de ressarcimento ou agradecimento simbólico; contudo, deve-se avaliar se há risco de indução indevida à participação, considerando o contexto socioeconômico e a natureza da atividade.

Nos termos da regulamentação brasileira, trata-se de pesquisa com coleta de dados sensíveis e uso de imagem, o que torna obrigatória a submissão prévia ao Comitê de Ética em Pesquisa.

Uma configuração metodológica eticamente estruturada incluiria a apresentação de Termo de Consentimento Livre e Esclarecido por escrito, com descrição detalhada da finalidade do estudo, dos dados coletados, dos riscos potenciais, das medidas de segurança adotadas, do prazo ou critério de retenção dos dados e do direito de retirada a qualquer momento. O uso de imagem deveria ser objeto de consentimento específico e destacado, com possibilidade de participação no estudo independentemente da autorização para divulgação de fotografia. Os principais ajustes são sintetizados no Quadro 2.5.4, que orientou a elaboração da versão com configuração eticamente estruturada correspondente.

Quadro 2.5.4: Cenário 4 – Síntese comparativa

Elemento	**Versão original**	**Redesenho adequado**	**Princípio**
Dados coletados	Humor e ansiedade	Dados sensíveis com proteção reforçada	Beneficência / LGPD
Consentimento	Explicação verbal	TCLE escrito e detalhado	Respeito
Uso de imagem	Registro para mural institucional	Consentimento específico e facultativo	Respeito / LGPD
Seleção da população	Não justificada	Pertinência científica explicitada	Justiça
Benefício	Vale-café	Ressarcimento simbólico não vinculante	Justiça
Proteção psicológica	Não descrita	Protocolo de orientação e encaminhamento	Beneficência

Cenário hipotético 4 (configuração metodológica eticamente estruturada): Aplicativo para pessoas idosas e saúde mental

Uma equipe de pesquisa desenvolveu aplicativo para acompanhamento de aspectos relacionados ao humor e à ansiedade em pessoas idosas participantes de programa de extensão universitária. A atividade foi planejada como sessão voluntária, com convite individualizado e esclarecimento pré-

vio sobre objetivos, procedimentos e natureza da coleta de dados sensíveis. A definição da população participante foi fundamentada na pertinência do estudo ao grupo etário, evitando recorte meramente conveniente ou excludente. Antes da participação, foi apresentado Termo de Consentimento Livre e Esclarecido por escrito, contendo descrição detalhada dos dados coletados, finalidade da pesquisa, medidas de confidencialidade, critérios de armazenamento e descarte, além da garantia de retirada a qualquer momento sem qualquer prejuízo. O tratamento das informações foi estruturado com pseudonimização e acesso restrito à equipe responsável. O registro de imagem foi tratado como procedimento independente, submetido a consentimento específico e facultativo, não condicionado à participação no estudo. Foram previstos mecanismos de orientação e encaminhamento em caso de identificação de sofrimento psíquico relevante. Eventual oferta de vale-café foi caracterizada como gesto de cortesia institucional, não condicionado à adesão nem configurado como incentivo à participação.

2.5.5. Cenário hipotético 5: Pesquisa com egressos e termo de consentimento

Cenário hipotético 5: Pesquisa com egressos e termo de consentimento

Uma equipe de pesquisa investigou a prática profissional de egressos de cursos de Computação que atuavam nas áreas de Web e Multimídia. Para a coleta de dados, foram enviados questionários online a ex-discentes de uma universidade, com o objetivo de levantar informações sobre suas experiências profissionais, desafios enfrentados e habilidades aplicadas no mercado de trabalho. Os convites para participação foram encaminhados por e-mail, contendo um link para o formulário de pesquisa.
Como etapa inicial do questionário, foi apresentado um Termo de Consentimento Livre e Esclarecido (TCLE), com o qual as pessoas participantes deveriam concordar para prosseguir. O TCLE era sucinto, foi avaliado apenas pela própria equipe de pesquisa e não detalhava os objetivos ou os riscos do estudo, tampouco as providências adotadas para a minimização desses riscos. O questionário incluía perguntas sobre experiências, conhecimentos e rotinas educacionais, atividades profissionais atuais, competências desenvolvidas e dificuldades encontradas no exercício da profissão.

O cenário envolve pesquisa com seres humanos por meio de coleta estruturada de informações pessoais relativas à trajetória acadêmica e profissional. Embora não haja indicação de dados sensíveis, as informações podem permitir identificação indireta, especialmente quando associadas a áreas específicas de atuação ou contextos profissionais restritos.

A existência formal de um TCLE não assegura, por si só, adequação ética. O princípio do respeito às pessoas exige consentimento efetivamente esclarecido, com descrição adequada dos objetivos do estudo, dos riscos envolvidos, das garantias de confidencialidade, dos direitos das pessoas participantes e das condições de retirada. A insuficiência informacional compromete a qualidade do consentimento.

O princípio da beneficência orienta a análise dos riscos reputacionais e profissionais decorrentes da eventual divulgação ou uso inadequado das informações coletadas. O princípio da justiça exige atenção ao modo de recrutamento, garantindo voluntariedade e ausência de expectativa institucional implícita.

Nos termos da regulamentação brasileira, trata-se de pesquisa com seres humanos e requer submissão prévia ao Comitê de Ética em Pesquisa, ainda que classificada como risco mínimo.

Do ponto de vista da LGPD, configura-se tratamento de dados pessoais, exigindo definição clara de finalidade, base legal, medidas de segurança e critérios de retenção e descarte.

Uma configuração metodológica eticamente estruturada incluiria a elaboração de Termo de Consentimento Livre e Esclarecido detalhado, contendo objetivos específicos da pesquisa, descrição dos dados coletados, riscos potenciais – inclusive reputacionais –, garantias de confidencialidade, critérios de retenção e descarte e direito de retirada a qualquer momento. Também demandaria submissão prévia do protocolo a Comitê de Ética em Pesquisa, definição explícita de estratégias para evitar identificação indireta e divulgação exclusivamente agregada dos resultados. Esses ajustes são sintetizados no Quadro 2.5.5, que orientou a elaboração da versão com configuração eticamente estruturada correspondente.

Quadro 2.5.5: Cenário 5 – Síntese comparativa

Elemento	**Versão original**	**Configuração eticamente estruturada**	**Princípio**
TCLE	Sucinto e pouco detalhado	TCLE completo e submetido ao CEP	Respeito
Riscos	Não explicitados	Riscos reputacionais descritos	Beneficência
Anonimização	Não especificada	Estratégias para evitar identificação indireta	Beneficência / LGPD
Seleção da população	Não justificada	Pertinência científica explicitada	Justiça
Recrutamento	Convite por e-mail	Participação voluntária sem vínculo institucional	Justiça
Proteção de dados	Não descrita	Finalidade, retenção e descarte definidos	LGPD

Cenário hipotético 5 (configuração metodológica eticamente estruturada): Pesquisa com egressos e termo de consentimento

Uma equipe de pesquisa desenvolveu estudo com egressos de cursos de Computação, com coleta de dados por questionário online sobre trajetória acadêmica, inserção e prática profissional, desafios enfrentados e competências utilizadas no mercado. O recrutamento foi realizado por convite eletrônico individual, com caráter explicitamente voluntário e sem qualquer vínculo com avaliação institucional, benefícios acadêmicos ou obrigações associa-

das à condição de egresso. A seleção dos egressos foi justificada pela pertinência do grupo ao objetivo da investigação, dado que a análise incidia especificamente sobre trajetórias profissionais de formados no curso. Antes do acesso ao questionário, foi apresentado Termo de Consentimento Livre e Esclarecido (TCLE) detalhado, contendo objetivos específicos do estudo, descrição do tipo de informações coletadas, riscos potenciais – inclusive reputacionais e profissionais –, garantias de confidencialidade, medidas de segurança e controle de acesso, critério de retenção e procedimento de descarte, direito de recusa e de retirada a qualquer momento e canais de contato da equipe e do Comitê de Ética em Pesquisa. O protocolo e o TCLE foram submetidos previamente à apreciação por CEP. O instrumento foi estruturado para reduzir possibilidade de identificação indireta, evitando combinações excessivamente específicas, com uso de categorias agregadas quando apropriado. A análise foi conduzida com pseudonimização e acesso restrito à equipe, e os resultados foram divulgados exclusivamente de forma agregada, sem referência individualizada a trajetórias pessoais.

2.5.6. Estruturas analíticas em cenários fictícios

Os cenários apresentados nesta seção são construções hipotéticas, elaboradas com finalidade exclusivamente didática e analítica. Não correspondem a casos reais nem fazem referência a indivíduos, instituições ou publicações específicas.

Cada cenário foi estruturado para representar configurações possíveis de pesquisa em Computação com participação humana direta. A intenção não é afirmar a ocorrência ou frequência dessas configurações no campo científico, mas explicitar, de maneira controlada, como determinadas combinações de decisões metodológicas podem gerar questões normativas relevantes.

Nos exemplos apresentados, são combinados elementos como coleta de dados com participantes, aplicação de instrumentos de pesquisa, validação empírica ou tratamento de informações potencialmente identificáveis. O foco recai sobre a análise do enquadramento normativo dessas configurações, especialmente quanto à necessidade de apreciação ética formal e às implicações relacionadas ao tratamento de dados.

Ao trabalhar com narrativas fictícias, é possível examinar os critérios de decisão envolvidos sem expor situações concretas ou atribuir responsabilidades específicas. O objetivo é oferecer ao leitor um instrumento conceitual para reconhecer, em contextos reais, quando uma determinada configuração metodológica demanda análise ética estruturada.

Para discussões aprofundadas sobre o panorama metacientífico da explicitação de aspectos éticos na Computação brasileira, recomenda-se a consulta à tese de doutorado de Luiz Paulo Carvalho, que desenvolve análise sistemática sobre o tema [Carvalho 2024].

2.6. Experiência empírica prévia das autoras

A reflexão desenvolvida neste tutorial apoia-se na trajetória acumulada das autoras na condução de pesquisas empíricas em Computação envolvendo participação humana direta, grupos em situação de vulnerabilidade e tratamento de dados pessoais, inclusive

sensíveis. Em todos os casos, os estudos foram previamente aprovados por instâncias institucionais de avaliação ética, conforme as normas vigentes nos respectivos países.

No campo das intervenções digitais em saúde e educação, destacam-se pesquisas sobre autoria web e métodos de intervenção programada [Cunha et al. 2021, Cunha et al. 2019, Rodrigues et al. 2018], incluindo o desenvolvimento do Método de Amostragem de Experiência e Intervenção Programada (ESPIM) e de sua infraestrutura para planejamento e monitoramento de intervenções mHealth. Esses estudos envolveram entrevistas com especialistas, estudos de caso reais e interação direta com participantes, exigindo atenção a consentimento informado, confidencialidade e proporcionalidade entre riscos e benefícios. De modo complementar, o sistema *Media Parcels* foi desenvolvido e avaliado no Reino Unido, em colaboração com pesquisador da University of Sussex [Zaine et al. 2019] e com aprovação por comitê de ética institucional.

Em contextos de vulnerabilidade, foram conduzidos estudos com jogos digitais para crianças [Verhalen and Rodrigues 2024], incluindo iniciativas baseadas em design participativo e *codesign* com docentes do ensino fundamental [Rodrigues et al. 2020], bem como aplicações educacionais e de alfabetização digital com pessoas idosas [Cliquet et al. 2023], investigações educacionais remotas com adolescentes [Eusebio and Pimentel 2025] e pesquisas com estudantes universitários [Sanches et al. 2022]. Essas investigações demandaram especial cuidado com consentimento de responsáveis legais, assentimento de menores, avaliação de riscos psicossociais e proteção da dignidade e da privacidade.

No âmbito de tecnologias assistivas, incluem-se pesquisas com interfaces baseadas em *eye-tracking* para pessoas com deficiência motora severa [Pedrosa et al. 2015a, Pedrosa et al. 2015b], conduzidas nos Estados Unidos e aprovadas por *Institutional Review Board* (IRB), envolvendo coleta de dados biométricos e rastreamento ocular. Também integram essa trajetória estudos sobre interfaces para TV Digital interativa [Rodrigues et al. 2011], diretrizes para aplicações IoT voltadas a pessoas idosas [Rodrigues et al. 2024b], dispositivos vestíveis para reabilitação pós-AVC [Pereira et al. 2025, Pimentel et al. 2025] e análises baseadas em bases de dados sensíveis anonimizadas na área de Gerontologia [Lima et al. 2021]. Esses estudos abrangeram desde pesquisas de risco mínimo até contextos clínicos e de reabilitação com maior complexidade ética, exigindo avaliação institucional quanto à proporcionalidade, monitoramento e proteção de participantes.

Os exemplos citados são representativos – embora não exaustivos – da experiência das autoras em pesquisas submetidas à apreciação ética institucional em diferentes contextos regulatórios. Essa trajetória fundamenta a abordagem adotada neste tutorial, que articula princípios normativos, exigências institucionais e desafios concretos da investigação científica em Computação.

2.7. Considerações finais

O conjunto de discussões realizadas no tutorial evidencia que a ética em pesquisa em Computação não se reduz ao cumprimento burocrático de normas, mas envolve decisões situadas, justificadas e transparentes ao longo de todo o processo de pesquisa. Essas decisões se materializam tanto em escolhas metodológicas quanto na observância consci-

ente das estruturas regulatórias que orientam pesquisas com participação humana direta e aquelas baseadas no tratamento de dados pessoais.

Ao trabalhar com cenários fictícios, o tutorial buscou promover uma postura reflexiva e responsável, capacitando pesquisadores a reconhecer quando a avaliação ética é necessária em estudos com interação direta com pessoas, quais questões devem ser consideradas e como integrar a ética ao desenho metodológico dos estudos, de forma consistente com as exigências do sistema nacional de ética em pesquisa, reconhecido e consolidado pela Lei nº 14.874/2024, e com a legislação vigente, em especial a LGPD.

Embora o tutorial não tenha abordado exemplos concretos de pesquisas publicadas, nem explorado de forma aprofundada cenários baseados em dados pessoais obtidos indiretamente, o exercício realizado fornece uma base conceitual e normativa sólida para que pesquisadores possam analisar criticamente seus próprios projetos, justificar suas decisões éticas e dialogar de maneira mais informada, transparente e responsável com Comitês de Ética em Pesquisa e demais instâncias institucionais.

2.8. Cuidados Éticos e Proteção de Dados

Este trabalho não envolveu pesquisa empírica com seres humanos, nem coleta, tratamento ou análise de dados pessoais, diretos ou indiretos, nos termos das resoluções do Conselho Nacional de Saúde e da Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/2018).

As discussões apresentadas têm caráter exclusivamente conceitual, normativo e didático. Os cenários analisados são hipotéticos e não correspondem a estudos reais conduzidos pelas autoras. Não houve interação com participantes humanos, nem uso de bases de dados pessoais ou rastros digitais associados a pessoas naturais. Dessa forma, o trabalho não se enquadra como pesquisa envolvendo seres humanos, não requer apreciação pelo Sistema Nacional de Ética em Pesquisa e não configura tratamento de dados pessoais nos termos da LGPD.

2.9. Agradecimentos

As autoras agradecem às pesquisadoras e aos pesquisadores, colaboradoras e coautoras com quem desenvolveram projetos, publicações acadêmicas e atividades de formação envolvendo participação humana e tratamento de dados pessoais, os quais demandaram submissão e apreciação por Comitês de Ética em Pesquisa. As reflexões apresentadas neste texto resultam dessa trajetória acumulada.

Agradecem, de modo particular, a André Pimenta Freire e Luiz Paulo Carvalho pela parceria em publicações anteriores referenciadas ao longo do texto, apresentadas no WebMedia 2023 e no GranDIHC-BR 2025–2035 (*GC2: Ethics and Responsibility*), cujas contribuições foram relevantes para a consolidação das análises aqui desenvolvidas [Pimentel et al. 2023, Rodrigues et al. 2024a].

Um modelo de linguagem de grande porte (LLM) foi utilizado como ferramenta auxiliar de apoio linguístico. Todas as decisões conceituais, estruturais e argumentativas, bem como a responsabilidade pelo conteúdo, são exclusivamente das autoras.

Referências

[Albres and Sousa 2019] Albres, N. A. and Sousa, D. V. C. (2019). Termo de assentimento livre e esclarecido: uso de história em quadrinhos em pesquisas com crianças. *Revista Sinalizar*, 4.

[Belmont 1979] Belmont (1979). *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research.

[Carvalho 2024] Carvalho, L. P. (2024). *Ética Computacional no Brasil: uma análise metacientífica da prática acadêmico-científico computacional brasileira*. Tese de doutorado, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

[Caspar 2021] Caspar, E. A. (2021). A novel experimental approach to study disobedience to authority. *Scientific Reports*, 11(1):22927.

[Cliquet et al. 2023] Cliquet, L. O. B. V., Pimentel, M. d. G. C., Batistoni, S. S. T., Rodrigues, K. R. d. H., Zaine, I., and Cachioni, M. (2023). Use of smartphones by older adults: characteristics and reports of students enrolled at a University of the Third Age (U3A). *PerCursos*, 24:1–30.

[CNS 1996] CNS (1996). *Resolução CNS nº 196, de 10 de outubro de 1996*. Conselho Nacional de Saúde (Brasil). Diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos.

[CNS 2012] CNS (2012). *Resolução CNS nº 466, de 12 de dezembro de 2012*. Conselho Nacional de Saúde (Brasil). Diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos.

[CNS 2016] CNS (2016). *Resolução CNS nº 510, de 7 de abril de 2016*. Conselho Nacional de Saúde (Brasil). Define as especificidades éticas das pesquisas em Ciências Humanas e Sociais.

[CNS 2022] CNS (2022). *Resolução CNS nº 674, de 6 de maio de 2022*. Conselho Nacional de Saúde (Brasil). Atualiza o Sistema CEP/CONEP e orienta sobre pesquisa com dados pessoais e digitais.

[CONEP 2022] CONEP (2022). *Ofício Circular nº 17, de 5 de julho de 2022: Orientações acerca do artigo 1º da Resolução CNS nº 510, de 7 de abril de 2016*. Comissão Nacional de Ética em Pesquisa (CONEP). Atualizado em 23 abr. 2025.

[Cunha et al. 2019] Cunha, B. C., Rodrigues, K. R., and Pimentel, M. d. G. C. (2019). Synthesizing guidelines for facilitating elderly-smartphone interaction. In *Proceedings of the 25th Brazillian Symposium on Multimedia and the Web*, pages 37–44.

[Cunha et al. 2021] Cunha, B. C. R., Rodrigues, K. R. D. H., Zaine, I., da Silva, E. A. N., Viel, C. C., and Pimentel, M. D. G. C. (2021). Experience Sampling and Programmed Intervention Method and System for Planning, Authoring, and Deploying Mobile Health Interventions: Design and Case Reports. *J Med Internet Res*, 23(7):e24278.

[Decreto12651 2025] Decreto12651 (2025). *Decreto nº 12.651, de 7 de outubro de 2025 – Regulamenta a Lei nº 14.874, de 28 de maio de 2024, que dispõe sobre a pesquisa com seres humanos e institui o Sistema Nacional de Ética em Pesquisa com Seres Humanos*. Brasil. Publicado no Diário Oficial da União em 8 de outubro de 2025.

[Dickerson 2015a] Dickerson, C. (2015a). The VA's Broken Promise To Thousands Of Vets Exposed To Mustard Gas. National Public Radio.

[Dickerson 2015b] Dickerson, C. (2015b). U.S. Troops Were Tested by Race in Secret World War II Chemical Experiments. National Public Radio.

[Eusebio and Pimentel 2025] Eusebio, J. and Pimentel, M. (2025). Avaliação Retrospectiva da Experiência de uma Coorte de Egressas de um Curso de Introdução à Programação para Alunas do Ensino Médio e Concluintes. In *Anais do XIX Women in Information Technology*, pages 275–286, Porto Alegre, RS, Brasil. SBC.

[Gaśińska 2015] Gaśińska, A. (2015). The contribution of women to radiobiology: Marie Curie and beyond. *Reports of Practical Oncology and Radiotherapy*, 21(3):250–258.

[Kühne 2018] Kühne, T. (2018). *Nazi Morality*, chapter 13, pages 215–229. John Wiley & Sons, Ltd.

[Lei14874 2024] Lei14874 (2024). *Lei nº 14874, de 28 de maio de 2024 – Dispõe sobre a pesquisa com seres humanos e institui o Sistema Nacional de Ética em Pesquisa com Seres Humanos*. Brasil.

[LGPD 2018] LGPD (2018). *Lei nº 13.709, de 14 de agosto de 2018 – Lei Geral de Proteção de Dados Pessoais (LGPD)*. Brasil. Redação dada pela Lei nº 13.853, de 2019.

[Lima et al. 2021] Lima, A. P. d., Dantas, L. A., Manzato, M. G., Pimentel, M., Orlandi, B., and Castro, P. (2021). An Interpretable Recommendation Model for Gerontological Care. In *Proceedings of the 15th ACM Conference on Recommender Systems*, RecSys '21, page 620–626, New York, NY, USA. Association for Computing Machinery.

[Mandal et al. 2011] Mandal, J., Acharya, S., and Parija, S. C. (2011). Ethics in human research. *Tropical Parasitology*, 1(1):2–3.

[Mellanby 1947] Mellanby, K. (1947). Medical Experiments on Human Beings in Concentration Camps in Nazi Germany. *British Medical Journal*, 1(4490):148–150. Published 25 January 1947.

[Milgram 1963] Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, 67(4):371–378.

[Museum 2025] Museum, U. S. H. M. (2025). Nazi medical experiments. Accessed Feb. 28, 2026.

[Pedrosa et al. 2015a] Pedrosa, D., Pimentel, M. d. G., and Truong, K. N. (2015a). Filteryedping: A dwell-free eye typing technique. In *Proceedings of the 33rd annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 303–306.

[Pedrosa et al. 2015b] Pedrosa, D., Pimentel, M. D. G., Wright, A., and Truong, K. N. (2015b). Filteryedping: Design Challenges and User Performance of Dwell-Free Eye Typing. *ACM Trans. Access. Comput.*, 6(1).

[Pereira et al. 2025] Pereira, A. P., Machado Neto, O. J., Elui, V. M. C., and Pimentel, M. d. G. C. (2025). Wearable Smartphone-Based Multisensory Feedback System for Torso Posture Correction: Iterative Design and Within-Subjects Study. *JMIR Aging*, 8:e55455.

[Pimentel et al. 2023] Pimentel, M. d. G. C., Freire, A. P., Carvalho, L. P., and Rodrigues, K. R. d. H. (2023). Submissão de Projetos de Pesquisa Web e Multimídia a Comitês de Ética em Pesquisa. In Marques Neto, M. C., Pimentel, M. d. G., Willrich, R., Macedo, A. A., and Goularte, R., editors, *Minicursos do XXIX Simpósio Brasileiro de Sistemas Multimídia e Web*. Sociedade Brasileira de Computação, Porto Alegre, RS, Brasil.

[Pimentel et al. 2025] Pimentel, M. d. G. C., Pereira, A. P., Machado Neto, O. J., Zimmermann, L. C., and Elui, V. M. C. (2025). Quantitative Evaluation of Postural SmartVest's Multisensory Feedback for Affordable Smartphone-Based Post-Stroke Motor Rehabilitation. *International Journal of Environmental Research and Public Health*, 22(7).

[Riedel 2005] Riedel, S. (2005). Edward Jenner and the History of Smallpox and Vaccination. *Proceedings (Baylor University Medical Center)*, 18(1):21–25.

[Rodrigues et al. 2024a] Rodrigues, K., Carvalho, L., Freire, A., and Pimentel, M. (2024a). GranDIHC-BR 2025-2035 - GC2: Ethics and Responsibility: Principles, Regulations, and Societal Implications of Human Participation in HCI Research. In *Anais do XXIII Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais*, pages 969–987, Porto Alegre, RS, Brasil. SBC.

[Rodrigues et al. 2018] Rodrigues, K. R., Cunha, B. C., Zaine, I., Viel, C. C., Scalco, L. F., and Pimentel, M. G. (2018). ESPIM system: interface evolution to enable authoring and interaction with multimedia intervention programs. In *Proceedings of the 24th Brazilian Symposium on Multimedia and the Web*, pages 125–132.

[Rodrigues et al. 2011] Rodrigues, K. R., Pereira, S. S., Quinelato, L. G., Melo, E. L., Neris, V. P., and Teixeira, C. A. (2011). Enhancing TV experience with additional content and multiple device interactivity. In *Proceedings of the 17th Brazilian Symposium on Multimedia and the Web on Brazilian Symposium on Multimedia and the Web - Volume 1*, WebMedia 2011, page 18–25, Porto Alegre, BRA. Brazilian Computer Society.

[Rodrigues et al. 2020] Rodrigues, K. R. d. H., Silva Junior, J. M., Nespule, L., Maia, M. S., Garcia, L. E., and Galati, H. B. (2020). Lume: Um jogo digital para incentivar a

leitura de lendas do folclóre brasileiro. In *Anais do XIX Simpósio Brasileiro de Jogos e Entretenimento Digital*, pages 1240–1249.

[Rodrigues et al. 2024b] Rodrigues, S., Fortes, R., and Rodrigues, K. (2024b). Guidelines for designing iot applications for older adults. In *Anais do XXIII Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais*, pages 472–486, Porto Alegre, RS, Brasil. SBC.

[Sanches et al. 2022] Sanches, R., Ponti, M., and Rodrigues, K. (2022). Evasão universitária e estratégias para retenção de alunos com base em intervenções remotas. In *Anais Estendidos do XXI Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais*, pages 84–87, Porto Alegre, RS, Brasil. SBC.

[Shuster 1997] Shuster, E. (1997). Fifty years later: The significance of the nuremberg code. *The New England Journal of Medicine*, 337(20):1436–1440.

[Varkey 2021] Varkey, B. (2021). Principles of clinical ethics and their application to practice. *Medical Principles and Practice*, 30(1):17–28.

[Verhalen and Rodrigues 2024] Verhalen, A. E. C. and Rodrigues, K. R. d. H. (2024). The use of serious digital games to talk about grief and finitude with children: reports on the design and evaluation process of two games. *Journal on Interactive Systems*, 15(1):926–941.

[WMA 2024] WMA (2024). *(World Medical Association) Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Participants*. World Medical Association.

[Zaine et al. 2019] Zaine, I., Frohlich, D. M., Rodrigues, K. R. D. H., Cunha, B. C. R., Orlando, A. F., Scalco, L. F., and Pimentel, M. D. G. C. (2019). Promoting Social Connection and Deepening Relations Among Older Adults: Design and Qualitative Evaluation of Media Parcels. *J Med Internet Res*, 21(10):e14112.

Capítulo 3

UX Research Envolvendo Pessoas Com Deficiência: Um Caminho Para Uma Webmídia Inclusiva

Daniela Cardoso Tavares, Marcelo Martins da Silva, Sara Lobato, Pryscila Barbosa, Marcelo Bustamante, Mariana Faria, André Pimenta

Resumo

A web consolidou-se como um meio central para as relações sociais e o exercício da cidadania, alcançando 84% da população brasileira em 2024. Contudo, essa ampliação do acesso não garante condições equitativas de uso, visto que apenas 2,9% dos sites do país possuem condições adequadas para pessoas com deficiência. Nesse cenário, a webmídia inclusiva e a acessibilidade digital emergem como direitos fundamentais e desafios contemporâneos na computação para promover a autonomia e a inclusão social em conformidade com o Estatuto da Pessoa com Deficiência e a Agenda 2030 da ONU. Apesar dos avanços normativos, a acessibilidade ainda é tratada de forma secundária no desenvolvimento de sistemas. Para mitigar esse problema, a pesquisa em Experiência do Usuário (UX Research) apresenta-se como uma abordagem essencial para identificar barreiras e orientar decisões de projeto de forma inclusiva. Diante disso, o capítulo introduzido discute o papel da UX Research na promoção da acessibilidade em webmídia, dividindo-se em seções que abordam aspectos históricos, legais, conceituais e os desafios metodológicos de envolver pessoas com deficiência no processo de desenvolvimento.

3.1 Introdução

A web pode ser considerada, na contemporaneidade, um dos principais meios de comunicação da sociedade, ao possibilitar a produção, o acesso e o compartilhamento de informações sem a necessidade de que as pessoas estejam no mesmo espaço físico ou região geográfica (Póvoa, 2000). Nesse contexto, a Internet assume um papel central na mediação das relações sociais, educacionais, culturais, econômicas e políticas, consolidando-se como um espaço estratégico para o exercício da cidadania e da participação social (Castells, 2003). No Brasil, de acordo com dados da pesquisa TIC Domicílios 2024, conduzida pelo Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br), aproximadamente 159 milhões de pessoas (o que corresponde a 84% da população) possuem acesso à internet em suas residências (NIC.br, 2025). Todavia, a ampliação do acesso não garante, por si só, condições equitativas de

uso e participação, especialmente para pessoas com deficiência. Os ambientes digitais, assim como outros espaços sociais, ainda refletem desigualdades estruturais. No campo da acessibilidade digital, uma pesquisa realizada pela BigDataCorp, em parceria com o Movimento Web para Todos, revelou que apenas 2,9% dos *sites* brasileiros apresentam condições adequadas de uso para pessoas com deficiência. Em 2024, foram avaliados 26,3 milhões de sites, evidenciando limitações significativas na acessibilidade web brasileira (Bigdatacorp; Movimento Web para Todos, 2024). Esse cenário reforça a acessibilidade digital como um dos desafios contemporâneos da computação e da inclusão social, alinhando-se às discussões apresentadas nos Grandes Desafios da Computação no Brasil 2025–2035 (SBC, 2025).

Diante desse cenário, a webmídia inclusiva emerge como um campo voltado à promoção da equidade informacional e da inclusão social nos ambientes digitais (Tavares, 2019). Essa abordagem está alinhada às preocupações globais relacionadas ao desenvolvimento sustentável e à redução das desigualdades, conforme estabelecido pela Agenda 2030 da Organização das Nações Unidas (Brasil, 2016). A acessibilidade proporciona qualidade de vida, autonomia, participação social e inclusão, beneficiando diferentes perfis de usuários, como pessoas com deficiência, idosos, indivíduos com mobilidade reduzida e pessoas com baixo letramento. Nesse contexto, o inciso primeiro do artigo terceiro da Lei nº 13.146/2015 - Estatuto da Pessoa com Deficiência (Brasil, 2015) define acessibilidade como:

> possibilidade e condição de alcance para utilização, com segurança e autonomia, de espaços, mobiliários, equipamentos urbanos, edificações, transportes, informação e comunicação, inclusive seus sistemas e tecnologias, bem como de outros serviços e instalações abertos ao público, de uso público ou privados de uso coletivo, tanto na zona urbana como na rural, por pessoa com deficiência ou com mobilidade reduzida (Brasil, 2015).

No campo da Computação, a acessibilidade pode ser entendida como a possibilidade de utilização plena e autônoma de sistemas, recursos e serviços digitais por diferentes perfis de usuários. Essa perspectiva envolve o desenvolvimento de soluções que possibilitem uma interação eficiente, segura e confortável para o maior número possível de pessoas, considerando a diversidade de habilidades, contextos e formas de uso (Horton; Quesenbery, 2014). Entretanto, apesar dos avanços normativos e tecnológicos, ainda se observa uma lacuna significativa na incorporação efetiva da acessibilidade nos processos de desenvolvimento de sistemas digitais. Em muitos casos, a acessibilidade é tratada como um aspecto secundário, o que compromete a qualidade da experiência e limita a participação de usuários diversos.

Nesse contexto, a pesquisa em Experiência do Usuário (UX *Research*) apresenta-se como uma abordagem fundamental para compreender as necessidades, limitações e expectativas dos usuários, permitindo que decisões de projeto sejam orientadas por evidências, e não por suposições. Ao envolver diretamente os usuários no processo de desenvolvimento, a UX *Research* contribui para a identificação de barreiras de acessibilidade e para a construção de soluções mais inclusivas. No entanto, observa-se que a aplicação de métodos de UX *Research* envolvendo pessoas com deficiência ainda permanece pouco explorada, especialmente no contexto da webmídia. Essa lacuna evidencia a necessidade de aprofundar discussões e práticas relacionadas à condução de pesquisas mais inclusivas, capazes de contemplar diferentes perfis de usuários, múltiplas formas de interação e estratégias metodológicas adaptadas às demandas de acessibilidade e inclusão. Diante desse cenário, este capítulo tem como objetivo discutir o papel da UX Research na promoção da acessibilidade em Webmedia, com ênfase na condução de pesquisas envolvendo pessoas com deficiência. Para isso, são abordados aspectos históricos, legais e conceituais da acessibilidade digital, bem como desafios relacionados ao desenvolvimento de experiências digitais mais inclusivas. Assim, este capítulo está organizado

da seguinte forma: esta seção 1 introduz o tema e destaca a relevância e o objetivo do capítulo; a seção 2 apresenta a acessibilidade digital como um direito fundamental, abordando sua evolução histórica e legal. A seção 3 discute o papel da UX *Research* no desenvolvimento de webmídias inclusivas. A seção 4 explora a condução de pesquisas envolvendo pessoas com deficiência. A seção 5 aborda os principais desafios da acessibilidade em webmídias. A seção 6 apresenta um panorama das pesquisas sobre acessibilidade publicadas nos anais do WebMedia. A Seção 7 reúne as considerações finais. Por fim, as seções 8 e 9 apresentam, respectivamente, a declaração sobre o uso de inteligência artificial e as referências utilizadas na elaboração deste capítulo.

3.2. Acessibilidade Digital como Direito Fundamental

Historicamente, as pessoas com deficiência acompanharam as transformações sociais, culturais e tecnológicas ao longo do tempo. No entanto, apesar dos avanços políticos e legislativos, persistem desafios significativos para a promoção da equidade e da plena inclusão social. Esses desafios estão, em grande medida, relacionados a fatores culturais e à forma como a deficiência foi compreendida e interpretada em diferentes períodos históricos (Figueira, 2021) (Tavares, 2017). Durante a Idade Média, a deficiência era frequentemente associada a maldições, sendo a pessoa com deficiência entendida como um sujeito excluído da vida social. A partir do século XIX, com o fortalecimento do pensamento científico, consolidou-se o modelo médico da deficiência (Júnior, 2010).

Nesse paradigma, o foco recai sobre o indivíduo e suas limitações funcionais, sendo a deficiência interpretada como um problema inerente ao corpo, classificado a partir de critérios clínicos. Esse modelo exerceu forte influência sobre políticas públicas e práticas sociais durante longo período. No entanto, ao centrar-se exclusivamente nas limitações individuais, desconsidera-se o papel do ambiente e das estruturas sociais na geração da exclusão. A partir da segunda metade do século XX, impulsionada pelo fortalecimento dos movimentos sociais das pessoas com deficiência e pela consolidação dos direitos humanos, surge o modelo social da deficiência. Nesse paradigma, a deficiência deixa de ser compreendida exclusivamente como uma condição individual e passa a ser entendida como resultado da interação entre o indivíduo e as barreiras presentes na sociedade. Dessa forma, a exclusão não decorre apenas de impedimentos físicos, sensoriais ou cognitivos, mas principalmente de fatores externos, como barreiras arquitetônicas, comunicacionais, tecnológicas e atitudinais. Essa mudança de perspectiva desloca o foco da adaptação do indivíduo para a transformação do ambiente, enfatizando a responsabilidade coletiva na promoção da inclusão (Bartalotti, 2006). Na legislação brasileira, essa concepção manifesta-se de forma explícita no Decreto nº 5.296/2004 (Brasil, 2004), que regulamenta as Leis nº 10.048/2000 (Brasil, 2000a) e nº 10.098/2000 (Brasil, 2000b). O inciso 1 do art. 5º do referido decreto define a pessoa com deficiência como:

> […]a que possui limitação ou incapacidade para o desempenho de atividade e se enquadra nas seguintes categorias:
>
> a) deficiência física: alteração completa ou parcial de um ou mais segmentos do corpo humano, acarretando o comprometimento da função física, apresentando-se sob a forma de paraplegia, paraparesia, monoplegia, monoparesia, tetraplegia, tetraparesia, triplegia, triparesia, hemiplegia, hemiparesia, ostomia, amputação ou ausência de membro, paralisia cerebral, nanismo, membros com deformidade congênita ou adquirida, exceto as deformidades estéticas e as que não produzam dificuldades para o desempenho de funções;

b) deficiência auditiva: perda bilateral, parcial ou total, de quarenta e um decibéis (dB) ou mais, aferida por audiograma nas freqüências de 500Hz, 1.000Hz, 2.000Hz e 3.000Hz;

c) deficiência visual: cegueira, na qual a acuidade visual é igual ou menor que 0,05 no melhor olho, com a melhor correção óptica; a baixa visão, que significa acuidade visual entre 0,3 e 0,05 no melhor olho, com a melhor correção óptica; os casos nos quais a somatória da medida do campo visual em ambos os olhos for igual ou menor que 60o; ou a ocorrência simultânea de quaisquer das condições anteriores;

d) deficiência mental: funcionamento intelectual significativamente inferior à média, com manifestação antes dos dezoito anos e limitações associadas a duas ou mais áreas de habilidades adaptativas, tais como:

1. comunicação;

2. cuidado pessoal;

3. habilidades sociais;

4. utilização dos recursos da comunidade;

5. saúde e segurança;

6. habilidades acadêmicas;

7. lazer; e

8. trabalho;

e) deficiência múltipla-associação de duas ou mais deficiências [Brasil 2004].

No inciso Segundo do artigo quinto do Decreto 5.296/2004 (Brasil, 2004), define-se a pessoa com mobilidade reduzida como:

> aquela que, não se enquadrando no conceito de pessoa portadora de deficiência, tenha, por qualquer motivo, dificuldade de movimentar-se, permanente ou temporariamente, gerando redução efetiva da mobilidade, flexibilidade, coordenação motora e percepção [Brasil 2004].

Essas definições evidenciam uma abordagem centrada na fisiologia humana, característica do modelo médico da deficiência. Ainda assim, o Decreto nº 5.296/2004 (Brasil, 2004) representa um avanço na garantia de direitos das pessoas com deficiência. Cabe destacar que a terminologia utilizada no Decreto nº 5.296/2004 (Brasil, 2004) reflete o contexto histórico de sua elaboração. Nesse sentido, observa-se o uso da expressão “deficiência mental”. Na contemporaneidade, essa terminologia vem sendo substituída por expressões mais alinhadas ao modelo social da deficiência, como “deficiência intelectual”.

De forma semelhante, a expressão “pessoa portadora de deficiência”, também presente em legislações anteriores, vem sendo progressivamente substituída pelo termo “pessoa com deficiência”. Essa mudança não se limita a uma questão terminológica, mas reflete uma transformação na forma de compreender a deficiência, que deixa de ser vista como algo que a pessoa “porta” e passa a ser entendida como resultado da interação entre suas limitações e barreiras sociais.

Dessa forma, neste capítulo, adota-se a terminologia contemporânea, em consonância com a Convenção sobre os Direitos das Pessoas com Deficiência e com a Lei Brasileira de Inclusão da Pessoa com Deficiência, de acordo com a Lei nº 13.146/2015 (Brasil, 2015), que reafirma uma abordagem baseada em direitos humanos, inclusão social e respeito à diversidade.

No que diz respeito ao contexto da acessibilidade digital, o Art. 47 coloca:

> No prazo de até doze meses a contar da data de publicação deste Decreto, será obrigatória a acessibilidade nos portais e sítios eletrônicos da administração pública na rede mundial de computadores (internet), para o uso das pessoas portadoras de deficiência visual, garantindo-lhes o pleno acesso às informações disponíveis.
>
> §1o Nos portais e sítios de grande porte, desde que seja demonstrada a inviabilidade técnica de se concluir os procedimentos para alcançar integralmente a acessibilidade, o prazo definido no caput será estendido por igual período.
>
> §2o Os sítios eletrônicos acessíveis às pessoas portadoras de deficiência conterão símbolo que represente a acessibilidade na rede mundial de computadores (internet), a ser adotado nas respectivas páginas de entrada (Brasil, 2004).

A transição paradigmática intensifica-se a partir da segunda metade do século XX, impulsionada pela Declaração Universal dos Direitos Humanos (1948) e pelo fortalecimento dos movimentos sociais das pessoas com deficiência. Nesse modelo, a exclusão resulta da interação entre pessoas com deficiência e as barreiras impostas pela sociedade.

Esse entendimento é incorporado à legislação brasileira por meio da Convenção sobre os Direitos das Pessoas com Deficiência, internalizada com status constitucional pelo Decreto nº 6.949/2009. Nesse sentido, destaca-se que o Decreto 6.949/2009 promulga a Convenção Internacional sobre os Direitos das Pessoas com Deficiência e seu Protocolo Facultativo, assinados em Nova York, em 30 de março de 2007. Assim, a Convenção tem o seguinte propósito, presente no Art. 1 do Decreto 6949/2009 (Brasil, 2009):

> […]promover, proteger e assegurar o exercício pleno e eqüitativo de todos os direitos humanos e liberdades fundamentais por todas as pessoas com deficiência e promover o respeito pela sua dignidade inerente (Brasil 2009).

Nesse artigo, também estabelece-se a seguinte definição referente à deficiência:

> Pessoas com deficiência são aquelas que têm impedimentos de longo prazo de natureza física, mental, intelectual ou sensorial, os quais, em interação com diversas barreiras, podem obstruir sua participação plena e efetiva na sociedade em igualdades de condições com as demais pessoas (Brasil 2009).

A consolidação desse paradigma ocorre com a promulgação da Lei nº 13.146 (Brasil,2015 que atualiza o conceito de pessoa com deficiência. Em seu Art. 1 destaca-se que a respectiva legislação visa "[…]assegurar e a promover, em condições de igualdade, o exercício dos direitos e das liberdades fundamentais por pessoa com deficiência, visando à sua inclusão social e cidadania" (Brasil, 2015).

Também nesta legislação, o Artigo Segundo define a pessoa com deficiência como ":

aquela que tem impedimento de longo prazo de natureza física, mental, intelectual ou sensorial, o qual, em interação com uma ou mais barreiras, pode obstruir sua participação plena e efetiva na sociedade em igualdade de condições com as demais pessoas".

No que diz respeito à avaliação da deficiência, o Decreto nº 11.063, de 2022 (Brasil, 2022) destaca no primeiro parágrafo que "a avaliação da deficiência, quando necessária, será biopsicossocial, realizada por equipe multiprofissional e interdisciplinar e considerará: (Vigência) (Vide Decreto nº 11.063, de 2022)

I - os impedimentos nas funções e nas estruturas do corpo;

II - os fatores socioambientais, psicológicos e pessoais;

III - a limitação no desempenho de atividades; e

IV - a restrição de participação.) [Brasil 2015].

A LBI explicita a centralidade das barreiras no processo de exclusão social, definindo-as no inciso quarto do artigo terceiro como:

qualquer entrave, obstáculo, atitude ou comportamento que limite ou impeça a participação social da pessoa, bem como o gozo, a fruição e o exercício de seus direitos à acessibilidade, à liberdade de movimento e de expressão, à comunicação, ao acesso à informação, à compreensão, à circulação com segurança, entre outros, classificadas em:

a) barreiras urbanísticas: as existentes nas vias e nos espaços públicos e privados abertos ao público ou de uso coletivo;

b) barreiras arquitetônicas: as existentes nos edifícios públicos e privados;

c) barreiras nos transportes: as existentes nos sistemas e meios de transportes;

d) barreiras nas comunicações e na informação: qualquer entrave, obstáculo, atitude ou comportamento que dificulte ou impossibilite a expressão ou o recebimento de mensagens e de informações por intermédio de sistemas de comunicação e de tecnologia da informação;

e) barreiras atitudinais: atitudes ou comportamentos que impeçam ou prejudiquem a participação social da pessoa com deficiência em igualdade de condições e oportunidades com as demais pessoas;

f) barreiras tecnológicas: as que dificultam ou impedem o acesso da pessoa com deficiência às tecnologias (Brasil 2015).

Barreiras à acessibilidade e iniciativas de enfrentamento em diferentes países

No contexto da web, destacam-se especialmente as barreiras atitudinais, tecnológicas e comunicacionais. Nesse cenário, a busca por uma comunicação digital acessível impulsionou o desenvolvimento de iniciativas voltadas à acessibilidade em diferentes países. Canadá, Estados Unidos, Austrália e Portugal estiveram entre os primeiros a estabelecer parâmetros de acessibilidade, buscando adequar os ambientes digitais às exigências legais e ampliar o acesso à informação (Moura, 2009). Nesse contexto, destaca-se a criação da W*eb Accessibility Initiative* (WAI), em 1997, pelo *World Wide Web Consortium* (W3C), com o objetivo de

promover recomendações e recursos voltados à acessibilidade digital. Essa iniciativa reúne representantes da indústria, governos, instituições de apoio às pessoas com deficiência e organizações de pesquisa, atuando no desenvolvimento de estratégias e diretrizes relacionadas à acessibilidade na Web (W3C Brasil, 2015). Em 1999, foi publicada a primeira versão das Web Content Accessibility Guidelines (WCAG 1.0), voltada à orientação da acessibilidade em conteúdos web. Desde então, as diretrizes passaram por revisões e atualmente encontram-se na versão 2.2. As WCAG são utilizadas por diferentes públicos, incluindo desenvolvedores, designers, educadores, entre outros (W3C Brasil, 2015). Para atender a esses diferentes perfis, as diretrizes organizam-se em princípios, diretrizes gerais, critérios de sucesso verificáveis e técnicas de implementação (W3C, 2024). As WCAG estruturam-se em quatro princípios fundamentais da acessibilidade web, conhecidos pelo acrônimo POUR (Perceptible, Operable, Understandable e Robust) conforme Figura 1.

Figura 1 – Os quatro pilares da acessibilidade WEB

Fonte: próprio autor.

Conforme descrito na Figura 1, são quatro os pilares da acessibilidade WEB:

• Perceptível — as informações e os elementos da interface devem ser disponibilizados de maneira que possam ser identificados e percebidos pelos usuários;

• Operável — os recursos de interação e navegação devem possibilitar utilização por diferentes formas de acesso e interação;

• Compreensível — o conteúdo e os mecanismos de funcionamento da interface devem apresentar clareza e previsibilidade para os usuários;

• Robusto — o conteúdo deve ser compatível com diferentes navegadores, dispositivos e tecnologias assistivas, favorecendo sua interpretação e utilização em distintos contextos tecnológicos (NIC.br, 2020).

Abaixo dos princípios encontram-se as diretrizes, complementadas por critérios de sucesso verificáveis e técnicas de implementação que auxiliam os desenvolvedores na aplicação prática das recomendações (W3C, 2024). Os critérios de sucesso das WCAG são organizados em três níveis de conformidade:

• Nível A — contempla requisitos básicos relacionados às barreiras de acessibilidade mais críticas, representando o nível mínimo de conformidade;

• Nível AA — corresponde a um nível intermediário de conformidade, abrangendo requisitos que ampliam significativamente a acessibilidade e favorecem o uso do conteúdo pela maioria dos usuários em diferentes contextos;

• Nível AAA — representa o nível mais elevado de conformidade, envolvendo recomendações adicionais voltadas ao aprimoramento da acessibilidade (W3C, 2024).

Além dos critérios de sucesso, as WCAG disponibilizam orientações práticas para auxiliar a implementação da acessibilidade, incluindo técnicas consideradas essenciais e recomendações complementares. Essas recomendações adicionais ampliam as possibilidades de inclusão e favorecem experiências mais acessíveis para diferentes perfis de usuários. Em conjunto, princípios, diretrizes, critérios de sucesso e técnicas compõem uma estrutura integrada voltada à promoção de uma comunicação digital mais inclusiva (W3C, 2024). No Brasil, as diretrizes de acessibilidade digital vêm sendo incorporadas por normas da Associação Brasileira de Normas Técnicas (ABNT), como a ABNT NBR 17225, que estabelece requisitos de acessibilidade para conteúdo e aplicações web, alinhando os aspectos técnicos às disposições da Lei Brasileira de Inclusão e aos compromissos assumidos pelo Brasil no âmbito da Convenção sobre os Direitos das Pessoas com Deficiência (ABNT, 2025). Nesse sentido, o Art. 63 da Lei nº 13.146/2015 estabelece que:

> É obrigatória a acessibilidade nos sítios da internet mantidos por empresas com sede ou representação comercial no País ou por órgãos de governo, para uso da pessoa com deficiência, garantindo-lhe acesso às informações disponíveis, conforme as melhores práticas e diretrizes de acessibilidade adotadas internacionalmente.
> **§ 1º Os sítios devem conter símbolo de acessibilidade em destaque (BRASIL, 2015).**

Dessa forma, a análise cronológica da legislação evidencia que a transição do modelo médico para o modelo social da deficiência constitui um eixo fundamental para a compreensão dos desafios contemporâneos da acessibilidade web. Nesse sentido, considera-se que a adoção de diretrizes nacionais e internacionais, integradas ao design centrado no ser humano e à UX Research envolvendo pessoas com deficiência, constitui estratégia fundamental para o desenvolvimento de sistemas digitais que contemplem a diversidade humana (Horton; Quesenbery, 2014).

3.3 Por que realizar pesquisas de UX (UX Research)?

Muitas organizações enfrentam dificuldades em oferecer produtos e serviços que atendam às necessidades e expectativas de seu público. Em grande parte dos casos, essas dificuldades estão relacionadas ao baixo nível de conhecimento sobre demandas, comportamentos e formas de

interação dos usuários com produtos e serviços. Além disso, quando o feedback obtido ao longo do processo de desenvolvimento é desconsiderado, perde-se uma importante fonte de informações para os processos de melhoria contínua, inovação e aperfeiçoamento da experiência do usuário (Norman, 2013; Garrett, 2011). Por essa razão, faz-se necessária a adoção da UX Research, com o objetivo de compreender necessidades, percepções e experiências que possam contribuir para o desenvolvimento de soluções mais adequadas ao contexto de uso. A desconsideração desses aspectos pode resultar em insatisfação e comprometer oportunidades de inovação e melhoria contínua.

Em contextos de webmídia, essa lacuna de compreensão tende a aprofundar barreiras de acesso e uso, especialmente para públicos diversos, reforçando práticas excludentes e comprometendo a construção de experiências digitais verdadeiramente inclusivas (W3C BRASIL, 2015).

A pesquisa em Experiência do Usuário (UX Research) surge como uma abordagem sistemática voltada à compreensão do comportamento, das motivações, das necessidades e das limitações dos usuários durante a interação com produtos e serviços digitais. Sem uma investigação estruturada desses aspectos, as organizações correm o risco de investir recursos financeiros, tempo e esforço no desenvolvimento de soluções que não atendem às demandas reais do público, comprometendo sua adoção e uso.

A oferta de experiências significativas depende do desenvolvimento de sistemas digitais que sejam eficazes, úteis, seguros, agradáveis, fáceis de aprender e utilizar (PREECE; ROGERS; SHARP, 2002). Além disso, a UX Research possibilita aprofundar a compreensão sobre as necessidades, expectativas e experiências dos usuários, envolvendo-os diretamente no processo de desenvolvimento e colocando-os no centro das decisões de projeto, em consonância com os princípios do Design Centrado no Ser Humano (ISO 9241-210:2019).

Quando aplicada ao domínio da webmídia, a UX Research assume papel central na identificação de obstáculos relacionados à acessibilidade, à compreensão da informação e à diversidade de contextos de uso. Ao considerar diferentes perfis de usuários, capacidades e formas de interação, essa prática possibilita o desenvolvimento de soluções mais inclusivas e alinhadas às reais condições de uso. Além disso, ao adotar métodos adequados de pesquisa em UX, as organizações deixam de presumir o que os usuários desejam e passam a compreender como eles pensam, sentem e se comportam em contextos reais de interação. Consequentemente, essa abordagem contribui para a redução de custos, ao possibilitar a identificação precoce de problemas e a realização de ajustes ainda nas etapas iniciais do desenvolvimento, fortalecendo o relacionamento com os usuários e favorecendo a tomada de decisões fundamentadas em evidências empíricas, e não apenas em suposições (ROHRER, 2014) (ISO, 2019).

No âmbito da webmídia inclusiva, essas evidências orientam decisões de projeto que consideram a diversidade de usuários, contribuindo para experiências mais equitativas e acessíveis. A UX Research também se relaciona diretamente com a geração de valor para o usuário e com o fortalecimento da relação entre usuários e sistemas digitais. Ao compreender de forma mais aprofundada as necessidades, expectativas e experiências dos usuários, as organizações conseguem desenvolver produtos e serviços mais alinhados às demandas reais do público, aumentando a satisfação e a percepção de qualidade. Essa aproximação contribui para a construção de relações mais duradouras entre usuários e serviços digitais, favorecendo a confiança e o engajamento contínuo. Além disso, experiências positivas, acessíveis e centradas no usuário podem influenciar a percepção de valor associada a uma marca, bem como a preferência dos usuários em relação a produtos e serviços concorrentes. Em um cenário de crescente competitividade no mercado digital, a capacidade de oferecer experiências inclusivas configura-se como um diferencial estratégico relevante (NORMAN, 2013).

As motivações para a realização de pesquisas em UX são diversas e incluem o desenvolvimento de produtos, serviços e experiências mais eficazes, a melhoria de soluções existentes e a adaptação às mudanças no comportamento e nas expectativas dos usuários ao longo do tempo. Tanto em organizações públicas quanto privadas, observa-se um crescimento significativo da demanda por pesquisas voltadas à experiência do usuário, impulsionado pela transformação digital e pela necessidade de oferecer serviços mais eficientes, acessíveis e centrados nas pessoas (STICKDORN et al., 2018). Nesse cenário, a UX Research destaca-se como um elemento estratégico para o desenvolvimento de webmídias inclusivas. Embora a pesquisa em UX seja tradicionalmente associada ao papel do pesquisador especializado, na prática ela pode ser conduzida por diferentes perfis profissionais, como designers, desenvolvedores, entre outros. Em muitos casos, esses profissionais incorporam princípios e métodos de UX Research em suas atividades, buscando compreender e avaliar as interações entre pessoas, produtos e serviços digitais. Dessa forma, o aprendizado e a aplicação de métodos de UX Research beneficiam diferentes áreas envolvidas no desenvolvimento de experiências digitais centradas no usuário (GOODMAN; KUNIAVSKY; MOED, 2012). A ampliação desse repertório metodológico mostra-se particularmente relevante para equipes envolvidas na concepção, no desenvolvimento e na manutenção de webmídias.

Sob essa perspectiva, a incorporação sistemática de métodos de UX Research torna-se essencial para a construção de webmídias mais inclusivas, especialmente ao considerar a diversidade de usuários e a necessidade de compreender diferentes formas de interação, navegação e acesso à informação. Essa discussão torna-se ainda mais relevante em pesquisas envolvendo pessoas com deficiência, temática que será aprofundada na seção seguinte.

3.4 UX Research Envolvendo Pessoas com Deficiência

A partir da compreensão do papel estratégico da UX Research, sua aplicação em contextos envolvendo pessoas com deficiência demanda cuidados metodológicos específicos A UX Research pode ser realizada em diferentes etapas do ciclo de desenvolvimento de produtos e serviços digitais. Em fases iniciais, essa abordagem é fundamental para o levantamento de requisitos, permitindo a identificação do perfil dos usuários, de suas necessidades, expectativas e dificuldades. Nesse estágio, torna-se possível compreender as tarefas que os usuários desejam realizar — sejam elas funcionais, sociais ou emocionais —, bem como as dificuldades e desafios envolvidos nesse processo (OSTERWALDER et al., 2015). Em um segundo momento, a pesquisa assume um caráter avaliativo, voltado à validação de propostas de solução, produtos ou protótipos, verificando se as tarefas são realizadas de forma eficaz e se as soluções agregam valor ao usuário. Esse processo é iterativo, podendo gerar ajustes ao longo do projeto (PREECE; ROGERS; SHARP, 2002). Nesse contexto, diferentes tipos de pesquisa podem ser empregados, como pesquisas exploratórias, avaliativas e comportamentais. Além disso, abordagens quantitativas e qualitativas se complementam, permitindo tanto a análise estatística quanto a compreensão aprofundada das experiências de uso (CRESWELL; POTH, 2018). Quando se trata de pesquisas envolvendo pessoas com deficiência, a adoção de princípios do Design Centrado no Ser Humano torna-se ainda mais relevante. Envolver diretamente essas pessoas desde as fases iniciais do processo de design é fundamental para garantir acessibilidade, reconhecer a diversidade de necessidades e evitar suposições capacitistas (LAZAR; GOLDSTEIN; TAYLOR, 2015). Pessoas com diferentes tipos de deficiência — sensoriais, motoras e cognitivas — devem ser consideradas, promovendo experiências digitais mais inclusivas e equitativas. O recrutamento de participantes exige critérios bem definidos, considerando o uso de tecnologias assistivas, a experiência dos participantes e suas formas de comunicação. Esse recrutamento pode ocorrer por meio de instituições ou pela técnica de

amostragem em bola de neve, sendo influenciado pelo prazo disponível e pelas possibilidades de acesso aos participantes (TAVARES, 2019). A adaptação dos instrumentos de coleta de dados e do Termo de Consentimento Livre e Esclarecido (TCLE) é essencial para garantir acessibilidade, incluindo compatibilidade com tecnologias assistivas, linguagem simples e disponibilização em múltiplos formatos.

3.4.1 Tecnologias assistivas no contexto da UX Research

A utilização de tecnologias assistivas constitui um elemento fundamental na interação de pessoas com deficiência com ambientes digitais, influenciando diretamente a experiência do usuário e os resultados de pesquisas em UX. Essas tecnologias possibilitam a mediação da interação humano-computador, promovendo maior autonomia, acessibilidade e inclusão em diferentes contextos de uso. Segundo o Comitê de Ajudas Técnicas (CAT), a tecnologia assistiva é definida como:

> uma área do conhecimento, de característica interdisciplinar, que engloba produtos, recursos, metodologias, estratégias, práticas e serviços que objetivam promover a funcionalidade, relacionada à atividade e participação, de pessoas com deficiência, incapacidades ou mobilidade reduzida, visando sua autonomia, independência, qualidade de vida e inclusão social (Bersch, 2017).

Nesse sentido, destaca-se os seguintes exemplos de tecnologia assistiva:

- NVDA (NonVisual Desktop Access): Leitor de tela gratuito e de código aberto utilizado no sistema Windows, que permite que pessoas com deficiência visual utilizem o computador por meio de áudio. O NVDA lê em voz alta os elementos exibidos na tela e possibilita a navegação pelo teclado, facilitando o acesso a conteúdos, aplicações e páginas da web sem o uso da visão;

- Dosvox: Sistema desenvolvido no Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais (NCE/UFRJ), voltado para pessoas com deficiência visual. O Dosvox constitui-se como um ambiente computacional com interação realizada por meio de síntese de voz, permitindo que os usuários possam desempenhar atividades como estudar, trabalhar, navegar na internet, entre outras. Diferentemente dos leitores de tela, que interpretam e reproduzem por voz as informações exibidas na interface gráfica do sistema operacional, o Dosvox oferece um ambiente próprio de interação, baseado em abreviaturas de comandos e menus sonoros acionados por meio do teclado, favorecendo a autonomia e a inclusão digital;

- Braille Fácil: sistema desenvolvido para a digitação, tradução, produção e impressão de textos em Braille, seguindo as regras da língua portuguesa. É o programa oficial de impressão braille do Brasil e Portugal;

- Musibraille: programa que permite a digitação e simulação musical de partituras escritas em braile;

- Orca: Leitor de tela gratuito e de código aberto utilizado em sistemas Linux por pessoas com deficiência visual. O Orca lê em voz alta os elementos da tela e possibilita a navegação pelo teclado, facilitando o acesso a aplicações e conteúdos sem o uso da visão;

- Sintetizadores de voz: Recursos de tecnologia assistiva que convertem texto em fala, permitindo que informações digitais sejam apresentadas em áudio. São utilizados por leitores de tela, interfaces adaptativas como o sistema dosvox, sistemas de comunicação alternativa como o prancha fácil e outras aplicações acessíveis. Para pessoas com deficiência visual, deficiência motora ou dificuldades de leitura, os sintetizadores de voz contribuem para o acesso à informação, à navegação e à interação com sistemas digitais;

- Speech Dispatcher: Sistema utilizado em ambientes Linux que gerencia a saída de voz das aplicações, permitindo a comunicação entre programas e sintetizadores de voz. Ele organiza e distribui as mensagens de áudio, facilitando o uso de leitores de tela e outros recursos acessíveis;

- VoiceOver: Leitor de tela nativo dos sistemas iOS e macOS, utilizado por pessoas com deficiência visual para interação com o dispositivo por meio de áudio. O VoiceOver lê em voz alta os elementos da tela e permite a navegação por gestos ou teclado, facilitando o acesso a aplicações e conteúdos sem o uso da visão;

- TalkBack: Leitor de tela nativo do sistema Android, utilizado por pessoas com deficiência visual para interação com o dispositivo por meio de áudio. O TalkBack lê em voz alta os elementos da tela e permite a navegação por gestos, facilitando o acesso a aplicações e conteúdos sem o uso da visão;

- Audiodescrição: recurso de acessibilidade comunicacional que traduz imagens em palavras, possibilitando o acesso a conteúdos imagéticos presentes em diferentes mídias e contextos. Esse recurso pode ser aplicado em filmes, teatros, redes sociais, entre outros. A audiodescrição é utilizada geralmente por indivíduos com deficiência visual, podendo também ser adotada por pessoas com deficiência intelectual e baixo letramento, favorecendo a compreensão, a autonomia e o acesso à informação;

- Motrix: sistema de controle do mouse e da digitação, acionado por comandos de voz. É adequado para pessoas com tetraplegia cuja fala esteja preservada;

- Switch Control (Controle por Interruptores): Recurso presente em sistemas computacionais que permite a interação por meio de dispositivos de entrada alternativos, como botões ou switches adaptados. É utilizado por pessoas com limitações motoras, possibilitando a navegação e o controle do sistema de forma acessível;

- Teclado Virtual e Teclado Adaptado: Interfaces que permitem a digitação por meio de dispositivos alternativos, como mouse, toque ou dispositivos assistivos. São utilizados por pessoas com deficiência motora ou mobilidade reduzida, facilitando a entrada de texto em sistemas digitais;

- Microfênix: trata-se de uma tecnologia desenvolvida para facilitar o uso do computador por pessoas com deficiência física grave que não conseguem utilizar as mãos nem produzir fala, como distrofia muscular, esclerose lateral amiotrófica (ELA) e outras patologias que acarretam limitações motoras severas. O sistema emprega um menu de varredura automática, no qual as opções são realçadas sequencialmente e a seleção da opção desejada é realizada mediante a detecção de ruídos, sinais de sensores ou o movimento de piscar os olhos;

- Prancha Fácil: Ferramenta digital utilizada por pessoas com deficiência na comunicação, especialmente em contextos de Comunicação Aumentativa e Alternativa (CAA). O sistema permite a criação de pranchas digitais com símbolos, imagens e textos, auxiliando a expressão de ideias e necessidades de forma acessível, com foco em pessoas com paralisia cerebral, autismo não verbal, distúrbios de fala e deficiência motora grave;

- Hand Talk: Aplicação que utiliza inteligência artificial para traduzir conteúdos digitais para a Língua Brasileira de Sinais (Libras), sendo voltada principalmente para pessoas com deficiência auditiva. A ferramenta contribui para o acesso à informação em ambientes digitais por meio de um avatar virtual que realiza a interpretação em sinais;

- VLibras: Conjunto de ferramentas que traduz conteúdos digitais (texto, áudio e vídeo) para Libras, promovendo acessibilidade para pessoas surdas em ambientes web e aplicações digitais. Pode ser integrado a sites e sistemas, ampliando o acesso à informação;

- Eye Tracking (Rastreamento Ocular): Tecnologia assistiva voltada principalmente para pessoas com deficiência motora severa, que permite a interação com sistemas computacionais por meio do movimento dos olhos. Essa tecnologia possibilita a navegação, seleção de elementos e comunicação sem o uso das mãos, sendo amplamente utilizada em contextos de acessibilidade e pesquisa em UX;

- Closed Caption (Legendas Ocultas): Recurso de acessibilidade utilizado por pessoas com deficiência auditiva, que apresenta transcrição textual de conteúdos audiovisuais, incluindo falas, efeitos sonoros e identificação de sons relevantes. Contribui para a compreensão de conteúdos multimídia em diferentes contextos;

- Simplificação de Interface (Leitura Fácil): Estratégia e conjunto de ferramentas voltadas para pessoas com deficiência intelectual, que envolvem a adaptação de conteúdos com linguagem simples, uso de ícones e organização visual clara. Essas soluções facilitam a compreensão e a navegação em ambientes digitais;

- Aplicativos de Apoio à Rotina e Organização (ex.: Time Timer, Choiceworks): Ferramentas utilizadas por pessoas com Transtorno do Espectro Autista (TEA) que auxiliam na organização de tarefas, gestão do tempo e previsibilidade de atividades. Esses aplicativos utilizam recursos visuais, temporizadores e sequenciamento de ações, contribuindo para a redução da ansiedade e para uma melhor compreensão das interações em ambientes digitais.

Assim, a diversidade de tecnologias assistivas evidencia que a acessibilidade não se restringe apenas à percepção visual ou auditiva. No contexto da UX Research, a consideração dessas tecnologias é fundamental para a condução de estudos inclusivos, garantindo que diferentes perfis de usuários sejam contemplados tanto no design quanto na avaliação de sistemas.

3.4.2. Desafios práticos

A realização de estudos quantitativos com perfis específicos pode ser limitada pela dificuldade de recrutamento dos participantes. Além disso, quanto mais específico for o recorte da pesquisa, maiores tendem a ser o tempo e os custos necessários para sua condução. A burocracia institucional também pode impactar os processos de coleta de dados e recrutamento. Nesse contexto, a técnica de amostragem em bola de neve pode representar uma alternativa viável, embora dependa diretamente das redes de contato dos participantes. Diante desse cenário, a participação de pessoas com deficiência como pesquisadoras torna-se estratégica, alinhando-se ao lema “nada sobre nós sem nós” (Tavares, 2019). O quadro comparativo apresentado a seguir evidencia diferenças importantes entre a realização de pesquisas qualitativas e quantitativas em termos de objetivo, número de participantes, prazo, recrutamento e adequação às especificidades das pessoas com deficiência. Enquanto abordagens qualitativas permitem maior profundidade e flexibilidade, abordagens quantitativas exigem maior escala e enfrentam desafios significativos de recrutamento, o que pode impactar diretamente a viabilidade dos estudos.

Quadro 1 – Comparação entre abordagens qualitativa e quantitativa em UX Research

Critério / Dimensão	**Pesquisa Qualitativa**	**Pesquisa Quantitativa**
Objetivo	Compreender experiências, percepções e necessidades dos participantes	Generalizar resultados e identificar padrões estatísticos
Tipo de pergunta	"Como?", "Por quê?"	"Quanto?", "Com que frequência?"
Número de participantes	Pequeno	Grande
Profundidade dos dados	Elevada	Moderada
Generalização dos resultados	Limitada	Elevada (quando a amostra é representativa)
Prazo de execução	Geralmente mais curto	Geralmente mais longo

Critério / Dimensão	Pesquisa Qualitativa	Pesquisa Quantitativa
Custo da pesquisa	Menor	Maior
Complexidade do recrutamento	Moderada	Alta, especialmente para perfis específicos
Dependência de amostras numerosas	Baixa	Elevada
Adequação a perfis raros ou específicos	Alta	Limitada
Flexibilidade metodológica	Alta	Menor
Necessidade de adaptações durante a pesquisa	Facilmente incorporadas	Podem comprometer a padronização do estudo
Uso de tecnologias assistivas	Facilmente acomodado durante a coleta	Exige maior padronização e controle experimental
Adequação à diversidade das pessoas com deficiência	Elevada	Dependentes da representatividade da amostra
Impacto da burocracia institucional	Menor	Maior, devido ao tempo necessário para recrutamento e coleta
Aplicação da amostragem em bola de neve	Alta aplicabilidade	Uso limitado, pois não produz amostras probabilísticas
Participação de pesquisadores com deficiência	Fortemente favorecida, enriquecendo a compreensão do contexto	Relevante, porém com menor influência sobre o desenho metodológico
Principais métodos utilizados	Entrevistas, grupos focais, observação, think-aloud, estudos de campo	Questionários, escalas, experimentos controlados, surveys

Critério / Dimensão	Pesquisa Qualitativa	Pesquisa Quantitativa
Principal vantagem	Profundidade e compreensão contextual das experiências	Produção de evidências generalizáveis
Principal limitação	Menor possibilidade de generalização	Dificuldade de recrutar amostras suficientes e representativas
Adequação para UX Research inclusiva	Muito alta	Moderada, dependendo da disponibilidade de participantes
Fonte: elaborado pelos autores.		

Descrição do Quadro

O quadro apresenta uma comparação entre abordagens qualitativas e quantitativas no contexto da UX Research envolvendo pessoas com deficiência. A pesquisa qualitativa destaca-se pela flexibilidade e pela capacidade de compreender experiências em profundidade, sendo particularmente adequada para contextos com amostras reduzidas e perfis específicos.

Por outro lado, a pesquisa quantitativa permite generalização dos resultados, mas exige maior número de participantes, o que pode representar um desafio significativo em estudos com pessoas com deficiência. A dificuldade de recrutamento, aliada a questões institucionais e logísticas, pode impactar diretamente a viabilidade dessa abordagem.

Além disso, observa-se que a técnica de amostragem em bola de neve é mais adequada para estudos qualitativos, enquanto sua aplicação em pesquisas quantitativas é limitada. Dessa forma, a escolha metodológica deve considerar não apenas os objetivos da pesquisa, mas também aspectos práticos como prazo, acesso aos participantes e viabilidade da coleta de dados.

Por fim, no contexto da Webmedia — caracterizada pela integração de diferentes mídias, como texto, áudio, vídeo, imagens e interatividade — a UX Research desempenha um papel estratégico na promoção de uma internet mais inclusiva e democrática. Garantir que pessoas com deficiência possam acessar, compreender e interagir com conteúdos digitais não é apenas um desafio técnico, mas também ético e social, sendo fundamental para a construção de experiências digitais mais justas, acessíveis e socialmente responsáveis [Henry et al. 2014].

3.4.3 Adaptação metodológica em UX Research inclusiva

A condução de pesquisas em UX envolvendo pessoas com deficiência exige a adoção de estratégias metodológicas que considerem a diversidade de perfis, capacidades e formas de interação dos participantes. Nesse contexto, a adaptação metodológica não deve ser

compreendida apenas como um ajuste técnico dos procedimentos de pesquisa, mas como parte de uma abordagem participativa voltada à construção colaborativa do conhecimento e à ampliação da participação significativa dos usuários ao longo de todo o processo de design.

Assim, a acessibilidade na pesquisa não se limita à remoção de barreiras operacionais, envolvendo também a promoção de processos mais democráticos, inclusivos e centrados nas experiências dos participantes. No campo da Interação Humano-Computador, abordagens como o Design Participativo (Participatory Design – PD) têm se destacado por promover o envolvimento ativo dos usuários no processo de design, não apenas como participantes, mas também como coparticipantes na construção das soluções desenvolvidas. Essa abordagem busca incorporar as experiências, necessidades e valores dos usuários ao longo de todo o ciclo de desenvolvimento, contribuindo para a criação de sistemas mais alinhados às suas realidades de uso (SCHULER; NAMIOKA, 1993).

No contexto das pessoas com deficiência, o Design Participativo torna-se ainda mais relevante ao possibilitar a redução de diferenças entre desenvolvedores e usuários, favorecendo processos mais inclusivos, colaborativos e democráticos. Estudos recentes indicam que a participação ativa de pessoas com deficiência nas etapas de pesquisa, design e avaliação pode contribuir para a identificação de barreiras e para o desenvolvimento de soluções mais adequadas às necessidades dos usuários (WACNIK; DALY; VERMA, 2023). Além disso, o Design Participativo fundamenta-se em princípios como colaboração, aprendizagem mútua e compartilhamento de poder no processo de tomada de decisão, permitindo uma compreensão mais profunda das práticas, experiências e contextos de uso dos participantes (SCHULER; NAMIOKA, 1993).

Nesse sentido, a participação dos usuários não se limita à validação das soluções propostas, mas envolve sua atuação ativa na definição de problemas, na ideação e na avaliação das alternativas desenvolvidas. No planejamento dos estudos, é fundamental considerar a adequação dos métodos às características dos participantes. Entrevistas, por exemplo, podem demandar adaptações na linguagem utilizada, priorizando construções mais simples, diretas e acessíveis, especialmente em contextos envolvendo pessoas com deficiência cognitiva. Além disso, o tempo de duração das sessões pode precisar ser ampliado, respeitando o ritmo dos participantes e garantindo condições adequadas de compreensão e resposta. Nos testes de usabilidade, a presença de tecnologias assistivas requer atenção especial à configuração dos ambientes de teste. Ferramentas como leitores de tela, ampliadores de tela, softwares de comunicação alternativa e dispositivos de entrada adaptados podem influenciar diretamente a forma como os usuários interagem com os sistemas. Dessa forma, torna-se necessário garantir que o ambiente experimental esteja devidamente preparado para acomodar essas tecnologias, evitando interferências que possam comprometer a validade dos resultados (LAZAR; GOLDSTEIN; TAYLOR, 2015).

A condução de estudos remotos também apresenta desafios e oportunidades no contexto da acessibilidade. Por um lado, a pesquisa remota pode ampliar o acesso a participantes geograficamente distribuídos e reduzir barreiras logísticas. Por outro, pode exigir maior suporte técnico, adaptação de ferramentas e acompanhamento mais próximo para garantir que os participantes consigam interagir adequadamente com as plataformas utilizadas.

Outro aspecto relevante refere-se à coleta de dados. Métodos tradicionais, como observação e think-aloud, podem necessitar de adaptações, especialmente quando utilizados com pessoas que fazem uso de tecnologias assistivas. Em alguns casos, a verbalização simultânea pode interferir na interação dos participantes com o sistema, exigindo abordagens alternativas, como o registro automatizado das interações seguido de entrevistas realizadas após a conclusão das tarefas,

permitindo que os participantes relatem suas percepções, dificuldades e experiências de uso sem comprometer a interação com o sistema durante a execução das atividades. Adicionalmente, a análise dos dados deve considerar as especificidades dos participantes, evitando interpretações baseadas exclusivamente em padrões normativos de interação. A literatura aponta que a participação dos usuários nos processos de decisão e avaliação é central para a compreensão das práticas de uso e para o desenvolvimento de soluções mais inclusivas e alinhadas às necessidades dos usuários (ROBERTSON; SIMONSEN, 2012).

Por fim, destaca-se que a adaptação metodológica em UX Research inclusiva está diretamente relacionada ao compromisso ético com a participação significativa das pessoas com deficiência. Isso implica não apenas garantir acessibilidade, mas também reconhecer esses participantes como agentes ativos no processo de construção do conhecimento e no desenvolvimento das soluções propostas, contribuindo para a criação de tecnologias mais inclusivas, eficazes e alinhadas às reais necessidades dos usuários.

3.5 Webmedia inclusiva: desafios de UX envolvendo pessoas com deficiência

A acessibilidade digital constitui um dos principais desafios contemporâneos no campo da UX, especialmente diante da diversidade de perfis e necessidades dos usuários. Diferentes mídias e contextos de interação demandam estratégias específicas para garantir o acesso equitativo à informação, à comunicação e à interação digital (PREECE; ROGERS; SHARP, 2002). Nesse contexto, a aplicação de padrões internacionais de acessibilidade, como as Diretrizes de Acessibilidade para Conteúdo Web (Web Content Accessibility Guidelines – WCAG 2.2), desempenha papel fundamental na orientação do desenvolvimento de sistemas digitais mais inclusivos(W3C, 2024).

Essas diretrizes contemplam não apenas aspectos relacionados à conformidade técnica, mas também fatores associados à qualidade da interação e da UX. Recursos como legendas para conteúdos audiovisuais, audiodescrição, contraste adequado entre texto e fundo e navegação por teclado constituem exemplos de práticas que contribuem para a redução de barreiras e para a ampliação do acesso aos conteúdos digitais. Entretanto, a acessibilidade em ambientes digitais não deve ser compreendida exclusivamente sob uma perspectiva técnica ou normativa.

Para além do atendimento aos critérios estabelecidos pelas diretrizes, torna-se necessário considerar aspectos relacionados à usabilidade e à UX como um todo. Interfaces consideradas tecnicamente acessíveis podem ainda apresentar dificuldades de uso, elevada carga cognitiva ou baixa eficiência na realização de tarefas, evidenciando a necessidade de integração entre acessibilidade, usabilidade e UX. Assim, um sistema verdadeiramente acessível deve ser perceptível, operável, compreensível e robusto (princípios fundamentais das WCAG) resultando em experiências satisfatórias para diferentes perfis de usuários, independentemente de suas habilidades ou limitações (NORMAN, 2013; ISO 9241-210, 2019).

No contexto brasileiro, essas limitações tornam-se ainda mais evidentes. Há mais de duas décadas, pesquisadores nacionais e internacionais vêm investigando dificuldades relacionadas à acessibilidade digital com o objetivo de compreender os desafios enfrentados por pessoas com deficiência no uso dessas tecnologias, bem como propor soluções voltadas à promoção de experiências mais inclusivas (BYERLEY; CHAMBERS, 2002) (MONTEIRO, 2011 (MITTAL et al., 2023). Esses estudos evidenciam que as dificuldades de interação em ambientes digitais são diversas e não se restringem à ausência de conformidade técnica.

Entre os principais desafios identificados destacam-se conteúdos sem estrutura semântica adequada, ausência de descrições textuais para imagens, falta de legendas e de recursos em

língua de sinais, incompatibilidade com leitores de tela, limitações na navegação por teclado, uso inadequado de componentes interativos, além de interfaces complexas, inconsistentes ou com elevada carga cognitiva (LAZAR; GOLDSTEIN; TAYLOR, 2015) (W3C, 2024) (ISO 9241-210, 2019). Além disso, pesquisas demonstram que a acessibilidade digital continua sendo um desafio para a Computação evidenciando que muitas das limitações identificadas nas últimas décadas permanecem presentes em diferentes contextos digitais (DE FÁTIMA GRANATTO et al., 2016) (DA COSTA NUNES et al., 2023).

Esse cenário reforça a necessidade de integrar acessibilidade, usabilidade e design centrado no ser Humano na concepção de sistemas digitais mais inclusivos (HORTON; QUESENBERY, 2014). Essa perspectiva demanda a adoção de práticas de design centrado no ser humano, nas quais a diversidade seja reconhecida como parte intrínseca do processo de desenvolvimento. Nesse sentido, testes e avaliações com usuários diversos tornam-se fundamentais para identificar problemas que frequentemente não são detectados por ferramentas automatizadas. Métodos como entrevistas, testes de usabilidade, observação e análise de tarefas possibilitam compreender dificuldades relacionadas à navegação, inconsistências de interface e elevada carga cognitiva durante a interação. Assim, a UX Research contribui para compreender como diferentes usuários percebem, interpretam e interagem com os ambientes digitais, permitindo o desenvolvimento de soluções mais alinhadas às necessidades reais dos usuários (LAZAR; GOLDSTEIN; TAYLOR, 2015).

Um dos desafios da UX Research no contexto de uma Webmedia inclusiva pode estar associada à ausência de formação específica, à limitação de conhecimento técnico ou à priorização de prazos e custos em detrimento da inclusão. Pesquisas evidenciam que mitos e o desconhecimento técnico ainda dificultam a incorporação sistemática da acessibilidade nos processos de desenvolvimento digital (BI et al., 2021; PARTHASARATHY et al., 2025). Como consequência, a acessibilidade tende a ser tratada como um aspecto secundário, comprometendo a construção de ambientes digitais mais democráticos e limitando a compreensão das necessidades e experiências desse público. Além disso, observa-se a dificuldade de recrutamento de participantes que atendam a perfis específicos de pesquisa, o que pode comprometer a constituição de amostras alinhadas aos objetivos do estudo.

Essa limitação torna-se particularmente sensível no contexto acadêmico, no qual a credibilidade da pesquisa pode ser questionada em razão de quantitativos reduzidos de participantes. Tal cenário evidencia uma tensão recorrente entre validade interna, associada à profundidade das análises qualitativas, e validade externa, relacionada à possibilidade de generalização dos resultados. Assim, torna-se fundamental adotar estratégias metodológicas alinhados com os objetivos da pesquisa, reconhecendo as especificidades dos estudos envolvendo pessoas com deficiência. Assim, pensar em acessibilidade em Webmedia vai além do simples cumprimento de legislações e diretrizes técnicas. Trata-se do reconhecimento do direito de todas as pessoas ao acesso à informação e à participação plena no ambiente digital.

Esse compromisso possui uma dimensão ética, social e organizacional alinhada aos princípios do design inclusivo e da cidadania digital. Além disso, a acessibilidade configura-se como um fator estratégico de inovação, ampliando o alcance das soluções digitais e contribuindo para o desenvolvimento de experiências mais centradas no ser humano (GARRETT, 2011). Ao integrar acessibilidade, usabilidade e UX de forma estratégica, os ambientes digitais podem consolidar-se como espaços mais democráticos, inclusivos e representativos da diversidade humana.

3.6. Panorama da Acessibilidade Digital no WebMedia – 2020/2025

No âmbito do WebMedia — principal evento brasileiro relacionado à pesquisa em tecnologias web, observa-se uma ampliação gradual das discussões sobre acessibilidade e inclusão. Nesse contexto, esta seção apresenta um panorama dos trabalhos relacionados à acessibilidade digital publicados entre 2020 e 2025, considerando diferentes trilhas e modalidades do evento.

3.6.1 Distribuição dos trabalhos por trilha

A análise da distribuição dos trabalhos permite compreender como a temática da acessibilidade digital vem sendo incorporada pela comunidade de pesquisa do WebMedia.

Os estudos identificados foram publicados nas seguintes trilhas: Principal, WFA (Workshop de Ferramentas e Aplicações), WTVDI (Workshop de TV Digital Interativa), CTD, CTIC e W4E (WebMedia for Everyone). Essa distribuição evidencia o caráter transversal da acessibilidade digital, que dialoga com áreas como interação humano-computador, multimídia, inteligência artificial e tecnologias assistivas.

Nos primeiros anos analisados, os trabalhos estavam concentrados principalmente em workshops e trilhas estendidas, especialmente aqueles relacionados à audiodescrição, assim como orientação e mobilidade. Entretanto, ao longo do período, observa-se uma presença crescente da temática na Trilha Principal, incluindo pesquisas sobre acessibilidade móvel, inteligência artificial aplicada à acessibilidade, reconhecimento de Libras e ambientes imersivos.

Esse cenário sugere uma consolidação gradual da acessibilidade digital como tema relevante dentro da comunidade científica do WebMedia, deixando de ocupar apenas espaços especializados e passando a integrar discussões centrais sobre desenvolvimento tecnológico.

Uma das categorias identificadas refere-se à acessibilidade audiovisual, especialmente envolvendo audiodescrição, descrição automática de conteúdos multimídia e recursos de acessibilidade para ambientes de mídia digital.

Entre os trabalhos encontrados destaca-se o artigo “Accessible Video using Interactive Audio Descriptions” (2020), publicado no WFA, que propõe mecanismos interativos de audiodescrição para ampliar a acessibilidade de vídeos para pessoas com deficiência visual. O estudo explora formas mais dinâmicas de interação com conteúdos audiovisuais acessíveis, buscando ampliar a autonomia dos usuários durante o consumo de mídia digital.

Na edição de 2021, o trabalho “An Approach for Automatic Description of Characters for Blind People”, publicado na Trilha Principal, investigou a descrição automática de personagens em vídeos por meio de técnicas computacionais voltadas à geração de descrições para pessoas cegas ou com baixa visão. O estudo evidencia o potencial da inteligência artificial e do processamento multimídia no apoio à acessibilidade audiovisual.

Já em 2023, o trabalho “Acessibilidade na TV 3.0 Brasileira a partir de mídias de legenda, glosa e áudio descrição”, publicado no WTVDI, ampliou essa discussão ao abordar diferentes recursos de acessibilidade previstos para a TV 3.0 brasileira, incluindo legendas, glosa em Libras e audiodescrição.

De forma geral, os estudos dessa categoria concentram-se no desenvolvimento de soluções técnicas voltadas à ampliação do acesso ao conteúdo multimídia. Entretanto, observa-

se uma presença reduzida de pesquisas voltadas à avaliação longitudinal da experiência de usuários com deficiência durante a utilização dessas tecnologias.

Outra categoria relevante refere-se à acessibilidade web e mobile, contemplando estudos relacionados à avaliação de acessibilidade, aplicações móveis e uso de inteligência artificial em processos automáticos.

Em 2021, o trabalho “Observatório da Acessibilidade da Web Brasileira”, publicado no WFA, apresentou uma iniciativa voltada ao monitoramento e análise da acessibilidade de sites brasileiros. Em 2022, o artigo “I can't pay! Accessibility analysis of mobile banking apps”, publicado na Trilha Principal, analisou a acessibilidade de aplicativos bancários móveis, evidenciando barreiras encontradas por usuários em serviços financeiros digitais.

Em 2024, o estudo “Investigating the accessibility of popular mobile Android apps”, também publicado na Trilha Principal, ampliou essa discussão ao investigar a acessibilidade de aplicativos Android amplamente utilizados.

Já em 2025, surgem pesquisas associando acessibilidade digital e inteligência artificial. O artigo “Image-to-UI Accessibility: Assessing ChatGPT's Effectiveness in Generating Accessible Android Screens from Figma Templates” investigou o uso do ChatGPT na geração de interfaces Android acessíveis. No mesmo ano, o trabalho “Um Estudo Sobre o Uso de Prompts LLM para a Avaliação da Acessibilidade Web dos Portais das Prefeituras da Região dos Sertões dos Crateús” analisou o potencial de modelos de linguagem de grande porte na avaliação automática da acessibilidade web.

Esses estudos demonstram uma tendência crescente de integração entre acessibilidade digital e inteligência artificial, especialmente em processos de automação da avaliação e desenvolvimento de interfaces acessíveis.

Os trabalhos relacionados à acessibilidade para pessoas surdas também apresentaram crescimento ao longo do período analisado.

Em 2024, o artigo “Automatic Time-aware Recognition of Brazilian Sign Language Based on Dynamic Time Warping” propôs mecanismos de reconhecimento automático de Libras utilizando técnicas de visão computacional. No mesmo ano, o trabalho “Uma Investigação sobre Técnicas de Data Augmentation Aplicadas à Tradução Automática Português-LIBRAS” explorou estratégias computacionais voltadas à tradução automática entre Português e Libras.

Em 2025, o estudo “LibrasDetec: um Componente de Detecção de Movimentos para Apoiar Jogos Digitais Acessíveis” apresentou uma solução para reconhecimento de movimentos em Libras aplicada ao contexto dos jogos digitais. Ainda nesse período, o artigo “Automatic 3D Animation Generation for Sign Language” investigou a geração automática de animações tridimensionais para representação de sinais. Complementando essa temática, o estudo “AI-Based Approaches for Brazilian Sign Language Recognition” apresentou uma revisão sobre abordagens baseadas em inteligência artificial aplicadas ao reconhecimento de Libras.

Essas pesquisas evidenciam a forte integração entre acessibilidade, inteligência artificial e visão computacional.

Outra temática identificada refere-se aos ambientes virtuais e à realidade virtual aplicados à acessibilidade.

Em 2022, o trabalho “Uma ferramenta para a customização de ambientes virtuais para práticas de Orientação e Mobilidade”, publicado no WFA, apresentou uma proposta voltada à criação de ambientes virtuais acessíveis para treinamento de orientação e mobilidade de pessoas com deficiência visual.

Em 2023, o estudo “Dashboards to Support Rehabilitation in Orientation and Mobility Activities” abordou o uso de dashboards para apoiar processos de reabilitação relacionados à orientação e mobilidade.

Na edição de 2024, o artigo “Virtual Reality for People with Visual Impairments in the Orientation and Mobility Context”, publicado em trilhas associadas ao CTD e WFA, discutiu o uso da realidade virtual para pessoas com deficiência visual. Já em 2025, o trabalho “Designing a Multimodal Feedback Component for VR Applications” apresentou mecanismos de feedback multimodal aplicados a ambientes de realidade virtual.

Também foram identificadas pesquisas voltadas à inclusão digital de pessoas idosas. Entre elas destacam-se os trabalhos “ARIA e Interactive Access: projetando chatbots para idosos” (CTIC, 2024) e “Práticas com smartphones para idosos: um projeto de extensão do ICMC/USP” (W4E, 2024). Esses estudos demonstram uma preocupação crescente com acessibilidade cognitiva, inclusão digital e participação social de pessoas idosas em ambientes digitais.

De modo geral, os trabalhos publicados entre 2020 e 2025 evidenciam o fortalecimento das discussões sobre acessibilidade digital no WebMedia. A análise realizada mostra uma ampliação progressiva da diversidade temática, bem como uma maior inserção da acessibilidade em diferentes trilhas do evento.

Apesar dos avanços observados, verifica-se que a maior parte das pesquisas ainda privilegia abordagens técnicas voltadas ao desenvolvimento de algoritmos, ferramentas e mecanismos automáticos de acessibilidade. Em contrapartida, estudos centrados na experiência de usuários com deficiência, UX Research, design participativo e participação ativa desses usuários nos processos de desenvolvimento permanecem menos frequentes.

Também são pouco recorrentes pesquisas longitudinais voltadas ao uso dessas tecnologias em contextos reais, considerando aspectos como autonomia, conforto, percepção subjetiva, contexto social e diversidade de perfis de usuários.

Por outro lado, destaca-se o crescimento de estudos envolvendo inteligência artificial, visão computacional e ambientes imersivos aplicados à acessibilidade digital. Esse cenário revela oportunidades promissoras para futuras investigações que integrem avanços tecnológicos e necessidades reais dos usuários.

Por fim, o panorama apresentado evidencia não apenas a evolução das discussões sobre acessibilidade digital no WebMedia, mas também desafios e oportunidades para o desenvolvimento de tecnologias mais inclusivas, alinhadas aos princípios do design centrado no usuário e da inclusão digital.

3.7. Considerações Finais

A acessibilidade digital constitui um elemento fundamental para a construção de uma sociedade mais inclusiva, equitativa e democrática, especialmente no contexto da crescente digitalização das relações sociais. Ao longo deste capítulo, buscou-se discutir a acessibilidade a partir de uma perspectiva que integra aspectos históricos, legais, técnicos e metodológicos, evidenciando sua

complexidade e sua relevância no campo da Webmedia. Nesse sentido, a UX Research destaca-se como um caminho promissor para a construção de webmídias mais inclusivas, ao possibilitar uma compreensão mais profunda das necessidades, experiências e barreiras enfrentadas pelos usuários. Ao integrar métodos qualitativos e quantitativos, bem como ao considerar a diversidade de perfis e contextos de uso, a UX Research contribui para o desenvolvimento de soluções mais alinhadas à realidade dos usuários.

Além disso, a inclusão de pessoas com deficiência no processo de pesquisa e desenvolvimento não deve ser compreendida apenas como uma exigência ética ou normativa, mas como uma oportunidade de inovação e melhoria da qualidade dessas tecnologias. A diversidade de perspectivas enriquece o processo de design e contribui para a criação de experiências mais significativas e acessíveis para todos. Por fim, destaca-se que uma WebMedia plenamente inclusiva requer um esforço contínuo e colaborativo, envolvendo pesquisadores, desenvolvedores, designers, gestores e usuários. Sua consolidação depende não apenas da adoção de diretrizes e abordagens de desenvolvimento acessível, mas da incorporação de uma cultura de inclusão que reconheça a diversidade humana como um princípio indispensável no projeto, desenvolvimento e avaliação dessas tecnologias.

3.8. Declaração sobre o uso de Inteligência Artificial

A ferramenta ChatGPT (OpenAI, modelo GPT-5.5) foi utilizada como apoio à revisão textual, bem como ao levantamento e à organização dos trabalhos apresentados neste capítulo, com base nos anais do WebMedia.

REFERÊNCIAS

ANDRADE, M. G.; RABELO, D. M. F.; MARTINS, R. de S.; VIANA, W. Investigating the accessibility of popular mobile Android apps: a prevalence, category, and language study. In: SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB (WEBMEDIA), 30., 2024. Anais [...]. Porto Alegre: SBC, 2024. p. 400-404.

ARCANJO, L. D. S.; COELHO, L. F.; GUIMARÃES, S. J. F.; PATROCÍNIO JÚNIOR, Z. K. do; CARDOSO, L. V. Automatic time-aware recognition of Brazilian sign language based on dynamic time warping. In: SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB (WEBMEDIA), 30., 2024. Anais [...]. Porto Alegre: SBC, 2024. p. 72-79.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. ABNT NBR 17225:2025: acessibilidade em conteúdo e aplicações web — requisitos. Rio de Janeiro: ABNT, 2025. Disponível em: https://mwpt.com.br/wp-content/uploads/2025/04/ABNT-NBR-17225-Acessibilidade-Digital.pdf. Acesso em: 12 maio 2026.

BARTALOTTI, Celina Camargo. Inclusão social das pessoas com deficiência: utopia ou possibilidade?. Paulus, 2006.

BYERLEY, Suzanne L.; BETH CHAMBERS, Mary. Accessibility of Web-based library databases: the vendors' perspectives. Library Hi Tech, v. 21, n. 3, p. 347-357, 2003.

BERSCH, Rita. Introdução à tecnologia assistiva. Porto Alegre: Assistiva Tecnologia e Educação, 2017.

BIGDATACORP; MOVIMENTO WEB PARA TODOS. Panorama da acessibilidade digital no Brasil. São Paulo: BigDataCorp, 2024. Disponível em: https://blog.bigdatacorp.com.br/acessibilidade-nos-sites-brasileiros-2024/. Acesso em: 8 maio 2026.

BI, Tingting et al. Accessibility in software practice: A practitioner's perspective. ACM Transactions on Software Engineering and Methodology (TOSEM), v. 31, n. 4, p. 1-26, 2022.

BRASIL. Lei nº 10.048, de 8 de novembro de 2000. Dá prioridade de atendimento às pessoas que especifica, e dá outras providências. Diário Oficial da União: Brasília, DF, 9 nov. 2000. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l10048.htm. Acesso em: 11 maio 2026.

BRASIL. Lei nº 10.098, de 19 de dezembro de 2000. Estabelece normas gerais e critérios básicos para a promoção da acessibilidade das pessoas portadoras de deficiência ou com mobilidade reduzida, e dá outras providências. Diário Oficial da União: Brasília, DF, 20 dez. 2000. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l10098.htm. Acesso em: 11 maio 2026.

BRASIL. Decreto nº 5.296, de 2 de dezembro de 2004. Regulamenta as Leis nº 10.048, de 8 de novembro de 2000, e nº 10.098, de 19 de dezembro de 2000. Diário Oficial da União: Brasília, DF, 3 dez. 2004. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/decreto/d5296.htm. Acesso em: 11 maio 2026.

BRASIL. Decreto nº 6.949, de 25 de agosto de 2009. Promulga a Convenção Internacional sobre os Direitos das Pessoas com Deficiência e seu Protocolo Facultativo, assinados em Nova York, em 30 de março de 2007. Diário Oficial da União: Brasília, DF, 26 ago. 2009. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2007-2010/2009/decreto/d6949.htm. Acesso em: 11 maio 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). Diário Oficial da União: seção 1, Brasília, DF, 7 jul. 2015. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13146.htm. Acesso em: 11 maio 2026.

BRASIL. Agenda 2030 para o Desenvolvimento Sustentável. Tradução do Centro de Informação das Nações Unidas para o Brasil (UNIC Rio). Brasília, DF, 2016. Disponível em: https://brasil.un.org/sites/default/files/2020-09/agenda2030-pt-br.pdf. Acesso em: 8 maio 2026.

CASTELLS, Manuel. A Galáxia Internet: reflexões sobre a Internet, negócios e a sociedade. Zahar, 2003.

COSTA, Rafael Rodrigues; OMAIA, Diego; ARAÚJO, Thiago Moreira; PEREIRA, José Otávio; COUTINHO, André Silva; CRUZ, Marcelo Pereira; FILHO, Gilson Luiz da Silva. Acessibilidade na TV 3.0 brasileira a partir de mídias de legenda, glosa e áudio descrição. In:

Simpósio Brasileiro de Sistemas Multimídia e Web, 29., 2023, Juiz de Fora. Anais do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2023). Porto Alegre: Sociedade Brasileira de Computação, 2023. p. 123-129.

COSTA, V.; TOMASZEWSKI, J. P.; PEREIRA, L.; SANTANA, B. S.; CORRÊA, G. AI-Based Approaches for Brazilian Sign Language Recognition: A Systematic Literature Review: Insights on Methods, Metrics, and Resources for LIBRAS Recognition. In: SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB (WEBMEDIA), 31., 2025. Anais [...]. Porto Alegre: SBC, 2025. p. 585-597.

CRESWELL, John W. Qualitative inquiry & research design: choosing among five approaches. 2. ed. Thousand Oaks: Sage Publications, 2007.

DA COSTA NUNES, Erik Henrique et al. Digital accessibility at the brazilian symposium on human factors in computing systems (IHC): An updated systematic literature review. In: Proceedings of the XXII Brazilian Symposium on Human Factors in Computing Systems. 2023. p. 1-15.

DE FÁTIMA GRANATTO, Cleusa; PALLARO, Marynea AP; BIM, Sílvia Amélia. Digital accessibility: systematic review of papers from the Brazilian symposium on human factors in computer systems. In: Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems. 2016. p. 1-10.

FIGUEIRA, Emílio. As Pessoas Com Deficiência na História do Brasil: uma trajetória de silêncio e gritos!. Wak, 2021.

FILHO, Itamar Rocha et al. An Approach for Automatic Description of Characters for Blind People. In: Proceedings of the Brazilian Symposium on Multimedia and the Web. 2021. p. 53-56.

FRANÇA, Lúcio Cauper Freitas de; ANDRADE, Lisieux Marie M. dos S.; VIANA, Windson. Um estudo sobre o uso de prompts LLM para a avaliação da acessibilidade web dos portais das prefeituras da região dos Sertões dos Crateús. In: SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB (WEBMEDIA), 31., 2025. Anais [...]. Porto Alegre: SBC, 2025. p. 516-525.

GARRETT, Jesse James. The elements of user experience: user-centered design for the web and beyond. 2. ed. Berkeley: New Riders, 2011.

GOODMAN, Elizabeth; KUNIAVSKY, Mike; MOED, Andrea. Observing the user experience: a practitioner's guide to user research. 2. ed. Amsterdam; Boston: Morgan Kaufmann Publishers, 2012.

HENRY, S. L.; ABOU-ZAHRA, S.; BREWER, J. The role of accessibility in a universal web. In: WEB FOR ALL CONFERENCE, 11., 2014, Seoul, Coreia. Proceedings [...]. New York: Association for Computing Machinery, 2014. Artigo 17. Disponível em: https://doi.org/10.1145/2596695.2596719. Acesso em: 9 fev. 2026.

HORTON, Sarah; QUESENBERY, Whitney. Uma web para todos: projetando experiências de usuário acessíveis . Rosenfeld Media, 2014.

ISO. ISO 9241-210:2019: ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems. Geneva: International Organization for Standardization, 2019. Disponível em: https://standards.iteh.ai/catalog/standards/sist/ec053ca4-2add-4e8c-9f0c-9664331ee35c/iso-9241-210-2019. Acesso em: 12 maio 2026.

JÚNIOR, Mário Cléber Martins Lanna (Ed.). História do movimento político das pessoas com deficiência no Brasil. Presidência da República, Secretaria de Direitos Humanos, Secretaria Nacional de Promocão dos Direitos da Pessoa com Deficiência, 2010.

JÚNIOR, J. M. L. de M.; ARAÚJO, M. de C. C.; FAÇANHA, A. R.; VIANA, W.; SÁNCHEZ, J. Uma ferramenta para a customização de ambientes virtuais para práticas de orientação e mobilidade. In: SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB (WEBMEDIA), 28., 2022. Anais [...]. Porto Alegre: SBC, 2022. p. 103-106.

LAZAR, Jonathan; GOLDSTEIN, Daniel F.; TAYLOR, Anne. Ensuring digital accessibility through process and policy. Waltham: Morgan Kaufmann, 2015.MOURA, Filipa Raquel Teixeira. Avaliação da Acessibilidade dos Sítios web das PME's Portuguesas. 2009.

LIMA, Manuella Aschoff; FRANÇA, Daniel de; SILVA NETO, Antônio da; SOUZA, Daniel Faustino L. de; OMAIA, Diego; ARAÚJO, Tiago Maritan Ugulino de. LibrasDetec: um componente de detecção de movimentos para karaokê em Libras. In: SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB (WEBMEDIA), 31., 2025. Anais [...]. Porto Alegre: SBC, 2025. p. 340-348.

LOPES, Renan; FAÇANHA, Agebson Rocha; VIANA, Windson. I can't pay! Accessibility analysis of mobile banking apps. In: Proceedings of the Brazilian Symposium on Multimedia and the Web. 2022. p. 253-257.

MAGALHÃES, Gabriel; FONSECA, Davi; AMORIM, Glauco; VIANA, Windson; SANTOS, Joel dos. Designing a Multimodal Feedback Component for VR Applications. In: BRAZILIAN SYMPOSIUM ON MULTIMEDIA AND THE WEB (WEBMEDIA), 31., 2025, Rio de Janeiro, RJ. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2025. p. 194-201.

MARTINS, Luisa; FERREIRA, Stênio; MARITAN, Tiago. Automatic 3D animation generation for Sign Language: A case study of Sign Language Machine Translation. In: Brazilian Symposium on Multimedia and the Web (WebMedia). SBC, 2025. p. 67-76.

MITTAL, Anant et al. Jod: Examining design and implementation of a videoconferencing platform for mixed hearing groups. In: Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility. 2023. p. 1-18.

MONTEIRO, I. T. Accessibility by mediation dialogues: development and evaluation of a web navigation helper. 2011. Dissertação (Mestrado em Informática) — Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2011.

MORAES, J. M.; CARMO FILHO, P.; ARAÚJO, M. de C. C.; COSTA, B.; AMORIM, M.; FAÇANHA, A. R.; et al. Dashboards to support rehabilitation in orientation and mobility for people who are blind. In: SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB (WEBMEDIA), 29., 2023. Anais [...]. Porto Alegre: SBC, 2023. p. 6-10.

MORAES JÚNIOR, José Martônio Lopes de; ARAÚJO, Maria do Carmo Cavalcante; FAÇANHA, Agebson Rocha; VIANA, Windson. Virtual Reality for People with Visual Impairments in the Orientation and Mobility Context. In: SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB (WEBMEDIA), 30., 2024. Anais Estendidos [...]. Porto Alegre: SBC, 2024.

MOURA, Filipa Raquel Teixeira. Avaliação da Acessibilidade dos Sítios web das PME's Portuguesas. 2009.

NIC.br. Cartilha de acessibilidade na Web: fascículo IV: tornando o conteúdo Web acessível [livro eletrônico]. Coordenação de Reinaldo Ferraz. Ilustrações de Mônica Lopes. São Paulo: Comitê Gestor da Internet no Brasil, 2020. Disponível em: https://www.w3c.br/pub/Materiais/PublicacoesW3C/cartilha-w3cbr-acessibilidade-web-fasciculo-IV.pdf. Acesso em: 9 maio 2026.

NIC.br. Pesquisa sobre o uso das tecnologias de informação e comunicação nos domicílios brasileiros: TIC Domicílios 2024 [livro eletrônico]. São Paulo: Comitê Gestor da Internet no Brasil, 2025. Disponível em: https://cetic.br/media/docs/publicacoes/2/20250512120132/tic_domicilios_2024_livro_eletronico.pdf. Acesso em: 8 maio 2026.

NORMAN, Don. The design of everyday things. Revised and expanded edition. New York: Basic Books, 2013.

OSTERWALDER, Alexander; PIGNEUR, Yves; BERNARDA, Gregory; SMITH, Alan; PAPADAKOS, Trish. Value proposition design: how to create products and services customers want. Hoboken: John Wiley & Sons, 2014.

PARDINI, Rafael; BÁRBARA, João; SCHEID, Henrique; PEREIRA, Ana Carolina; MEIRA JÚNIOR, Wagner; FERRAZ, Rogério; ROCHA, Bruno. Observatório da acessibilidade da web brasileira. In: SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB (WEBMEDIA), 27., 2021. Anais Estendidos [...]. Porto Alegre: SBC, 2021. p. 71-74.

PARTHASARATHY, P. D. et al. Skill, will, or both? understanding digital inaccessibility from accessibility professionals' viewpoint. In: Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems. 2025. p. 1-9.

PÓVOA, Marcello. Anatomia da Internet: investigações estratégicas sobre o universo digital. Casa da Palavra, 2000.

PREECE, Jennifer; ROGERS, Yvonne; SHARP, Helen. Interaction design: beyond human-computer interaction. New York: John Wiley & Sons, 2002.

RIBEIRO, Carlos Nery; OLIVEIRA, Cynthya Letícia Teles de; ARAUJO, Lucas Padilha Modesto de; RODRIGUES, Kamila Rios da Hora; MANZATO, Marcelo Garcia. ARIA e Interactive Access: projetando chatbots para idosos. In: CONCURSO DE TRABALHOS DE INICIAÇÃO CIENTÍFICA (CTIC), 4., 2024. Anais Estendidos [...]. Porto Alegre: SBC, 2024. p. 41-44.

ROBERTSON, Toni; SIMONSEN, Jesper. Challenges and opportunities in contemporary participatory design. Design Issues, Cambridge, v. 28, n. 3, p. 3-9, 2012.

RODRIGUES, Kamila Rios da Hora; SANTOS, Suzane Santos dos; GALLEGO, Daniele; MARTINS, Ketlen; MALPARTIDA, Katherin Felipa Carhuaz; VERHALEN, Aline Elias Cardoso; DEUS, João Pedro de. Práticas com smartphones para idosos: um projeto de extensão do ICMC/USP. In: WEBMEDIA FOR EVERYONE (W4E), 3., 2024. Anais Estendidos [...]. Porto Alegre: SBC, 2024. p. 239-243.

ROHRER, Christian. When to Use Which User-Experience Research Methods. Fremont: Nielsen Norman Group, 2014.

SCHULER, Douglas; NAMIOKA, Aki (org.). Participatory design: principles and practices. Hillsdale: Lawrence Erlbaum Associates, 1993.

SILVA, Marcos André Bezerra da; LIMA, Manuella Aschoff C. B.; SILVA, Diego Ramon Bezerra da; SOUZA, Daniel Faustino L. de; SILVA, Samuel de Moura Moreira da; ARAÚJO, Tiago Maritan Ugulino de. Uma investigação sobre técnicas de data augmentation aplicadas à tradução automática Português-LIBRAS. In: SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB (WEBMEDIA), 30., 2024. Anais [...]. Porto Alegre: SBC, 2024. p. 319-325.

SOCIEDADE BRASILEIRA DE COMPUTAÇÃO (SBC). Grandes Desafios da Computação no Brasil 2025–2035. Coordenação: André Luís de Medeiros Santos e Flávio Rech Wagner. Porto Alegre: Sociedade Brasileira de Computação, 2025.

STICKDORN, Marc; HORMESS, Markus; LAWRENCE, Adam; SCHNEIDER, Jakob. This is service design doing: applying service design thinking in the real world. Sebastopol: O'Reilly Media, 2018.

Tavares, Daniela Cardoso; Penha, Marcio Rogério; Borges, José Antonio dos Santos; Dias, Angélica Fonseca da Silva; Ferreira, Thiago de Melo; "GOOGLEVOX: UMA INTERFACE ADAPTATIVA PARA ALAVANCAR A INTERAÇÃO DE DEFICIENTES VISUAIS EM PESQUISAS NO GOOGLE", p-472-482. In: . São Paulo: Blucher, 2017.

TAVARES, Daniela Cardoso. Vídeos mediadores do conhecimento: uma abordagem para a comunicação acessível a pessoas com deficiência visual. 2019. Dissertação de Mestrado. Instituto Politecnico de Leiria (Portugal).

VIEIRA, Alex de Souza; GUEDES, Álan Lívio V.; MORAES, Daniel de Sousa; MADEIRA, Lucas Ribeiro; COLCHER, Sérgio; SOARES NETO, Carlos de Souza. ListeningTV: Accessible Video using Interactive Audio Descriptions. In: Simpósio Brasileiro de Sistemas Multimídia e Web, 26., 2020, São Luís. Anais Estendidos do Simpósio Brasileiro de Sistemas

Multimídia e Web (WebMedia 2020). Porto Alegre: Sociedade Brasileira de Computação, 2020. p. 71–74. DOI: 10.5753/webmedia_estendido.2020.13065.

WACNIK, Peter; DALY, Shanna; VERMA, Aditi. Participatory design: a systematic review and insights for future practice. In: Proceedings of the Design Society. Cambridge: Cambridge University Press, 2023. v. 3, p. 2285–2294. DOI: https://doi.org/10.1017/pds.2023.229.

W3C BRASIL. Cartilha acessibilidade na Web: fascículo 2: benefícios, legislação e diretrizes da acessibilidade na Web. São Paulo: Comitê Gestor da Internet no Brasil, 2015. Livro eletrônico. ISBN 978-85-5559-008-5. Disponível em: https://www.w3c.br/pub/Materiais/PublicacoesW3C/cartilha-w3cbr-acessibilidade-web-fasciculo-II.pdf. Acesso em: 11 maio 2026.

WORLD WIDE WEB CONSORTIUM (W3C). Diretrizes de Acessibilidade para Conteúdo Web (WCAG) 2.2. Recomendação do W3C, 12 dez. 2024. Disponível em: https://www.w3.org/TR/WCAG22/. Acesso em: 11 maio 2026.

XAVIER, L. E. O.; RABELO, D. M. F.; MARTINS, R. de S.; VIANA, W. Image-to-UI Accessibility: Assessing ChatGPT’s Effectiveness in Generating Accessible Android Screens from Figma Templates. In: SIMPÓSIO BRASILEIRO DE SISTEMAS MULTIMÍDIA E WEB (WEBMEDIA), 31., 2025. Anais [...]. Porto Alegre: SBC, 2025. p. 302-311.