# O papel da conscientização no enfrentamento ao *Cyberbullying*

Michel Gomes Nogueira[1]
Daniel Chaves Cafe[2]

**Resumo:** O estudo aborda os desafios éticos e sociais do *cyberbullying* na forma de *deepfakes*, práticas que ameaçam a privacidade e a integridade digital. A pesquisa foi conduzida por meio de implementação de um programa de conscientização em uma escola pública, envolvendo palestras, dinâmicas e materiais educativos. Os resultados indicaram alta satisfação dos participantes e aumento significativo do conhecimento sobre riscos e estratégias de prevenção, com destaque para a compreensão dos impactos psicológicos e jurídicos. A principal contribuição do trabalho é demonstrar a eficácia de ações educativas para promover letramento digital e responsabilidade no uso das tecnologias. Como limitações, destacam-se a amostra restrita a uma única instituição e a ausência de acompanhamento longitudinal para avaliar mudanças comportamentais duradouras.

**Palavras-chave:** Cyberbullying, Inteligência Artificial e Deepfake

**Abstract:** The study addresses the ethical and social challenges of cyberbullying in the form of deepfakes. practices that threaten privacy and digital integrity. The research was conducted through the implementation of an awareness program in a public school, involving lectures, interactive activities, and educational materials. The results indicated high participant satisfaction and a significant increase in knowledge about risks and prevention strategies, with particular emphasis on understanding the psychological and legal impacts. The main contribution of the work is to demonstrate the effectiveness of educational actions to promote digital literacy and responsible use of technologies. Limitations include the sample size restricted to a single institution and the lack of longitudinal follow-up to assess lasting behavioral changes.

**Keywords:** Cyberbullying, Artificial Intelligence e Deepfake

[1] Aluno da Pós-Graduação em Engenharia Elétrica pela Universidade de Brasília.

[2] Professor Magistério Superior do Departamento de Engenharia Elétrica pela Universidade de Brasília.

## 1. INTRODUÇÃO

O avanço das tecnologias de inteligência artificial (IA) tem impulsionado significativamente a geração de conteúdos. Esses conteúdos são produzidos por meio de redes neurais profundas, utilizando principalmente técnicas de aprendizado profundo (*Deep Learning*), como as *Generative Adversarial Networks (GANs)* (GOODFELLOW et al., 2014).

O *deep learning* é um subconjunto do aprendizado de máquina (*Machine Learning*) que utiliza redes neurais compostas por múltiplas camadas para representar a complexidade do processo de tomada de decisão do cérebro humano. Geralmente, essas redes possuem três ou mais camadas, podendo chegar a centenas ou milhares, para treinar os modelos. Diferencia-se do *machine learning* tradicional, cujos modelos são considerados não profundos, utilizando redes neurais simples, com apenas uma ou duas camadas computacionais. Atualmente, técnicas de *deep learning* estão presentes na maioria das aplicações de inteligência artificial que permeiam o cotidiano (HOLDSWORTH; SCAPICCHIO, [s.d.]).

Goodfellow et al. (2014) descrevem o conceito como uma estrutura composta por redes adversárias. Nessa abordagem, um modelo generativo é confrontado por um modelo adversário, que, nesse caso, é um modelo discriminativo. O modelo generativo pode ser comparado a uma equipe de falsificadores, cujo objetivo é produzir moedas falsas e utilizá-las sem ser detectado. Já o modelo discriminativo é análogo à polícia, que busca identificar as falsificações. A ideia central é criar uma competição entre os dois modelos, de modo que ambos aprimorem seus métodos continuamente, até que as falsificações se tornem praticamente indistinguíveis dos itens genuínos.

Essas falsificações podem ser utilizadas para a prática de atos ilícitos, como o *cyberbullying*. Entre as práticas mais notáveis e em ampla expansão está o uso de *deepfakes*, que, embora possam ter aplicações legítimas, como na indústria do entretenimento, também representam uma séria ameaça à segurança digital e à integridade da informação.

O uso de *deepfakes* configura uma forma de *cyberbullying*, cujo propósito envolve práticas ilícitas e obscuras, como inserir o rosto de pessoas em vídeos pornográficos sem qualquer tipo de consentimento ou conhecimento. Soma-se a isso a crescente facilidade de criação de conteúdos falsos de áudio e vídeo, utilizados para chantagem, intimidação e sabotagem. O alcance dessas práticas pode se expandir para áreas como política e relações internacionais, favorecendo a divulgação de informações falsas com o objetivo

de incitar violência, desacreditar líderes e instituições ou até mesmo influenciar processos eleitorais (CHESNEY; CITRON, 2019).

Consequentemente, o *cyberbullying* emerge como uma forma de violência digital que se intensifica com o uso das redes sociais e outras plataformas online. Diferentemente do *bullying* tradicional, o *cyberbullying* ocorre em ambientes virtuais, sem limitações temporais ou espaciais, podendo atingir as vítimas de forma contínua e global. Entre suas características estão o anonimato, maior impacto, ampla audiência, individualidade na aplicação, entre outras (FERREIRA; DESLANDES, 2018).

Essa prática está associada a impactos psicológicos severos, como depressão, ansiedade e ideação suicida, além de consequências sociais e educacionais, como evasão escolar, afastamento social e isolamento (ARAÚJO et al., 2024).

É fundamental compreender que o *bullying* e o *cyberbullying* são fenômenos complexos que transcendem a esfera jurídica. Mais do que medidas punitivas, a prevenção eficaz requer educação e diálogo sobre os impactos dessas práticas. A construção de uma Internet segura e saudável depende do engajamento coletivo, envolvendo toda a sociedade na orientação sobre os benefícios e riscos do uso das tecnologias, promovendo conscientização e responsabilidade digital (SAFERNET, 2024).

Os desafios éticos e sociais decorrentes dessas práticas são amplamente discutidos na literatura acadêmica. Segundo Westerlund (2019), a crescente sofisticação dessas falsificações representa uma ameaça à confiança pública na mídia, podendo ser utilizada para manipular e influenciar a opinião pública. Além disso, muitos desses ataques têm como objetivo causar danos psicológicos, políticos, financeiros e até físicos (MIRSKY; LEE, 2021).

A relação entre *deepfakes* e *cyberbullying* é particularmente preocupante, sobretudo em ambientes escolares, pois tem como objetivo atacar a honra e a imagem de crianças e adolescentes, configurando novas modalidades de violência digital (FIDALGO, 2024).

A proliferação do *cyberbullying* por meio do uso de *deepfakes* tem resultado em incidentes significativos que evidenciam os riscos associados a essa tecnologia. Estudos recentes indicam que 98% dos vídeos de *deepfake* disponíveis online envolvem conteúdo pornográfico, e 99% desses materiais afetam especialmente mulheres. Em 2019, o número total de vídeos *deepfake* registrou um aumento de 550% (JUANATEY, 2025).

Em resposta ao aumento de *deepfakes* pornográficos, autoridades têm buscado reforçar a legislação para penalizar essas práticas. Na Espanha, o Ministério Público solicitou que vídeos sexuais com rostos substituídos sejam considerados delitos, destacando a necessidade de uma reforma legislativa diante do uso crescente da Inteligência Artificial para criar conteúdos que prejudiquem indivíduos (CORBACHO; CEDEIRA, 2024).

Na Coreia do Sul, *deepfakes* pornográficos têm impactado significativamente a juventude, com jovens atuando tanto como vítimas quanto como perpetradores. *Bots* no Telegram facilitam a criação desses conteúdos, permitindo que adolescentes produzam *deepfakes* de colegas, refletindo uma cultura sexista na qual esses atos são frequentemente tratados como brincadeiras (VAN, 2024).

O governo australiano apresentou uma lei que tipifica como crime a criação ou o compartilhamento de imagens pornográficas *deepfake* sem consentimento. A legislação prevê pena de até sete anos de prisão para quem produzir ou divulgar imagens sexualmente explícitas alteradas digitalmente sem autorização. A nova norma se aplica apenas a material sexual envolvendo adultos, enquanto conteúdos relacionados a crianças permanecem sob legislação específica. A medida integra um conjunto de ações voltadas para combater a violência contra mulheres e enfrentar o papel da tecnologia na propagação de atitudes misóginas (MERCER, 2024).

A legislação brasileira sobre *cyberbullying* foi fortalecida com a sanção da Lei nº 14.811/2024, que incluiu no Código Penal o artigo 146-A, tipificando *bullying* e *cyberbullying* como crimes. A pena para o crime de intimidação sistemática virtual varia de dois a quatro anos de reclusão, além de multa, caso a conduta não configure delito mais grave. A lei também instituiu a Política Nacional de Prevenção e Combate à Violência contra Crianças e Adolescentes, dando ênfase especial aos ambientes educacionais (BRASIL, 2024).

A problemática do estudo evidencia que os *deepfakes* e o *cyberbullying* representam alguns dos maiores desafios tecnológicos, éticos e sociais da atualidade. Pesquisas apontam que o *cyberbullying* está associado a diversos problemas de ordem moral, seus impactos podem gerar consequências psiquiátricas que comprometem a saúde mental e o desempenho escolar, especialmente entre adolescentes vítimas dessas práticas. Segundo Ferreira e Deslandes (2018), “aquele que é intimidado com cyberbullying teria

aproximadamente oito vezes mais chances de levar uma arma para a escola do que outros estudantes que não tiveram essa experiência".

Embora os *deepfakes* representem riscos significativos, eles não são necessariamente irremediável. As tecnologias de detecção tendem a evoluir, e ações judiciais poderão, ocasionalmente, resultar em vitórias contra criadores de falsificações prejudiciais. Paralelamente, as principais plataformas de mídia social devem tornar-se gradualmente mais eficazes na identificação e remoção de conteúdos fraudulentos. Além disso, soluções baseadas na verificação de procedência digital, se amplamente adotadas, poderão oferecer uma resposta mais duradoura no futuro (CHESNEY; CITRON, 2019).

No entanto, atualmente existe uma lacuna significativa nas estratégias de detecção, regulamentação e ações educativas, que devem ser priorizadas para garantir um ambiente digital seguro e confiável.

Este estudo tem como objetivo geral promover a conscientização sobre os riscos e impactos do *cyberbullying*, especialmente na forma de *deepfakes*, desenvolvendo competências críticas e de letramento digital que permitam aos participantes utilizar as redes sociais de maneira mais segura e responsável. Além de propor medidas para reduzir os danos causados por essa tecnologia, é fundamental investir no desenvolvimento de ferramentas eficazes capazes de detectar automaticamente vídeos com rostos substituídos (KORSHUNOV; MARCEL, 2018).

O estudo também propõe alcançar diversos objetivos, entre eles compreender a tecnologia utilizada na criação e detecção de *deepfakes*, explorando seus principais mecanismos e aplicações. Busca desenvolver o pensamento crítico em adolescentes e jovens, capacitando-os a identificar *deepfakes* e questionar a veracidade das informações consumidas online, além de aprimorar o letramento digital dos participantes, incentivando práticas seguras e responsáveis no uso das redes sociais.

Pretende ainda reduzir incidentes de *cyberbullying* relacionados ao uso de *deepfakes*, por meio da conscientização sobre os danos emocionais e as implicações legais dessas ações. Outro objetivo é avaliar o impacto das ações de conscientização, utilizando questionários e dinâmicas de grupo para medir mudanças no comportamento e na percepção de riscos. O estudo também busca contribuir para a produção acadêmica, apoiando o desenvolvimento de artigos científicos na área de Segurança Cibernética, e

divulgar os resultados obtidos em revistas e congressos especializados, fortalecendo a discussão científica sobre o tema.

## 2. O CONTEXTO HISTÓRICO DA INTELIGÊNCIA ARTIFICIAL (IA) E AS DIVERSAS DEFINIÇÕES DE CYBERBULLYING

A Ilíada, embora originariamente pertencente à tradição oral e posteriormente fixada por escrito, incorpora elementos tecnológicos imaginários que lhe conferem um caráter singular na literatura arcaica grega, como os autômatos forjados nas oficinas do deus Hefesto. Esses dispositivos míticos antecipam concepções sobre máquinas autônomas e inteligência artificial, cujas raízes conceituais remontam à Antiguidade. No entanto, a IA consolidou-se como disciplina científica apenas no século XX, conforme apontado por McCorduck (2004).

Contudo, foi com Alan Turing, em 1950, que surgiu um marco decisivo: o Teste de Turing, proposto como critério para avaliar se uma máquina poderia exibir comportamento inteligente. Em seu trabalho, Turing levanta a questão fundamental: "as máquinas podem pensar?" e, posteriormente, sugere reformulá-la para "as máquinas podem imitar o comportamento humano?". Para isso, descreve o chamado jogo da imitação, no qual um interrogador deve distinguir, por meio de perguntas e respostas, entre um ser humano e uma máquina (TURING, 1950).

Da época de Turing até os dias atuais, a Inteligência Artificial evoluiu para um campo interdisciplinar dedicado ao desenvolvimento de sistemas capazes de simular processos cognitivos humanos, como raciocínio, aprendizado e percepção. Trata-se de uma das áreas mais recentes no âmbito das ciências e da engenharia, cujo termo "Inteligência Artificial" foi cunhado em 1956, marcando o início formal da disciplina (RUSSELL; NORVIG, 2021).

O *bullying* nas escolas começou a ser estudado de forma sistemática na década de 1970, principalmente na Escandinávia. Em 1973, o pesquisador Dan Olweus publicou um trabalho pioneiro sobre o tema, traduzido para o inglês em 1978 sob o título *Aggression in Schools: Bullies and Whipping Boys*. Nesse estudo, o *bullying* foi definido como um conjunto de comportamentos agressivos, físicos e verbais, direcionados de forma intencional e repetitiva contra indivíduos em situação de vulnerabilidade (SMITH, 2013).

Nos anos 1980, Dan Olweus desenvolveu um questionário para identificar casos de bullying, ferramenta que se tornou essencial para pesquisas subsequentes. Em 1983, juntamente com a primeira campanha nacional contra o *bullying* na Noruega, ele elaborou um programa de intervenção escolar. A avaliação do Programa de Prevenção ao *Bullying* de Olweus (1983-1985) indicou uma redução aproximada de 50% nos casos, resultado que impulsionou novas pesquisas e inspirou a criação de outros programas voltados ao combate ao *bullying* (SMITH, 2013).

Com o advento da internet, surgiu uma nova modalidade de *bullying* praticada por meios eletrônicos, especialmente por celulares e computadores, conhecida como *cyberbullying*. Esse fenômeno é definido como um ato agressivo e intencional, realizado por um indivíduo ou grupo, utilizando recursos digitais para estabelecer contato de forma repetida e prolongada contra uma vítima que apresenta dificuldade para se defender. O crescimento do *cyberbullying* está diretamente relacionado à ampla disseminação de computadores conectados à rede e dispositivos móveis entre os jovens (SMITH et al., 2008).

O avanço das tecnologias digitais introduziu características que favorecem práticas nocivas no âmbito do *cyberbullying*. Entre os fatores mais relevantes está a possibilidade de anonimato, que dificulta a identificação dos agressores. Ferramentas como contas de e-mail temporárias, pseudônimos em plataformas de bate-papo e aplicativos de mensagens permitem que os autores ocultem sua identidade. Essa condição oferece aos indivíduos um elevado grau de proteção, tornando mais complexos tanto a responsabilização quanto os mecanismos de defesa das vítimas (HINDUJA; PATCHIN, 2005).

A seguir, apresenta-se uma linha do tempo que destaca marcos históricos relacionados à evolução da Inteligência Artificial (IA), incluindo a IA generativa, bem como definições sobre o fenômeno do *cyberbullying*.

### 2.1. Linha do Tempo da Inteligência Artificial

A evolução da Inteligência Artificial é marcada por marcos históricos que transformaram a ciência da computação e impactaram profundamente a sociedade. A seguir, destacam-se os principais eventos:

**1943 – Primeiro neurônio artificial**

Warren McCulloch e Walter Pitts publicam o artigo *A Logical Calculus of Ideas Immanent in Nervous Activity*, propondo o primeiro modelo matemático de neurônio artificial, base para as redes neurais (MCCORDUCK, 2004).

**1950 – Teste de Turing**

Alan Turing apresenta o teste que avalia se uma máquina pode demonstrar comportamento inteligente semelhante ao humano. Para Turing, a questão "as máquinas podem pensar?" deve ser substituída por "as máquinas podem imitar um ser humano?" (TURING, 1950; PEGN, 2024).

**1956 – Conferência de Dartmouth**

John McCarthy cunha o termo "Inteligência Artificial" durante a conferência que marca o nascimento oficial da IA como disciplina científica (FLORES, 2025).

**1957-1958 – LISP e Perceptron**

John McCarthy cria a linguagem LISP, essencial para pesquisas em IA (MCCORDUCK, 2004). Paralelamente, Frank Rosenblatt desenvolve o Perceptron, uma rede neural simplificada para reconhecimento de padrões (PEGN, 2024).

**1965 – ELIZA**

Joseph Weizenbaum desenvolve o chatbot ELIZA, que simula um psicoterapeuta, demonstrando o potencial da IA no processamento de linguagem natural (MAZER, 2023).

**1970-1980 – "Inverno da IA"**

Devido a limitações tecnológicas e expectativas irreais, ocorre uma queda no financiamento e no interesse por IA (PEGN, 2024; ATRA, 2024).

**1980 – Revitalização da IA**

Surgem os sistemas especialistas, projetados para resolver problemas em domínios específicos (FLORES, 2025).

**1986 – Backpropagation**

Geoffrey Hinton e colegas popularizam o algoritmo de retropropagação, impulsionando o avanço das redes neurais (MAZER, 2023).

**1997 – Deep Blue**

O supercomputador da IBM derrota Garry Kasparov, campeão mundial de xadrez, evidenciando a capacidade das máquinas em tarefas cognitivas complexas (PEGN, 2024).

**2006 – Deep Learning**

Geoffrey Hinton reintroduz redes neurais profundas, revolucionando o aprendizado de máquina (MAZER, 2023).

**2011 – Siri**

A Apple lança o Siri, inaugurando a era dos assistentes virtuais inteligentes (PEGN, 2024).

**2022 – ChatGPT**

A OpenAI lança o ChatGPT, popularizando a IA generativa e transformando a interação homem-máquina (SOBREIRA, 2025).

## 2.2. IA Generativa: Um Novo Paradigma

A Inteligência Artificial Generativa (IAGen) é uma tecnologia capaz de criar conteúdos originais a partir de comandos em linguagem natural. Diferente da simples curadoria de informações, ela produz textos, imagens, vídeos, músicas e códigos. Seu funcionamento baseia-se em padrões aprendidos em grandes volumes de dados coletados online (HOLMES et al., 2024).

A geração ocorre por meio da análise estatística das distribuições de palavras, pixels e outros elementos. Assim, a IAGen identifica e replica padrões comuns para criar novos conteúdos (HOLMES et al., 2024).

Exemplos notáveis incluem GPT (Generative Pre-trained Transformer), voltado para texto, e [3]DALL·E, para imagens, que demonstram o potencial da IA em aplicações diversas, como na educação para criação de plataformas educacionais; saúde para criação de imagens médicas sintéticas para treinamento; entretenimento para produção de filmes, jogos e *streaming* de música (CAMPOI, 2025).

Segundo Machado et al. (2025), a integração da IA generativa é uma tendência irreversível, mas exige literacia digital crítica e diretrizes éticas claras para evitar abusos. Apesar das vantagens, a IA generativa apresenta riscos significativos, tais como:

> ***Deepfakes* e Desinformação:** criação de conteúdos falsos, como imagens e vídeos, com aparência realista para fins de manipular eleições (NUNAN, 2023).
>
> **Plágio e Direitos Autorais:** produção de textos, músicas e imagens que imitem obras protegidas por direitos autorais, sem qualquer forma de autorização, remuneração ou atribuição (CHAGAS, 2025). Consequentemente, risco de uma apatia cognitiva, o que pode gerar uma atrofia no desenvolvimento de competências humanas essenciais e dificuldade na capacidade de análise e síntese (WOTCKOSKI, 2025).
>
> **Quebra na Segurança da Informação:** uso de dados sensíveis por parte dos usuários em ferramentas de IA, absorvidos pelos modelos, sem controle de anonimização ou descarte (CARDOSO, 2025).
>
> **Viés Algorítmico:** inserção de vieses humanos e sociais em escala (FERNANDES; GOLDIM, 2023), o que pode refletir e ampliar desigualdades e preconceitos históricos e influenciar nas escolhas e decisões (MACHADO et al. 2025).

[3] O "DALL" em DALL-E é uma homenagem ao artista surrealista Salvador Dalí, enquanto o "E" se refere ao robô animado Wall-E da Pixar. Seu sucessor, o DALL-E 2, foi apresentado em abril de 2022 e foi projetado para gerar imagens mais fotorrealistas, em resoluções mais altas (AWAN, 2024).

> **Informações imprecisas:** conhecidas também como “alucinações” quando informações são geradas de forma errôneas ou completamente inventadas (como se tivesse alucinando), mesmo quando o modelo parece estar apresentando uma resposta coerente e convincente (ROWLAND, 2023 *apud* SILVA, 2025).

Diante desses desafios, a IA exige políticas nacionais e internacionais e marcos regulatórios para garantir benefícios à humanidade (UNESCO, 2025). A UNESCO (2021) aprovou a primeira recomendação global sobre ética da IA, publicada em português em 2022 para fomentar o debate no Brasil. O documento reconhece impactos significativos, positivos e negativos, da IA sobre a mente humana, interações sociais, processos decisórios e áreas como educação e ciências.

Algumas tendências de IA seguem transformado as sociedades e as economias ao redor do mundo, que incluem agentes autônomos capazes de agir de forma independente; modelos multimodais que integram texto, imagem e áudio; e sistemas com maior raciocínio apoiados por chips customizados. Destacam-se também IA humanoide para colaboração; pequenos modelos mais eficientes; IA explicável para transparência; regulação global como diferencial competitivo e mudanças no mercado de trabalho, com novas profissões surgindo e outras desaparecendo (NEXUX.AI, 2025).

### 2.3. Definições de Cyberbullying

O cyberbullying representa um fenômeno complexo que reflete as transformações sociais trazidas pela tecnologia. Sua compreensão demanda análise do contexto sociocultural e definição clara, pois envolve práticas que podem parecer inofensivas, mas geram impactos significativos na vida das vítimas. A consolidação do conceito é essencial para orientar pesquisas, políticas públicas e estratégias de prevenção.

Segundo Schreiber e Antunes (2015) esclarecem que quanto ao conceito de *cyberbullying* não há consenso único entre os estudiosos do assunto, mas as definições convergem para a ideia de uso intencional e repetido de tecnologias de comunicação para causar dano. Eles destacam alguns estudos, dentre eles, os estudos de Belsey, que descrevem como o uso de e-mails, celulares e mensagens para difamar ou prejudicar indivíduos. Quanto aos estudos da pesquisadora Willard, ela enfatiza o conceito baseado em conteúdos ofensivos e discriminatórios. Para Smith, ele define como ação agressiva e intencional realizada por um grupo ou por um indivíduo, com o uso de forma de contato

eletrônico, de forma repetida e ao longo de um período contra uma vítima que não consegue se defender com facilidade. Hinduja e Patchin ampliam para qualquer ofensa deliberada via texto eletrônico.

Apesar da diversidade conceitual, todas as definições destacam elementos comuns: intencionalidade, repetição e exclusão, utilizando ferramentas tecnológicas para causar danos. Assim, o *cyberbullying* é uma extensão do *bullying* tradicional para o ambiente digital, com características próprias que exigem atenção (SCHREIBER e ANTUNES, 2015).

Nos Estados Unidos, a partir de 2005, como resultado de vários casos extremos, o *cyberbullying* se popularizou em razão de uma massificação de debates sobre o assunto pela mídia tradicional (HENSON, 2012).

## 3. METODOLOGIA

A pesquisa foi conduzida por meio de uma revisão bibliográfica sistemática, utilizando bases de dados científicas como Google Acadêmico, arXiv, Portal de Periódicos da CAPES, além de jornais e revistas online. Foram selecionados artigos relevantes que abordam os temas *deepfakes* e *cyberbullying*, com o objetivo de identificar práticas consolidadas e ações semelhantes já aplicadas na área.

Além da revisão bibliográfica, a metodologia incluiu a implementação de um programa de conscientização sobre os riscos e impactos das *deepfakes* e do *cyberbullying*. Para avaliar a efetividade das ações, foram aplicados questionários imediatamente após palestras e eventos educativos. Questionários adicionais foram disponibilizados em panfletos, infográficos e vídeos, acessíveis por meio de QR codes, visando ampliar a participação da comunidade.

Essa abordagem permitiu não apenas a análise crítica da literatura existente, mas também a mensuração do impacto das estratégias de conscientização, garantindo uma integração entre teoria e prática.

A metodologia foi implementada conforme a proposta do projeto submetida e aprovada pelo Comitê de Enfrentamento da Desinformação, em atendimento ao Edital DEX/DEG/DPG/DPI nº 01/2025. O projeto teve como critérios: divulgar informações fundamentadas em evidências científicas; promover o uso responsável e ético de

tecnologias geradoras de conteúdos por inteligência artificial; e desenvolver uma campanha de conscientização.

Foram estabelecidos contatos com quatro escolas públicas e privadas do Distrito Federal. Entretanto, apenas o Centro Educacional GISNO manifestou interesse em participar, possibilitando a realização de duas palestras sobre os riscos e impactos das *deepfakes* e do cyberbullying para um público aproximado de 150 pessoas (alunos, professores e equipe de apoio).

A escolha dessa instituição se justificou por ser uma escola pública que atende estudantes do ensino médio, com faixa etária predominante entre 15 e 18 anos, público diretamente exposto às tecnologias digitais e às práticas sociais abordadas pelo projeto.

Cabe destacar que, por a campanha ter sido realizada exclusivamente no Centro Educacional GISNO, os resultados obtidos refletem apenas a realidade dessa instituição, não sendo representativos do conjunto das escolas do Distrito Federal. Para alcançar achados mais robustos e generalizáveis, seria necessária a participação de um número maior de instituições de ensino do DF.

### 3.1. GISNO

O Centro Educacional GISNO (CED GISNO), localizado na Asa Norte (SGAN 907, Bloco A), em Brasília-DF, é uma instituição pública estadual vinculada à Secretaria de Estado de Educação do Distrito Federal (SEEDF). Criado em 1971 como Ginásio do Setor Noroeste, passou a se chamar Centro Educacional 02 de Brasília Norte em 1977 e, em 1979, adotou o nome atual (BRASÍLIA MINHA, 2025).

A escola oferece as seguintes modalidades de ensino: Ensino Fundamental, Ensino Médio e Educação de Jovens e Adultos (EJA), funcionando nos turnos diurno, vespertino e noturno (ESCOL.AS, 2025). Atende cerca de 488 matrículas, conforme dados do Censo Escolar (QEDU, 2025). A infraestrutura inclui biblioteca, sala de leitura, laboratórios de informática e ciências, quadra de esportes, cozinha, recursos de acessibilidade e internet banda larga, além de sala de recursos multifuncionais para atendimento educacional especializado (CED GISNO, 2019).

### 3.2. Realização da palestra de conscientização.

No dia 07 de novembro de 2025, foi realizada no auditório do Centro Educacional GISNO uma palestra de conscientização cujo tema foi: “*Deepfakes e Cyberbullying*”.

O evento contou com duas sessões destinadas aos alunos do ensino médio, ocorridas às 09h30 e às 11h00, cada uma com duração aproximada de 1 hora e 20 minutos.

As palestras foram conduzidas pelo professor Daniel Chaves Café, pelo pós-graduando Michel Gomes Nogueira e pelos graduandos da Universidade de Brasília (UnB): Mariana da Silva Reis, Maria Luiza Vasconcelos do Nascimento e Kauan Henrique da Silva Rodrigues. A graduanda Alicia Rodrigues de Souza Barbosa participou na elaboração dos panfletos, nas reuniões de planejamento e revisão dos documentos produzidos.

Todos são integrantes do projeto de ensino, pesquisa, extensão e inovação voltado ao desenvolvimento de estratégias para enfrentamento da desinformação, no âmbito do Edital DEX/DEG/DPG/DPI nº 01/2025 – Comitê UnB de Enfrentamento da Desinformação.

Devido à limitação de assentos no auditório, as palestras foram divididas em duas apresentações, abordando os mesmos tópicos:

- É IA ou é real?;
- O que é deepfake?;
- Como são criados os *deepfakes;*
- Impactos negativos dos *deepfakes;*
- Crimes relacionados ao *cyberbullying* e ao compartilhamento de materiais difamatórios;
- O uso de *deepfakes;*
- Como identificar *deepfakes;*
- Como denunciar *cyberbullying;*
- Infográfico orientativo sobre *bullying* e *cyberbullying;*
- Espaço para dúvidas e perguntas; e
- Convite para participação da avaliação da palestra

O professor Daniel Café iniciou a palestra apresentando a equipe do projeto e contextualizando os impactos psicológicos e morais que essas práticas podem causar na vida dos adolescentes. Foram exibidos vídeos e imagens geradas por inteligência artificial para demonstrar as possibilidades atuais da tecnologia.

Além disso, foi realizada uma dinâmica interativa com os participantes, na qual eram exibidos fotos e vídeos e os alunos deveriam identificar se eram reais ou produzidos por IA.

Ao final, os alunos receberam um infográfico orientativo, (formato A4), com informações sobre como lidar com situações de bullying e *cyberbullying*, conforme ilustrado na figura 3.1.

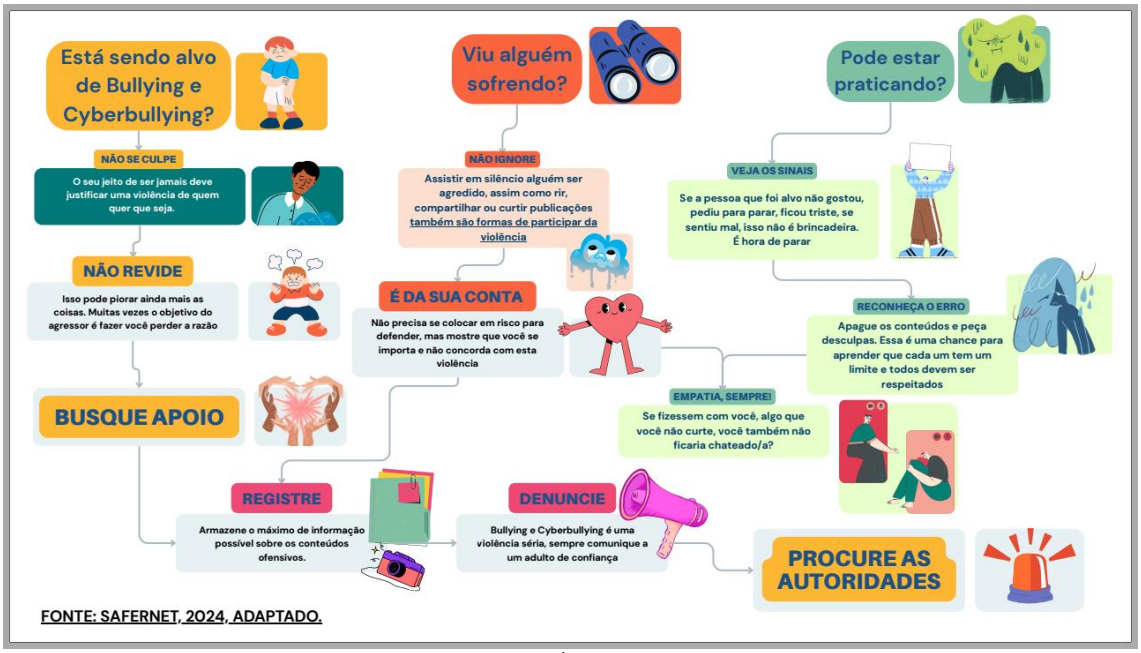



Fig. 3.1 – Infográfico orientativo [4](SAFERNET, 2024, adaptado).

O infográfico foi apresentado e detalhado aos alunos, com orientações sobre como agir em casos de *bullying* ou *cyberbullying*.

O conteúdo destacou a importância de buscar apoio e ajuda caso sejam vítimas dessas práticas ou conheçam alguém que esteja passando por situações semelhantes. Para aqueles que praticam *bullying* ou *cyberbullying*, o infográfico trouxe recomendações sobre como reconhecer o erro, pedir desculpas e adotar uma postura empática em relação à pessoa alvo da ação difamatória.

Além disso, foram entregues à coordenação pedagógica do GISNO dois infográficos em tamanho A1, contendo alertas sobre os riscos dos *deepfakes* e do *cyberbullying*, bem como orientações para identificar sinais de manipulação digital, conforme ilustrado na figura 3.2.

[4] Cedido pelo graduando Kauan Henrique da Silva Rodrigues


Fig. 3.2 – Panfleto orientativo (2025)[5]

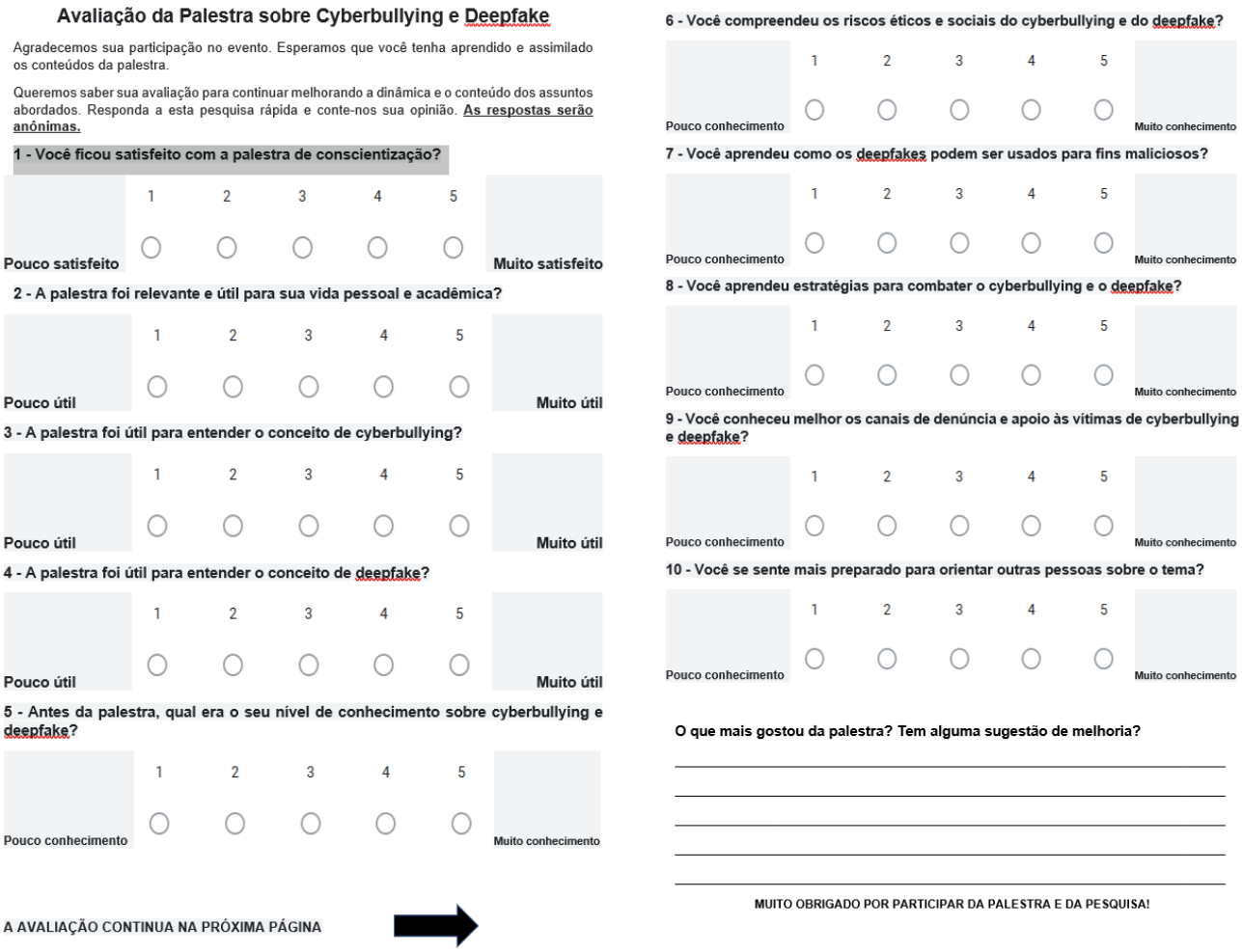
Ao término das apresentações, foi aberto um espaço para perguntas e esclarecimento de dúvidas. Em seguida, alunos e professores receberam uma folha de avaliação da palestra, composta por 10 perguntas de múltipla escolha e um espaço para comentários, onde os participantes puderam registrar o que mais gostaram e sugerir melhorias.

As perguntas foram elaboradas para avaliar: a qualidade da palestra; o conhecimento prévio e o aprendizado adquirido. A avaliação utilizou uma escala de cinco níveis, contemplando os seguintes critérios:

- Satisfação: de pouco satisfeito a muito satisfeito;
- Utilidade: de pouco útil a muito útil; e
- Conhecimento: de pouco conhecimento a muito conhecimento.

[5] Cedido pela graduanda Alicia Rodrigues de Souza Barbosa.

A figura 3.3 ilustra o modelo utilizado na avaliação.



**Avaliação da Palestra sobre Cyberbullying e Deepfake**

Agradecemos sua participação no evento. Esperamos que você tenha aprendido e assimilado os conteúdos da palestra.

Queremos saber sua avaliação para continuar melhorando a dinâmica e o conteúdo dos assuntos abordados. Responda a esta pesquisa rápida e conte-nos sua opinião. **As respostas serão anônimas.**

**1 - Você ficou satisfeito com a palestra de conscientização?**

Pouco satisfeito 1 2 3 4 5 Muito satisfeito

**2 - A palestra foi relevante e útil para sua vida pessoal e acadêmica?**

Pouco útil 1 2 3 4 5 Muito útil

**3 - A palestra foi útil para entender o conceito de cyberbullying?**

Pouco útil 1 2 3 4 5 Muito útil

**4 - A palestra foi útil para entender o conceito de deepfake?**

Pouco útil 1 2 3 4 5 Muito útil

**5 - Antes da palestra, qual era o seu nível de conhecimento sobre cyberbullying e deepfake?**

Pouco conhecimento 1 2 3 4 5 Muito conhecimento

A AVALIAÇÃO CONTINUA NA PRÓXIMA PÁGINA

**6 - Você compreendeu os riscos éticos e sociais do cyberbullying e do deepfake?**

Pouco conhecimento 1 2 3 4 5 Muito conhecimento

**7 - Você aprendeu como os deepfakes podem ser usados para fins maliciosos?**

Pouco conhecimento 1 2 3 4 5 Muito conhecimento

**8 - Você aprendeu estratégias para combater o cyberbullying e o deepfake?**

Pouco conhecimento 1 2 3 4 5 Muito conhecimento

**9 - Você conheceu melhor os canais de denúncia e apoio às vítimas de cyberbullying e deepfake?**

Pouco conhecimento 1 2 3 4 5 Muito conhecimento

**10 - Você se sente mais preparado para orientar outras pessoas sobre o tema?**

Pouco conhecimento 1 2 3 4 5 Muito conhecimento

**O que mais gostou da palestra? Tem alguma sugestão de melhoria?**

MUITO OBRIGADO POR PARTICIPAR DA PALESTRA E DA PESQUISA!

Fig. 3.3 – Avaliação da Palestra (autor, 2025)

## 4. RESULTADOS

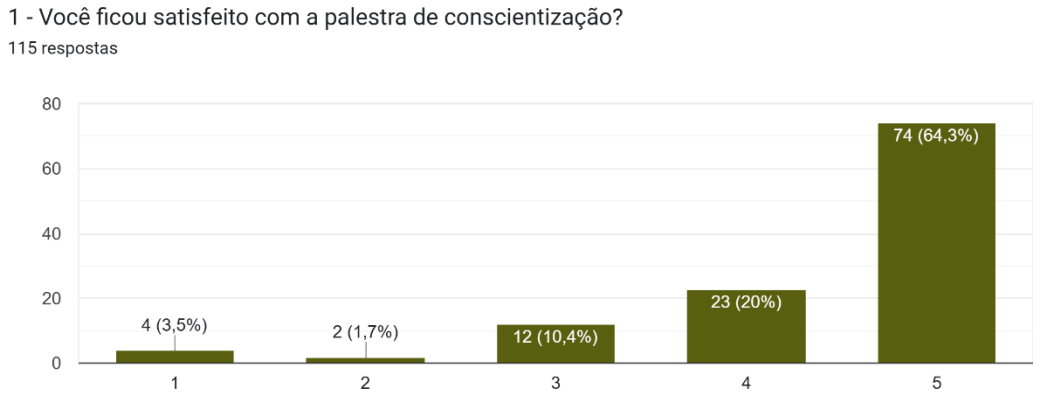
Ao todo, foram respondidos manualmente 115 formulários por alunos, professores e apoio. Posteriormente, as respostas foram inseridas em um formulário do *Google Forms*, elaborado pela equipe do projeto. O preenchimento manual de formulários impressos ocorreu devido à Lei Federal nº 15.100, de 13 de janeiro de 2025, que regulamenta o uso de aparelhos eletrônicos portáteis pessoais (incluindo celulares) por estudantes nos estabelecimentos de ensino da educação básica.

Considerou-se que, mesmo quando utilizados para fins pedagógicos ou didáticos, não seria conveniente permitir o uso desses dispositivos no auditório do GISNO, pois nem todos os celulares tinham acesso à rede Wi-Fi ou à internet móvel. Por esse motivo, optou-se pelo uso de formulários impressos para a coleta das respostas, em vez do formato digital, no caso via *QR Code*.

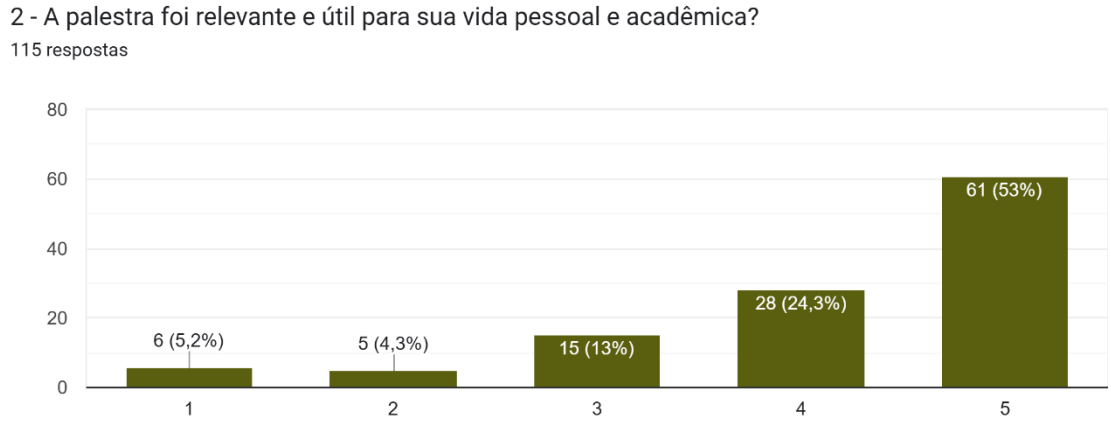
### 4.1. Análise dos dados da pesquisa

A análise dos dados de *feedback* da palestra de conscientização sobre *cyberbullying* e *deepfake* revela uma recepção geralmente muito positiva por parte dos participantes.

Satisfação Geral: A maioria dos participantes expressou alta satisfação com a palestra, com uma média de 4.40 e uma moda de 5 (a classificação mais frequente). O gráfico de barras "Satisfação com a Palestra de Conscientização" mostra que 74 dos 115 participantes deram a classificação máxima (5), e 23 deram 4, indicando que a grande maioria ficou muito satisfeita.



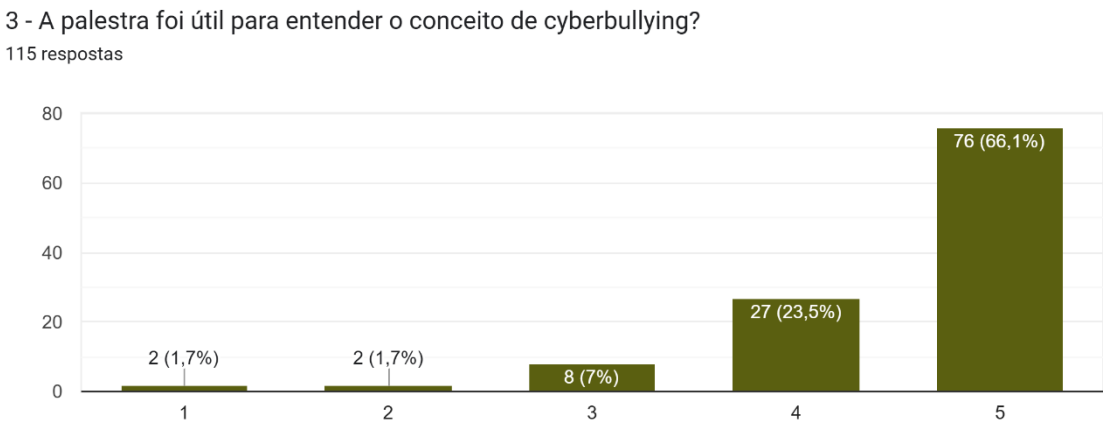

Relevância e Utilidade: A palestra foi considerada altamente relevante e útil para a vida pessoal e académica, com uma média de 4.16 e uma moda de 5. O gráfico “Relevância e Utilidade da Palestra" ilustra que 61 participantes atribuíram a classificação máxima, reforçando o impacto positivo do conteúdo.

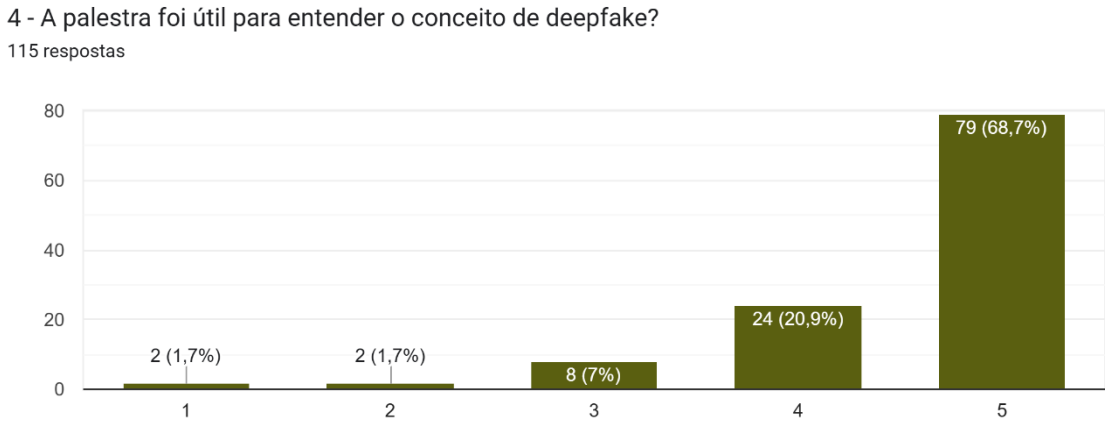


A análise dos gráficos para as perguntas 3 e 4 revela a percepção dos participantes sobre a utilidade da palestra para compreender os conceitos de *cyberbullying* e *deepfake*. A maioria dos participantes considerou a palestra muito útil para entender o conceito de *cyberbullying* e *deepfake*.

A classificação mais frequente (moda) é 5, indicando que um grande número de pessoas atribuiu a utilidade máxima. Esse resultado é particularmente importante, dado que o *deepfake* é um tópico mais recente e, por vezes, mais complexo para o público em geral.

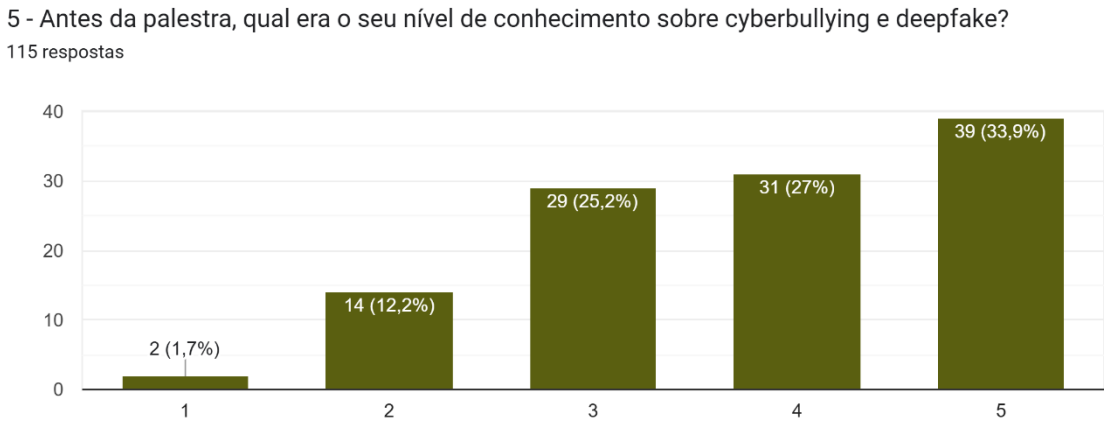
Em resumo, ambos os gráficos indicam que a palestra foi extremamente bem-sucedida em educar os participantes sobre os conceitos de *cyberbullying* e *deepfake*, com a maioria a atribuir classificações elevadas de utilidade. Isso reforça a eficácia da palestra em cumprir os seus objetivos de conscientização e educação sobre estes temas críticos.



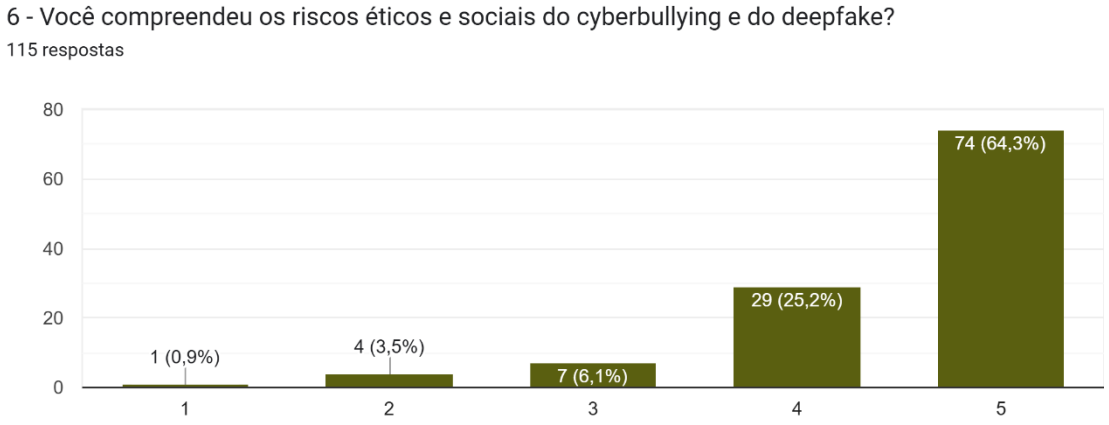




Nível de Conhecimento Prévio: A pergunta "5 - Antes da palestra, qual era o seu nível de conhecimento sobre *cyberbullying* e *deepfake*?" teve a pontuação média mais

baixa (3.79). Embora ainda seja uma pontuação razoável, a distribuição de respostas para esta pergunta é mais dispersa, com um número maior de pontuações 2 e 3 em comparação com outras perguntas. Isso sugere que, embora a palestra tenha sido eficaz em aumentar o conhecimento, havia uma base de conhecimento prévio mais variada entre os participantes.

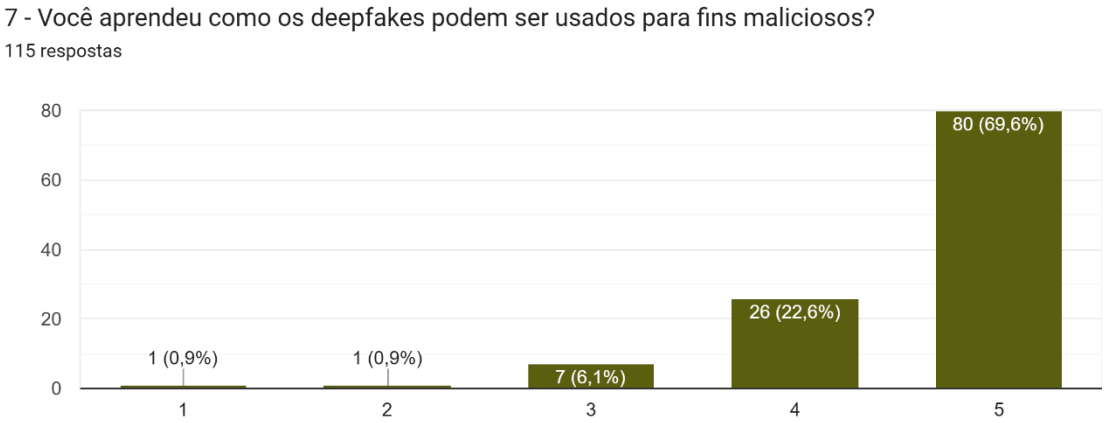



Compreensão dos Riscos: Houve uma forte compreensão dos riscos éticos e sociais do *cyberbullying* e do *deepfake*, com uma média de 4.49 e uma moda de 5. O gráfico "Compreensão dos Riscos Éticos e Sociais de *Cyberbullying* e *Deepfake*" destaca que 74 participantes classificaram a sua compreensão como 5, demonstrando a eficácia da palestra em transmitir informações relevantes.

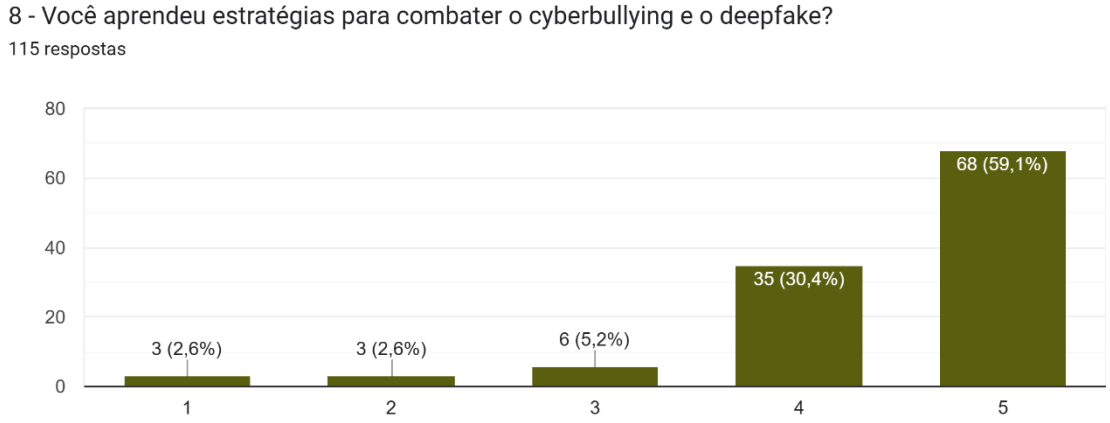



Aprendizagem sobre *Deepfakes* Maliciosos: Os participantes aprenderam como os *deepfakes* podem ser usados para fins maliciosos, com uma média de 4.59, a mais alta

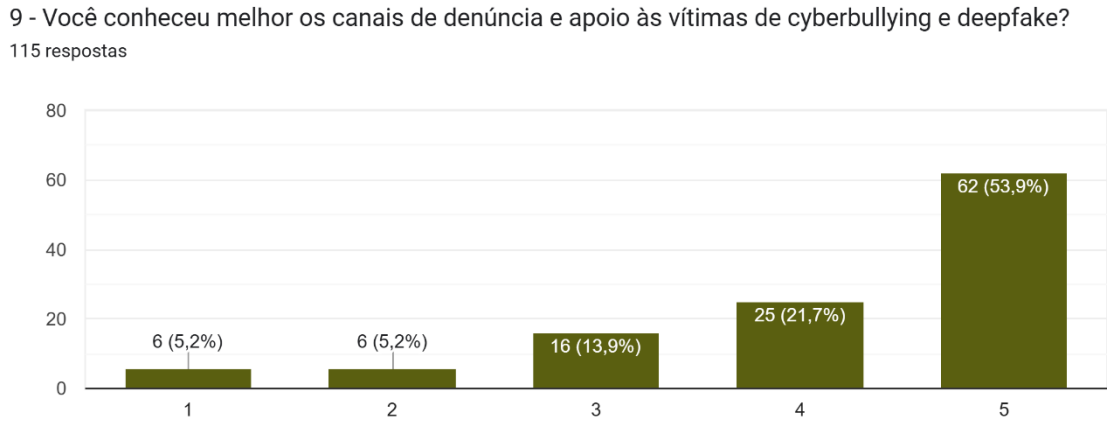
entre todas as perguntas. Isso sugere que a palestra foi particularmente eficaz em abordar este tópico sensível.



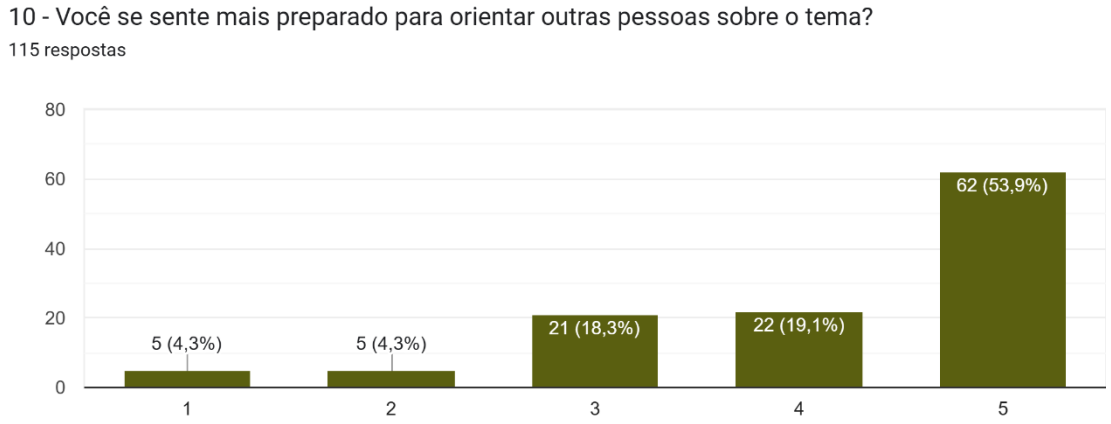

Estratégias de Combate: Houve uma boa aprendizagem de estratégias para combater o *cyberbullying* e o *deepfake*, com uma média de 4.41 e uma moda de 5. Em relação ao bom resultado desta aprendizagem pode se atribuir também ao fato da explicação e entrega dos panfletos orientativos sobre *cyberbullying* e *deepfake*.




Canais de Denúncia e Apoio: Embora a maioria dos aspetos da palestra tenha sido bem avaliada, a familiaridade com os canais de denúncia e apoio às vítimas de *cyberbullying* e *deepfake* teve uma média ligeiramente inferior (4.14). Isso sugere que pode haver uma oportunidade para reforçar a divulgação desses recursos em futuras iniciativas.

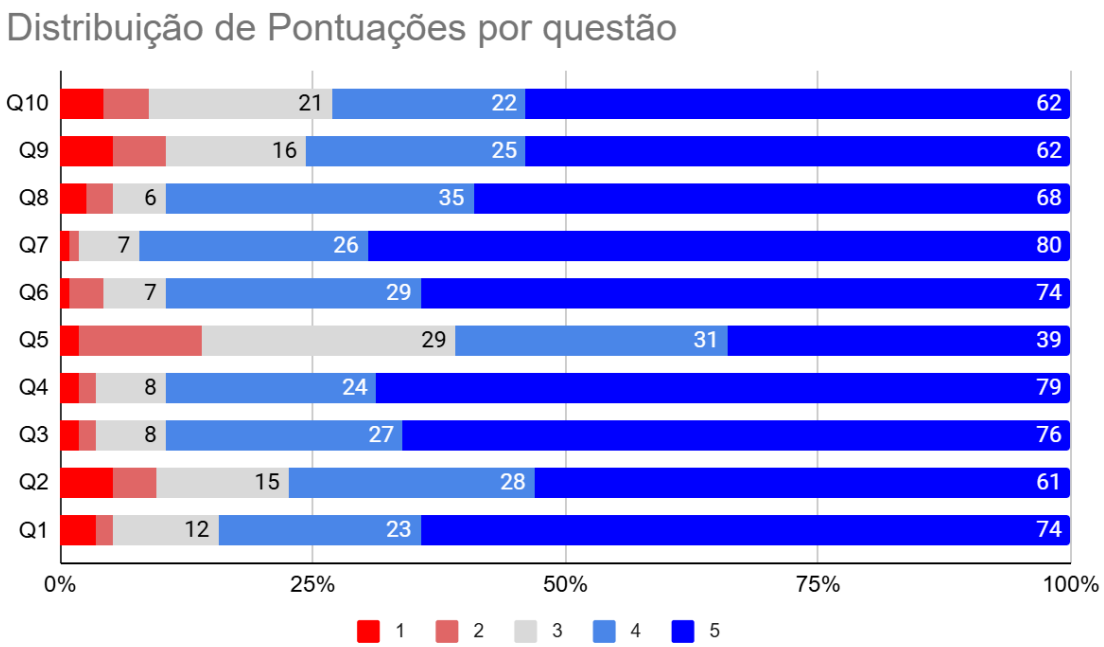



Os dados da pergunta 10 indicam que a palestra foi bem-sucedida em aumentar a confiança e a preparação dos participantes para orientar outras pessoas sobre os temas de *cyberbullying* e *deepfake*. A maioria dos participantes sente-se muito ou totalmente preparada. Este é um resultado positivo, pois demonstra que a palestra não só transmitiu conhecimento, mas também capacitou os participantes a serem agentes de mudança e conscientização nas suas comunidades.




O gráfico de barras "Distribuição de Pontuações por Questão" ilustra claramente a distribuição das respostas para cada questão. É possível observar visualmente as questões com maior concentração de pontuações altas e aquelas com uma distribuição mais variada.

As barras coloridas permitem identificar rapidamente quais pontuações são mais frequentes para cada questão, com as cores azuis representando pontuações mais altas e as cores vermelho e cinza representando pontuações mais baixas.




- **Alta Satisfação Geral:**

A maioria das questões (Q1, Q3, Q4, Q6, Q7, Q8) apresenta uma concentração significativa de respostas nas pontuações 4 e 5, indicando um alto nível de satisfação e concordância dos participantes.

A questão Q4, por exemplo, teve a maior contagem de pontuações 5, sugerindo que este aspecto específico da palestra foi extremamente bem recebido.

- **Áreas com Oportunidades de Melhoria:**

Q5 (Conhecimento Prévio): Esta questão mostra uma distribuição mais equilibrada entre as pontuações, com um número considerável de respostas nas pontuações 3 e 4, e até mesmo algumas nas pontuações 1 e 2. Isso sugere que os participantes tinham níveis de conhecimento prévio mais variados sobre o tema antes da palestra.

Q2, Q9 e Q10: Embora ainda com boas pontuações, estas questões apresentam um número um pouco maior de respostas nas pontuações 1, 2 e 3 em comparação com as questões de maior satisfação. Isso pode indicar que, para esses aspectos, há uma margem para aprimoramento ou para fornecer informações mais detalhadas.

Quanto a pergunta realizada sobre o que os participantes mais gostaram da palestra, segue a análise:

- **Clareza e Linguagem Acessível:** Muitos participantes elogiaram a forma como os palestrantes explicaram os conceitos, destacando a clareza e a acessibilidade da linguagem utilizada. Isso facilitou a compreensão dos temas complexos de *cyberbullying* e *deepfake*.
- **Interatividade e Engajamento:** A interação com o público, especialmente por meio de jogos, testes de fotos de IA e vídeos, foi um ponto alto. Esses elementos foram eficazes em prender a atenção dos alunos e tornar a palestra mais dinâmica e interessante.
- **Conteúdo Relevante e Informativo:** A palestra foi considerada muito informativa e completa, proporcionando aos participantes um maior conhecimento sobre *cyberbullying*, *deepfake* e inteligência artificial. A relevância dos temas para a vida pessoal e acadêmica foi frequentemente mencionada.
- **Importância da Conscientização:** Muitos reconheceram a importância da conscientização sobre o uso e a evolução rápida da tecnologia, bem como a necessidade de orientar outras pessoas sobre esses temas.
- **Foco em IA e *Deepfake*:** A abordagem sobre a inteligência artificial e o *deepfake*, incluindo como identificar imagens geradas por IA e seus usos maliciosos, foi particularmente apreciada e considerada muito útil.

Quanto às sugestões de melhoria:

- **Aprimorar a Identificação de IA/Deepfake:** Alguns participantes expressaram o desejo de receber mais informações e detalhes práticos sobre como identificar se uma imagem é real ou gerada por IA.
- **Qualidade do Áudio:** Houve uma sugestão para melhorar a qualidade do microfone ou o volume da fala, indicando que em alguns momentos a audibilidade pode ter sido um desafio.
- **Manter a Interação:** Embora a interação tenha sido um ponto forte, uma sugestão mencionou a necessidade de manter a interação para sustentar a atenção do público ao longo de toda a palestra.

- **Abordagem de Tópicos Adicionais:** Uma sugestão mais específica foi a de incluir a questão do gasto de energia no processamento da IA ao explicar seu funcionamento, e também trazer temas para pesquisa individual para aprofundar o aprendizado.

Em resumo, a palestra foi bem avaliada em geral, com a maioria dos participantes expressando alta satisfação e concordância. No entanto, a análise destaca a importância de considerar o nível de conhecimento prévio dos participantes e a possibilidade de aprimorar alguns aspectos específicos da palestra para garantir que todos os participantes se beneficiem ao máximo.

A palestra foi um importante em termos de conteúdo, clareza e engajamento, com os elementos interativos sendo um diferencial. As sugestões de melhoria são construtivas e podem ajudar a aprimorar ainda mais futuras edições, especialmente no que diz respeito a detalhes técnicos de identificação de *deepfakes* e à qualidade do áudio.

### 4.2. Relação entre achados da pesquisa e autores referenciados

Para complementar a análise dos resultados descritos no item 4.1, elaborou-se um mapeamento que relaciona os principais achados da pesquisa aos autores citados na fundamentação teórica. Esse quadro permite visualizar como as evidências empíricas dialogam com a literatura, reforçando a coerência entre os dados obtidos e as contribuições acadêmicas sobre *cyberbullying*, *deepfakes* e estratégias de prevenção.

Quadro 4.2 – Relação entre achados da pesquisa e autores referenciados

| Achado | Descrição resumida | Autores relacionados | Como se relaciona |
|---|---|---|---|
| Alta satisfação geral | Média 4,40 e moda 5; maioria muito satisfeita com a palestra. | Hinduja e Patchin (2005; 2015); Olweus (1978) | Programas educativos bem estruturados tendem a elevar satisfação e engajamento em prevenção ao *bullying/cyberbullying*; instrumentos de avaliação e intervenção escolar já consolidados. |
| Alta relevância/‘utilidade’ percebida | Média 4,16 e moda 5; conteúdo julgado útil para vida | UNESCO (2021; 2025); Machado et al. (2025) | Diretrizes para letramento digital e uso responsável de IA reforçam utilidade de ações educativas; integração da IA |

| | | | |
|---|---|---|---|
| | pessoal/acadêmica. | | generativa em educação exige práticas de conscientização. |
| Utilidade para entender *cyberbullying* e *deepfake* | Moda 5 em questões de compreensão dos conceitos. | Smith et al. (2008); Smith (2013); Chesney e Citron (2019); Mirsky e Lee (2021) | Conceitos e impactos de *cyberbullying* e *deepfakes* fundamentam a necessidade de conteúdos explicativos; revisão técnica sobre criação/detecção de *deepfakes*. |
| Conhecimento prévio mais baixo | Q5 com média 3,79; base prévia heterogênea. | Hinduja e Patchin (2005) | Literatura mostra variabilidade de conhecimento prévio entre estudantes sobre riscos digitais; necessidade de nivelamento. |
| Compreensão dos riscos éticos e sociais | Média 4,49; forte compreensão dos riscos. | Chesney e Citron (2019); Westerlund (2019); CNECV (2024); UNESCO (2021; 2025) | Autores discutem ameaça à confiança pública, manipulação e ética da IA; documentos de orientação política/ética reforçam análise de riscos. |
| Aprendizagem sobre usos maliciosos dos deepfakes | Maior média (4,59). | Korshunov e Marcel (2018); Mirsky e Lee (2021); Juanatey (2025); Mercer (2024); Van (2024); Corbacho e Cedeira (2024) | Estudos técnicos e casos recentes demonstram proliferação e danos dos *deepfakes* (pornográficos, chantagem, etc.), corroborando o foco educativo. |
| Aprendizagem de estratégias de combate | Média 4,41 e moda 5; reforçada por materiais (infográficos). | SaferNet (2024); Hinduja e Patchin (2005/2015); BRASIL – Lei 14.811/2024 | Estratégias práticas de denúncia/apoio e arcabouço legal brasileiro fortalecem a dimensão de enfrentamento e prevenção. |
| Familiaridade com canais de denúncia menor | Média 4,14; há espaço para reforçar divulgação de canais. | SaferNet (2024); CED GISNO (2019) | Materiais e políticas públicas indicam canais oficiais; recomendação é ampliar visibilidade e uso. |
| Confiança para orientar outras pessoas | Participantes se dizem preparados para orientar | Olweus (1978); Hinduja e Patchin | Programas de prevenção e letramento digital visam formar multiplicadores; |

| | | | |
|---|---|---|---|
| | a comunidade. | (2015); UNESCO (2025) | literatura de intervenção escolar valida esse efeito. |
| *Feedback* qualitativo: clareza, linguagem acessível, interatividade | Elogios à clareza, jogos/testes, vídeos; pedido de manter interação. | Machado et al. (2025) | Abordagens pedagógicas ativas e recursos multimodais elevam engajamento e compreensão em temas digitais. |
| Sugestões: mais técnicas para identificar IA/*Deepfake* | Demanda por maior detalhamento prático. | Mirsky e Lee (2021); Korshunov e Marcel (2018) | Revisões de detecção oferecem base para incluir módulos práticos (artefatos, pixels, áudio, inconsistências de iluminação/movimento). |
| Sugestões: incluir tema de gasto de energia/impactos ambientais da IA | Interesse dos alunos em ampliar escopo. | UNESCO (2025); CNECV (2024) | Documentos de ética e políticas de IA incluem preocupações socioambientais e governança; tema pode ser incorporado em futuras edições. |

## 5. CONCLUSÃO

Este estudo reforça a importância da conscientização sobre os riscos e impactos das *deepfakes* e do *cyberbullying*, evidenciando a necessidade de desenvolver competências críticas e de letramento digital para um uso mais seguro e responsável das redes sociais. Ao explorar os mecanismos de criação e detecção das *deepfakes*, bem como suas implicações sociais e legais, foi possível destacar a relevância do pensamento crítico como ferramenta essencial para identificar conteúdos manipulados e reduzir a propagação de informações falsas.

A metodologia adotada desempenhou papel fundamental para alcançar esses objetivos. A implementação de um programa de conscientização, aliado à aplicação de questionários após palestras e eventos educativos, permitiu avaliar a efetividade das ações propostas. Além disso, a disponibilização de questionários, distribuição de panfletos, infográficos e vídeos ampliou a participação da comunidade, garantindo maior alcance e engajamento.

Essa abordagem prática e interativa contribuiu para mensurar mudanças no comportamento e na percepção de riscos, fortalecendo a validade dos resultados obtidos.

As ações propostas, como a capacitação de adolescentes e jovens, a promoção de práticas digitais seguras e a sensibilização quanto aos danos emocionais e jurídicos do *cyberbullying*, contribuem para minimizar os impactos negativos dessas tecnologias. A avaliação das mudanças comportamentais e perceptivas dos participantes, aliada à produção acadêmica e à divulgação científica, fortalece o debate sobre segurança cibernética e incentiva soluções inovadoras, como sistemas de detecção baseados em IA e regulamentações específicas.

Como limitações da pesquisa, destaca-se a amostra restrita ao Centro Educacional GISNO, o que impede a generalização dos resultados para outras escolas do Distrito Federal. Além disso, não foi realizado acompanhamento longitudinal, impossibilitando avaliar mudanças comportamentais duradouras.

Sugere-se, para pesquisas futuras, a abordagem dos seguintes temas: (i) desenvolvimento de ferramentas automatizadas para detecção de *deepfakes* acessíveis ao público geral; (ii) realização de estudos longitudinais que permitam avaliar a eficácia das ações de conscientização ao longo do tempo; (iii) integração de programas educativos nos currículos escolares, ampliando a formação crítica e o letramento digital; (iv) estabelecimento de parcerias com plataformas digitais para implementação de políticas mais rigorosas contra conteúdos manipulados e práticas de cyberbullying; e (v) investigações sobre os impactos psicológicos das *deepfakes* e estratégias de mitigação voltadas às vítimas.

Dessa forma, este estudo não apenas cumpre seu papel educativo e preventivo, mas também se posiciona como um passo significativo para a construção de um ambiente digital mais ético, seguro e consciente.

## 1. REFERÊNCIAS BIBLIOGRÁFICAS

ARAÚJO, Gabriel da Silva et al. Violência digital: o impacto do cyberbullying nas redes sociais e suas consequências na saúde mental. Revista F&T, v. 29, n. 140, 2024. DOI: 10.69849/revistaft/cl10202411301207. Disponível em: < https://revistaft.com.br/violencia-digital-o-impacto-do-cyberbullying-nas-redes-sociais-e-suas-consequencias-na-saude-mental/>. Acesso em: 10 dez. 2025.

ATRA. A linha do tempo da Inteligência Artificial. Disponível em: <https://www.atra.com.br/2024/10/24/a-linha-do-tempo-da-inteligencia-artificial/>. Acesso em: 12 dez. 2025.

AWAN, Abid A. DATACAMP. O que é DALL-E? Disponível em: <https://www.datacamp.com/pt/blog/what-is-dall-e>. Acesso em: 04 dez. 2025.

BRASIL. Lei nº 14.811, de 12 de janeiro de 2024. Institui medidas de proteção à criança e ao adolescente contra a violência nos estabelecimentos educacionais ou similares. Disponível em: <https://www.planalto.gov.br/ccivil_03/_Ato2023-2026/2024/Lei/L14811.htm>. Acesso em: 02 dez. 2025.

BRASÍLIA MINHA. Gisno. Disponível em: <https://brasiliaminha.com.br/dw/doku.php?id=gisno>. Acesso em: 12 dez. 2025.

CAMPOI, D. Z. S. IA Generativa: Como modelos como o GPT e DALL-E estão transformando indústrias. LinkedIn, 2025. Disponível em: <https://www.linkedin.com/pulse/ia-generativa-como-modelos-o-gpt-e-dall-e-est%C3%A3o-danilo-6so4c>. Acesso em: 04 dez. 2025.

CARDOSO, Jeferson Dias. IA Generativa em Ambientes Corporativos: Análise dos Riscos de Segurança e Governança. Universidade Federal de Uberlândia, 2025. Disponível em: <https://repositorio.ufu.br/bitstream/123456789/45935/3/IAGenerativaAmbientes.pdf> Acesso em: 12 dez. 2025.

CED GISNO – Centro Educacional do Distrito Federal. Proposta Pedagógica do Centro Educacional GISNO – Ano Letivo 2019. Brasília: Secretaria de Estado de Educação do Distrito Federal (SEEDF), 2019. Disponível em: <https://www.educacao.df.gov.br/wp-conteudo/uploads/2018/07/pp_ced_gisno_plano_piloto-1.pdf>. Acesso em: 12 dez. 2025.

CHAGAS, Leandro. Os desafios éticos na utilização da IA generativa: entre o progresso e a responsabilidade. LinkedIn, 2025. Disponível em: <https://www.linkedin.com/pulse/os-desafios-%C3%A9ticos-na-utiliza%C3%A7%C3%A3o-da-ia-generativa-entre-chagas-stvzf/>. Acesso em: 12 dez. 2025.

CHESNEY, R.; CITRON, D. *Deepfakes and the new disinformation war*. Foreign Affairs, 2019. Disponível em: <https://www.foreignaffairs.com/articles/world/2018-12-11/*deepfakes*-and-new-disinformation-war>. Acesso em: 10 dez. 2025.

CNECV. Inteligência Artificial: Inquietações Sociais, Propostas Éticas e Orientações Políticas. Lisboa: Conselho Nacional de Ética para as Ciências da Vida, 2024. Disponível em: <https://www.cnecv.pt/files/1715104301_c3ad59f003de091ee280b6c153182c6b_cnecv-livro-branco-ia-maio-2024.pdf>. Acesso em: 13 dez. 2025.

CORBACHO, J; CEDEIRA, B. La Fiscalía pide que sean delito los vídeos sexuales con caras suplantadas y los ‘fakes’ para desinformar. 2024. Disponível em:

<https://www.elespanol.com/espana/20240905/fiscalia-pide-delito-videos-sexuales-caras-suplantadas-fakes-desinformar/883412285_0.html>. Acesso em: 11 dez. 2025.

ESCOL.AS. Ced Gisno - Brasília - DF. Disponível em: <https://www.escol.as/266983-ced-gisno>. Acesso em: 12 dez. 2025.

FERNANDES, Márcia; GOLDIM, José Roberto. Implicações éticas e Inteligência Artificial Generativa. Migalhas, 2023. Disponível em: <https://www.migalhas.com.br/coluna/migalhas-de-protecao-de-dados/382002/implicacoes-eticas-e-inteligencia-artificial-generativa--parte-ii>. Acesso em: 12 dez. 2025.

FERREIRA, Taiza Ramos de Souza Costa; DESLANDES, Suely Ferreira. Cyberbullying: conceituações, dinâmicas, personagens e implicações à saúde. Ciência & Saúde Coletiva, v. 23, n. 10, p. 3369-3379, 2018. DOI: 10.1590/1413-812320182310.13482018. Disponível em: < https://www.scielosp.org/article/ssm/content/raw/?resource_ssm_path=/media/assets/csc/v23n10/1413-8123-csc-23-10-3369.pdf> Acesso em: 10 dez. 2025.

FIDALGO, Adriano Augusto. As deepfakes e as deepnudes como modalidades de cyberbullying. Revista Arace, v. 6, n. 4, p. 135-150, 2024. DOI: 10.56238/arev6n4-135. Disponível em: < https://periodicos.newsciencepubl.com/arace/article/view/2076>. Acesso em: 10 dez. 2025.

FLORES, Diego. História e Evolução da Inteligência Artificial: Marcos, Pioneiros e Impactos. Quiker, 2025. Disponível em: <https://quiker.com.br/historia-da-inteligencia-artificial/>. Acesso em: 12 dez. 2025.

GOODFELLOW, I. et al. *Generative adversarial networks*. Advances in Neural Information Processing Systems, 2014. Disponível em: https://arxiv.org/abs/1406.2661. Acesso em: 10 dez. 2025.

HENSON, Billy. Bullying beyond the schoolyard: Preventing and responding to cyberbullying. Security Journal, v. 25, n. 1, p. 88-89, 2012. Disponível em: <https://link.springer.com/content/pdf/10.1057/sj.2011.25.pdf>. Acesso em: 08 dez. 2025.

HINDUJA, Sameer; PATCHIN, Justin W. Cyberbullying Fact Sheet. Retrieved on November, v. 10, p. 2010, 2005. Disponível em: <https://cyberbullying.org/cyberbullying_fact_sheet.pdf>. Acesso em: 03 dez. 2025.

HOLDSWORTH, Jim; SCAPICCHIO, Mark. O que é deep learning? [s.n.]. Disponível em: <https://www.ibm.com/br-pt/think/topics/deep-learning>. Acesso em: 01 dez. 2025.

HOLMES, Wayne et al. Guia para a IA generativa na educação e na pesquisa. UNESCO Publishing, 2024. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000390241. Acesso em: 04 dez. 2025.

JUANATEY, H. *Deepfakes porno digital: la IA para atacar mujeres políticas, activistas y periodistas*. 2025. Disponível em:

https://www.huffingtonpost.es/politica/*deepfakes*-porno-digital-falso-ia-atacar-mujeres-politicas-activistas-periodistas.html. Acesso em: 10 mar. 2025.

KORSHUNOV, P.; MARCEL, S. *Deepfakes: a new threat to face recognition? Assessment and detection*. arXiv preprint arXiv:1812.08685, 2018. Disponível em: <https://arxiv.org/abs/1812.08685>. Acesso em: 11 dez. 2025.

MACHADO, Raphael; DAVID, Rodrigo; SOUZA, Rodolfo. A Inteligência Artificial Generativa no Ecossistema Acadêmico: Uma Análise de Aplicações, Desafios e Oportunidades. arXiv, 2025. Disponível em: <https://arxiv.org/pdf/2507.03106v1>. Acesso em: 12 dez. 2025.

MAZER. Linha do Tempo da Inteligência Artificial. Disponível em: <https://www.mazer.dev/pt-br/inteligencia-artificial/curso-conceitos-ai/secao-1-visao-geral/linha-do-tempo-da-ia/>. Acesso em: 12 dez. 2025.

MCCORDUCK, Pamela; CFE, Cli. Machines who think: A personal inquiry into the history and prospects of artificial intelligence. AK Peters/CRC Press, 2004.

MERCER, P. Australia criminalizes distribution and creation of deepfake pornographic material. 2024. Disponível em: <https://www.voanews.com/a/australia-criminalizes-distribution-and-creation-of-deepfake-pornographic-material/7643430.html>. Acesso em: 11 dez. 2025.

MIRSKY, Yisroel; LEE, Wenke. The Creation and Detection of Deepfakes: A Survey. ACM Computing Surveys, v. 54, n. 1, p. 1-41, 2021. Disponível em: < https://arxiv.org/pdf/2004.11138>. Acesso em: 10 dez. 2025.

NEXUX.AI. 8 tendências de inteligência artificial que estão transformando o mercado de IA. 2025. Disponível em: <https://nexux.ai/blog/tendencias-e-futuro/as-8-tendencias-de-inteligencia-artificial-que-estao-transformando-o-mercado-de-ia>. Acesso em: 05 dez. 2025.

NUNAN, Vladimir. Inteligência Artificial Generativa: Entre a Criatividade e os Riscos Éticos. Eduvem, 2023. Disponível em: <https://eduvem.com/ia-generativa-criatividade-riscos-eticos/>. Acesso em: 12 dez. 2025.

PEGN. Inteligência artificial: relembre o surgimento e os principais marcos da tecnologia. Revista PEGN, 2024. Disponível em: <https://revistapegn.globo.com/tecnologia/noticia/2024/05/inteligencia-artificial-relembre-o-surgimento-e-os-principais-marcos-da-tecnologia.ghtml>. Acesso em: 04 dez. 2025.

QEDU. CED GISNO. Disponível em: <https://qedu.org.br/escola/53001044-ced-gisno>. Acesso em: 09 nov. 2025.

RUSSELL, Stuart; NORVIG, Peter. Inteligência Artificial: Uma Abordagem Moderna. 4. ed. Rio de Janeiro: Elsevier, 2021.

SAFERNET BRASIL. Estou enfrentando um problema de cyberbullying. Devo denunciar?, 2024. Disponível em: <https://new.safernet.org.br/content/estou-enfrentando-um-problema-de-cyberbullying-devo-denunciar>. Acesso em: 13 dez. 2025.

SCHREIBER, Fernando C. de Castro; ANTUNES, Maria Cristina. Cyberbullying: do virtual ao psicológico. Boletim Academia Paulista de Psicologia, v. 35, n. 88, p. 109-125, 2015.Disponível em: < https://www.redalyc.org/pdf/946/94640400008.pdf>. Acesso em: 08 dez. 2025.

SILVA, Lauren Simão da. Inteligência artificial generativa na produção científica: impactos positivos, desafios éticos e limitações. UNIFESP, 2025. Disponível em: <https://repositorio.unifesp.br/items/401fefe4-146c-42c1-9015-c8b7d3fd31f3>. Acesso em: 04 dez. 2025.

SMITH, P. K. et al. Cyberbullying: Its nature and impact in secondary school pupils. Journal of Child Psychology and Psychiatry, v. 49, n. 4, p. 376-385, 2008. Disponível em: < https://acamh.onlinelibrary.wiley.com/doi/epdf/10.1111/j.1469-7610.2007.01846.x>. Acesso em: 08 dez. 2025.

SMITH, Peter K. School bullying. Sociologia, problemas e práticas, n. 71, p. 81-98, 2013. Disponível em: < https://journals.openedition.org/spp/988>. Acesso em: 11 dez. 2025.

SOBREIRA, VICTOR. Um panorama da História da Inteligência Artificial e suas aplicações na pesquisa histórica. Varia Historia, v. 41, p. e25035, 2025. Disponível em: < https://www.scielo.br/j/vh/a/rBFnX4gCZ4N8fcRt6mjxbFQ/?format=pdf&lang=pt>. Acesso em: 04 dez. 2025.

TURING, A. M. Computing Machinery and Intelligence. Mind, v. 59, n. 236, p. 433-460, 1950.

UNESCO. Ética da Inteligência Artificial no Brasil. Brasília: UNESCO, 2025. Disponível em: <https://www.unesco.org/pt/fieldoffice/brasilia/expertise/artificial-intelligence-brazil> Acesso em: 5 dez. 2025.

UNESCO. Recomendação sobre a Ética da Inteligência Artificial. Paris: UNESCO, 2021. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000381137_por>. Acesso em: 5 dez. 2025.

UNIVERSIDADE DE BRASÍLIA. Edital DEX/DEG/DPG/DPI nº 1/2025-1, 2025. Disponível em: https://enfrentamentoadesinformacao.unb.br/wp-content/uploads/2025/03/Edital_DEX_DEG_DPG_DPI_N1_2025-1.pdf. Acesso em: 11 nov. 2025.

VAN, Quentin. En Corée du Sud, la jeunesse victime et bourreau des deepfakes pornographiques. 2024. Disponível em: <https://www.lemonde.fr/pixels/article/2024/08/30/en-coree-du-sud-la-jeunesse-victime-et-bourreau-des-*deepfakes*-pornographiques_6299665_4408996.html>. Acesso em: 11 mar. 2025.

WESTERLUND, Mika. The Emergence of Deepfake Technology: A Review. Technology Innovation Management Review, v. 9, n. 11, p. 39-52, 2019. Disponível em: <https://timreview.ca/sites/default/files/article_PDF/TIMReview_November2019%20-%20D%20-%20Final.pdf>. Acesso em: 10 dez. 2025.

WOTCKOSKI, R. B., Borges, M. C., Celeste, M., Andrade, S. T. de, & Oliveira, C. de F. (2025). Inteligência artificial generativa e seus desafios éticos no ensino superior: uma revisão sistemática da literatura. REVISTA DELOS, 18(70), e6187. <https://doi.org/10.55905/rdelosv18.n70-061>. Acesso em: 05 dez. 2025.