Rostos fabricados, danos reais

Um panorama dos deepfakes sexuais entre jovens brasileiros

Safer net | Safe Online | Tech Coalition

Índice

CLIQUE PARA IR DIRETO PARA O CONTEÚDO

Sumário executivo

Mensagem da Secretaria Nacional de Direitos Digitais (Sedigi/MJSP)

Mensagem do Ministério Público Federal

Mensagem da Agência Nacional de Proteção de Dados

Mensagem da SaferNet Brasil

1. Definições & tipologias

2. Violência sexual mediada por IA em escolas brasileiras

3. Pontos de vista de diferentes atores envolvidos no fenômeno

4. Ecossistema digital que facilita a violência sexual mediada por IA

5. Respostas de diferentes atores ao problema da violência sexual mediada por IA

6. Recomendações detalhadas para múltiplos atores

7. Considerações finais

Sobre esta pesquisa

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

Sumário executivo

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

QUAL O CONTEXTO DESTA PESQUISA?

A inteligência artificial (IA) generativa mudou a violência sexual on-line contra crianças e adolescentes. Ferramentas que antes exigiam conhecimento técnico especializado agora permitem que qualquer pessoa com acesso à internet produza imagens hiper-realistas de abuso em segundos.

Este relatório é resultado de dez meses de pesquisa pela SaferNet Brasil, com apoio do fundo SafeOnline e da Tech Coalition. Ele documenta o fenômeno no Brasil em três frentes: os casos em escolas, o ecossistema digital que facilita a violência e as perspectivas dos próprios jovens brasileiros sobre o problema.

QUAIS OS PRINCIPAIS ACHADOS BRASILEIROS DESTA PESQUISA?

DEEPFAKES SEXUAIS EM ESCOLAS BRASILEIRAS (CAPÍTULO 2)

Entre agosto de 2025 e fevereiro de 2026, a SaferNet Brasil mapeou 17 episódios de *deepfake*s sexuais em escolas de 11 estados brasileiros (AL, BA, CE, MG, MT, PA, PB, PE, RJ, RS e SP), envolvendo 61 infratores e 222 vítimas. A totalidade das vítimas era do gênero feminino — estudantes de 12 a 17 anos e, em alguns casos, professoras. A totalidade dos infratores era do gênero masculino, da mesma faixa etária. A proporção de quase quatro vítimas para cada infrator indica práticas coletivas ou seriais de produção e circulação de imagens em canais fechados de mensagens. Casos foram registrados tanto em escolas públicas quanto privadas, em capitais e no interior, evidenciando que o fenômeno é nacional e sistêmico, e não circunstancial ou regional.

A OPINIÃO DOS JOVENS BRASILEIROS (CAPÍTULO 3)

Uma enquete (*survey*) com 679 jovens (predominantemente entre 16 e 21 anos) revelou uma fratura de gênero profunda nas percepções sobre o problema. Enquanto 86,8% das meninas cisgênero afirmaram com certeza que criar um *deepfake* sexual de uma colega sem consentimento é crime, apenas 61,9% dos meninos deram a mesma resposta — e 9,4% deles afirmaram categoricamente que não é crime. Dos 501 meninos cisgênero da amostra, 62 (12,4%) admitiram já ter criado um *deepfake* sexual; entre as meninas cisgênero, esse número foi zero. Ao todo, 42,6% dos jovens entrevistados declararam produzir, conhecer alguém que produz ou ter ouvido falar de colegas que produzem esse tipo de conteúdo. O Telegram foi apontado como a principal plataforma onde esses materiais são criados (42,6% dos respondentes), seguido por pesquisas nos buscadores (33,9%) e por aplicativos de nudificação para celular (27,1%).

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

QUAIS AS PRINCIPAIS EVIDÊNCIAS DESTA PESQUISA?

UM ECOSSISTEMA DIGITAL DE ABUSO (CAPÍTULO 4)

Para esta pesquisa, foram selecionadas 264 URLs submetidas ao canal de denúncias da SaferNet Brasil, relacionadas a suspeitas de abuso ou exploração sexual infantil. A categoria mais prevalente foi a de conteúdos com crimes de abuso sexual contra crianças, identificada em 149 URLs (56,4% da amostra).

A análise aprofundada dos cinco estudos de caso identificou os principais vetores de risco. Por tipo de plataforma, a *dark web* apresentou a maior taxa de conteúdo ilícito (85,8% de suas 127 URLs). Já os aplicativos de nudificação, apesar de terem poucos links na amostra, destacaram-se pela extrema gravidade: 100% deles continham *deepfakes* de crianças e crimes contra crianças.

ROBÔS DE NUDIFICAÇÃO NO TELEGRAM

Entre 2023 e 2026, a SaferNet Brasil recebeu mais de 10 mil denúncias de potencial material ou conteúdo de violência sexual contra crianças e adolescentes com o domínio “t.me/”, do Telegram. *Bots* de nudificação operam com modelos freemium — imagens de teste gratuitas seguidas de acesso pago via Telegram Stars, criptomoedas ou PIX. As Stars podem ser convertidas em criptomoedas pela TON Network (blockchain vinculada ao Telegram) após 21 dias, criando um circuito financeiro que entra por vias convencionais e sai em ativos de difícil rastreamento. A cada transação, Apple e Google retêm US$ 3 e o Telegram, US$ 0,49 — três empresas lucrando sobre crimes contra crianças, com graus distintos de responsabilidade.

FÓRUNS E GERADORES NA DARK WEB

Dos 127 domínios da rede Tor analisados, 81,89% continham material real de abuso sexual infantil e 10,24% continham material artificial. A SaferNet Brasil monitorou um fórum criado exclusivamente para produção e disseminação de material gerado por IA, cujas publicações já ultrapassavam 20 mil visualizações em março de 2026. O fórum mantém centenas de guias e tutoriais — desde geração de textos descrevendo abuso até treinamento de modelos sob demanda —, com indicações que incluem ferramentas e vídeos acessíveis na internet aberta, evidenciando a porosidade entre a *surface web* e esses ecossistemas criminosos. A retórica sustenta que imagens artificiais seriam uma “alternativa ética” ao abuso real, narrativa que funciona como mecanismo de recrutamento e normalização da exploração sexual de crianças e de adolescentes.

Foi também analisado um gerador de imagens baseado em modelos de código aberto, dedicado exclusivamente à produção de material ou conteúdo de violência sexual contra crianças e adolescentes. Acesso anônimo (sem e-mail), com créditos acumulados ao avaliar mídias de outros usuários — mecanismo que configura aprendizado por reforço humano, aprimorando ativamente o modelo. Em março de 2026, havia produzido mais de 510 mil materiais com a palavra *girl* (menina, em inglês). Conteúdos referentes ao Brasil cresceram 37,6% em 60 dias (4.401 para 6.056 registros), com as idades mais frequentes nos *prompt*s sendo 12, 10, 7, 11 e 6 anos — e presença alarmante de idades de 1, 2 e 3 anos.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

PLATAFORMAS DE PORNOGRAFIA

Duas plataformas de conteúdo adulto foram analisadas como estudos de caso. Na primeira, a SaferNet Brasil acumula 8,6 mil denúncias desde 2018 (de 3 em 2018 para 2.108 em 2025); 27% dos links disponíveis continham material ou conteúdo de violência sexual contra crianças e adolescentes, e buscas por termos como *AI teen* (adolescente de IA, em inglês) retornaram mais de 5 mil resultados. Na segunda — atualmente sob investigação do regulador de comunicações do Reino Unido —, foram identificados 263 *deepfake*s envolvendo 22 vítimas, das quais 19 tinham menos de 18 anos no momento da captura das imagens originais; a pesquisa documentou táticas de ofuscação (números substituindo letras nos nomes das vítimas) para burlar moderação e impedir que as próprias vítimas localizem os materiais. Em ambos os estudos de caso, foram encontrados *deepfake*s de deputadas federais e estaduais brasileiras publicados com nomes completos, sem qualquer moderação — instrumentos de intimidação política contra mulheres em posições de poder. Nenhuma das plataformas exige verificação efetiva de idade além da autoidentificação.

COMO REAGEM AUTORIDADES PÚBLICAS E EMPRESAS?

O Brasil dispõe de instrumentos legais: o Estatuto da Criança e do Adolescente criminaliza produção, distribuição, posse e simulação de material ou conteúdo de violência sexual contra crianças e adolescentes (arts. 240, 241-A a 241-D); a Lei n.° 15.123/2025 incluiu o uso de IA na tipificação penal da violência psicológica contra mulheres; e o ECA Digital (Lei n.°15.211/2025) estabeleceu o dever de cuidado das plataformas. Entretanto, até maio de 2026, a SaferNet Brasil identificou 70 projetos de lei sobre *deepfake*s e IA no Congresso, de 15 partidos, sem coordenação unificada. A maioria deles não tem como foco direto a proteção infantil.

No plano executivo, a atuação coordenada da Autoridade Nacional de Proteção de Dados (ANPD), do Ministério Público Federal (MPF) e da Secretaria Nacional do Consumidor (Senacon) no caso Grok (janeiro de 2026) representou a resposta mais significativa, consolidando que *deepfake*s de pessoas identificáveis constituem dados pessoais sensíveis sob a Lei Geral de Proteção de Dados (LGPD). Mato Grosso tornou-se o primeiro estado a aprovar legislação específica (Lei 13.257/2026). Internacionalmente, o Reino Unido lidera com a criminalização da posse de guias para gerar material ou conteúdo de violência sexual contra crianças e adolescentes por IA, e os Estados Unidos aprovaram o *TAKE IT DOWN Act*.

Quanto às empresas de IA generativa, dez companhias aderiram a princípios voluntários de *Safety by Design* em 2024, com avanços em filtragem de *prompt*s, curadoria de dados e moderação. Entretanto, a disponibilização dos pesos de modelos de código aberto torna todas essas proteções estruturalmente reversíveis.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

QUAIS AS RECOMENDAÇÕES PRIORITÁRIAS DA SAFERNET BRASIL?

Ator	Educação e prevenção	Resposta à vitimização	Monitoramento e fiscalização	Cooperação institucional
Educadores	Integração curricular de temas como manipulação digital, consentimento, empatia e impactos psicossociais de conteúdos sexualizados não consensuais.	Protocolos institucionais de denúncia, acolhimento e encaminhamento sem revitimização.	-	Integração com famílias, conselhos tutelares e organizações da sociedade civil.
Poder Executivo	Campanhas nacionais de prevenção e educação digital voltadas a crianças, famílias e educadores.	Protocolo nacional de resposta rápida aos *deepfakes* sexuais e conteúdos artificiais de abuso e exploração sexual infantil.	Monitoramento de transações financeiras ligadas à exploração sexual infantil; criação de centro federal especializado em ameaças digitais.	-
Poder Legislativo	-	Previsão legal de procedimentos céleres para remoção de conteúdo e responsabilização.	-	Harmonização legislativa com padrões internacionais de proteção infantil.
Empresas	Design e comunicação de produtos que desestimulem a criação e circulação de conteúdo sexualmente manipulado.	-	Exigências regulatórias de segurança, transparência e auditoria independente para IA generativa; transparência sobre algoritmos, critérios de moderação e taxas de resposta.	-

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

Mensagem da Secretaria Nacional de Direitos Digitais (Sedigi/MJSP)

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

uso de ferramentas de IA generativa para fabricação de imagens falsas contra crianças e adolescentes organiza-se em uma cadeia mercadológica complexa. Aplicativos distribuídos em lojas oficiais, robôs em serviços de mensagens e sites acessíveis por navegador oferecem a mesma função, cobram por ela e se utilizam de intermediários de pagamento, hospedagem e busca para chegar ao usuário. Agir contra cada elo dessa cadeia isoladamente seria pouco efetivo. Os operadores dos sites frequentemente possuem domínios estrangeiros e, quando se derruba uma página, ela é rapidamente restabelecida em outro endereço. A persecução penal dos crimes atinge apenas o autor da imagem após o dano já ter se concretizado. Essa realidade torna indispensável da escolha regulatória de estabelecer pressupostos e balizas para responsabilização dos intermediários de conteúdo de terceiros, que são identificáveis e controlam os pontos por onde as ferramentas de modificação chegam ao público.

Esse foi o caminho adotado pelo Supremo Tribunal Federal no julgamento dos Temas 987 e 533 da Repercussão Geral. Ao declarar a inconstitucionalidade parcial do art. 19 do Marco Civil da Internet, o Tribunal reconheceu aos provedores de aplicações um dever de cuidado. Eles respondem por falha sistêmica quando não promovem a indisponibilização imediata de conteúdos que configurem crimes graves, listados em rol taxativo. O rol incluiu os crimes praticados contra a mulher em razão da condição do sexo feminino, entre eles a violência psicológica do art. 147-B do Código Penal, cuja pena é aumentada de metade quando há uso de inteligência artificial.

A partir do novo Decreto n.º 12.976, de 20 de maio de 2026, esse regime de responsabilização foi densificado, com foco na proteção de mulheres nos ambientes digitais. O art. 9º veda aos provedores de aplicações de internet a geração e a modificação de conteúdo íntimo de terceiro por inteligência artificial. O art. 10 impõe aos provedores baseados em funcionalidades de IA salvaguardas técnicas e procedimentais para identificar e bloquear essas solicitações, proporcionais ao volume de acessos e ao risco da aplicação. A obrigação recai sobre a geração do conteúdo, antes da circulação e antes que exista uma vítima. A fiscalização cabe à Agência Nacional de Proteção de Dados (ANPD), que já iniciou tomada de subsídios sobre as novas regras. A relevância desse novo regime já é sentida no ambiente digital. Após a edição do Decreto 12.976, a Secretaria Nacional de Direitos Digitais (Sedigi) realizou duas atividades de monitoramento com foco em aplicativos de nudificação. O primeiro examinou aplicativos de nudificação disponíveis nas lojas de aplicativos. O segundo identificou 32 sites que oferecem a mesma função na web aberta, em vários casos sem cadastro nem verificação de idade, e concluiu que a retirada dos aplicativos das lojas não basta quando o serviço permanece acessível pelo navegador. O material foi encaminhado à Polícia Federal e à ANPD. Em agosto, a partir de ofício da SaferNet Brasil, foi feito um novo pedido de avaliação de 80 sites. Em agosto de 2026, a Agência abriu processo de monitoramento das lojas de aplicativos e das principais ferramentas de IA generativa para verificar as medidas adotadas contra a geração de conteúdo íntimo não autorizado.

Diante de todas essas mudanças, a publicação deste relatório tem valor próprio para essa agenda. Em dez meses de pesquisa, a SaferNet Brasil reuniu em um só estudo os casos ocorridos nas escolas, o funcionamento do mercado que os abastece e a percepção dos próprios jovens sobre o problema. Parte da atuação descrita nesta mensagem nasceu de informações levadas pela entidade ao Ministério, e as recomendações que encerram o trabalho oferecem a escolas, empresas e órgãos públicos um ponto de partida comum. A Sedigi cumprimenta a equipe pela pesquisa e recomenda sua leitura a todos os que têm responsabilidade sobre o ambiente digital.

VICTOR OLIVEIRA FERNANDES
Secretário Nacional de Direitos Digitais
do Ministério da Justiça e Segurança Pública

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

Mensagem do Ministério Público Federal

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Desde 2006, o Ministério Público Federal mantém parceria com a SaferNet Brasil para o combate às violações de direitos humanos on-line com enfoque principal na violência sexual infantil, violência contra a mulher e racismo e discriminação, todos no ambiente digital. Essa parceria abrange tanto o aspecto da prevenção desses crimes quanto o aspecto criminal da persecução penal.

No âmbito da prevenção, inúmeras ações conjuntas culminaram entre 2015 e 2018 no Projeto MP pela educação digital nas escolas, num esforço coordenado também com o Comitê Gestor da Internet no Brasil, para trazer conscientização aos coordenadores de ensino de escolas públicas e privadas de todos os estados brasileiros e inserir de forma direta e transversal conteúdo sobre educação digital: como navegar com segurança na internet, respeitando as regras de cidadania que valem tanto offline como on-line e se protegendo para não ser vítima de violações on-line. Essa ação foi transformada em política pública do Ministério da Educação e hoje integra o currículo das escolas brasileiras.

No âmbito criminal, a SaferNet Brasil é o mais importante hotline brasileiro, recebendo mais de 87.689 denúncias ao ano sobre violações de direitos humanos on-line e integrando o INHOPE, a associação de provedores hotline que formam a rede global de combate ao abuso sexual on-line, o que propicia a amplitude da base de informações. Com isso, o seu repositório de denúncias agrega valioso material para as autoridades de justiça que lutam contra esse crime devastador representando uma importante ferramenta que auxilia na triagem e no cruzamento de denúncias, otimizando a investigação criminal.

O presente Relatório "Rostos fabricados, danos reais" se debruça sobre o fenômeno dos *deepfakes* sexuais criados com ferramentas de inteligência artificial envolvendo os jovens brasileiros num recorte feito entre agosto de 2025 e fevereiro de 2026.

O componente da inteligência artificial trouxe uma camada muito grave aos crimes sexuais on-line, já que a escala e rapidez na criação e distribuição dessas imagens de *deepfakes*, aumentou sobremaneira os riscos e as consequências a que as vítimas estão submetidas.

Ao se preocupar em fazer um diagnóstico preciso sobre esse fenômeno nas redes sociais, na *deep web* e em outros serviços on-line, o Relatório analisa as ferramentas de IA e técnicas empregadas, além de descrever o percurso criminoso, representando um importante material que pode ser incorporado às investigações e à persecução penal desses delitos.

Este estudo detalhado e abrangente traz dados essenciais para o planejamento das ações de política pública para lidar com esse fenômeno e o planejamento de ações para o combate aos crimes sexuais on-line incrementados com a inteligência artificial.

Mais uma vez, podemos contar com o olhar atento da SaferNet Brasil, que presta indispensável serviço à sociedade brasileira, nos auxiliando a identificar caminhos para proteger principalmente nossas crianças e adolescentes no ambiente digital, um ambiente que traz tantas oportunidades e inovações, mas que precisa de regras e responsabilidade de todos os atores envolvidos.

FERNANDA TEIXEIRA SOUZA DOMINGOS
Coordenadora do Grupo de Atuação Especial de Combate aos Crimes Cibernéticos (GACCTI – MPF)

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

Mensagem da Agência Nacional de Proteção de Dados (ANPD)

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

A pesquisa conduzida pela SaferNet Brasil oferece um necessário mapeamento de um fenômeno que, infelizmente, tem crescido no Brasil e no mundo: a violência sexual mediada por inteligência artificial. Valendo-se de sua experiência no monitoramento de denúncias de violações de direitos fundamentais e de crimes cometidos na Internet, a SaferNet Brasil realizou um minucioso diagnóstico e um verdadeiro trabalho investigativo sobre as deepfakes sexuais no contexto infantojuvenil.

O resultado é apresentação de um cenário preocupante: a circulação disseminada de deepfakes sexuais nas escolas brasileiras, a naturalização do tema entre crianças e adolescentes e a exposição de um ecossistema digital que incentiva essa prática. Isso tudo baseado em dados de alta granularidade e representativos de diversos estados brasileiros.

A pesquisa sobre a percepção de crianças e adolescentes acerca das deepfakes sexuais — o que sabem sobre o tema, o contato que têm com vítimas, a noção que possuem sobre a ilegalidade da conduta, entre outros aspectos - revela o quanto a prática já adentrou a realidade infantojuvenil, inclusive no ambiente escolar, e a disparidade dessa percepção entre meninos e meninas.

Igualmente relevantes são os dados relacionados às ferramentas e plataformas mais frequentemente associadas à criação e disseminação desse conteúdo e ao ecossistema de pagamentos que dá base para o desenvolvimento de um mercado em torno da circulação desses conteúdos.

Mais do que apresentar os problemas, a SaferNet Brasil aponta possíveis caminhos e soluções ancoradas em dados consistentes e na experiência de mais de 20 anos de atuação no Brasil. As recomendações oferecem valiosos pontos de partida aos diferentes atores envolvidos no enfrentamento da questão. São subsídios não apenas para ações imediatas, mas também para pesquisas e o desenvolvimento de políticas públicas futuras. Os achados indicam, por exemplo, a importância de direcionar políticas públicas de apoio a grupos que são desproporcionalmente afetados por tais práticas, como as meninas e mulheres.

Parabenizo a equipe pelo cuidado na elaboração desta pesquisa. Desejo uma excelente leitura e que este relatório sirva de base para debates e ações frutíferas em prol da proteção de crianças, adolescentes, mulheres e de todos os que possam ser prejudicados pela criação e disseminação destes conteúdos que têm impactado a vida de tantos.

LORENA GIUBERTI COUTINHO
Economista, doutora pela Universidade de Maastricht (Holanda) e mestre pela Universidade de Brasília, é diretora da Agência Nacional de Proteção de Dados

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

Mensagem da SaferNet Brasil

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

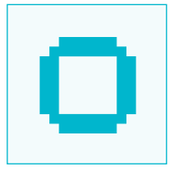
4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA



título deste relatório, *Rostos fabricados, danos reais*, explica muito sobre o que a sociedade brasileira agora presencia em 2026 com os desenvolvimentos de sistemas de IA generativa.

Na avaliação da SaferNet Brasil, os desafios trazidos por essa tecnologia colocam a sociedade numa encruzilhada histórica. O fenômeno das imagens de abuso e exploração sexual contra crianças e adolescentes geradas artificialmente traz algo mais distinto e complexo do que o mau uso de uma ferramenta por maus atores ou alguns indivíduos. Por esse motivo, acreditamos que a IA não representa apenas um novo capítulo de um problema antigo, mas uma forma inédita de violência sexual, sem precedentes na sociedade.

A escala e a velocidade com que esses conteúdos são gerados e disseminados alteram a própria natureza do dano. É uma nova modalidade de violência digital que, pela primeira vez, pode provocar dissociação nas vítimas, que, ao verem essas imagens, embora saibam racionalmente que não são elas, vivenciam o estigma e sofrimento de uma experiência traumática causada pela violência sexual. O trauma é real, mesmo que a imagem seja fabricada.

O que os dados desta pesquisa mostram é que estamos diante de um ecossistema desenhado para facilitar o abuso. Nesse sentido:

1. **Vemos a nudificação de estudantes que tiveram suas fotos comuns, postadas para fins legítimos, transformadas em materiais pornográficos para humilhação em grupos de colegas**

2. **Vemos o uso de rostos de vítimas reais e conhecidas para alimentar novas gerações de abusos sintéticos**

3. **Vemos a comercialização ativa desse conteúdo em redes e fóruns onde criminosos celebram o fato de que fantasias agora podem ser feitas “sob encomenda”**

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

É alarmante a tentativa de justificar a produção desses conteúdos, conforme observado pela SaferNet Brasil em publicações nas redes sociais, sob a falsa premissa de aceitabilidade moral por serem artificiais. Isso é inaceitável. O verdadeiro dano não reside na autenticidade da imagem, mas sim na brutal desumanização e na violação da autonomia sobre seu próprio corpo e sua própria imagem, da sua privacidade e da sua dignidade sexual.

A facilidade com que qualquer indivíduo pode violar a imagem de um colega de escola, sem exigir perícia técnica, representa um risco estrutural grave. Isso impõe novos desafios para a proteção de direitos fundamentais, especialmente no Brasil, onde a violência sexual contra crianças e adolescentes atinge níveis críticos, com uma vítima a cada oito minutos, conforme dados do Fundo das Nações Unidas para a Infância (UNICEF).[1]

Este relatório, realizado com o apoio do fundo internacional *SafeOnline* e da *Tech Coalition,* coalizão das empresas de tecnologia para acabar com o abuso sexual infantil, oferece as evidências necessárias para podermos sair da reação pontual e avançar para estratégias coordenadas. Neste documento, propomos caminhos concretos, como a asfixia financeira de redes criminosas e o fortalecimento de protocolos nas escolas, fugindo de soluções meramente punitivas que ignoram a complexidade do problema.

A experiência de pesquisa neste relatório nos ensinou que, quando trabalhamos para tornar a internet segura para crianças e adolescentes, estamos elevando o padrão de segurança e integridade digital para todas as pessoas.

E É ESTE PROPÓSITO QUE MOTIVA A SAFERNET BRASIL HÁ 20 ANOS.

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

O que fazer se você for vítima de violência sexual mediada por IA?

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

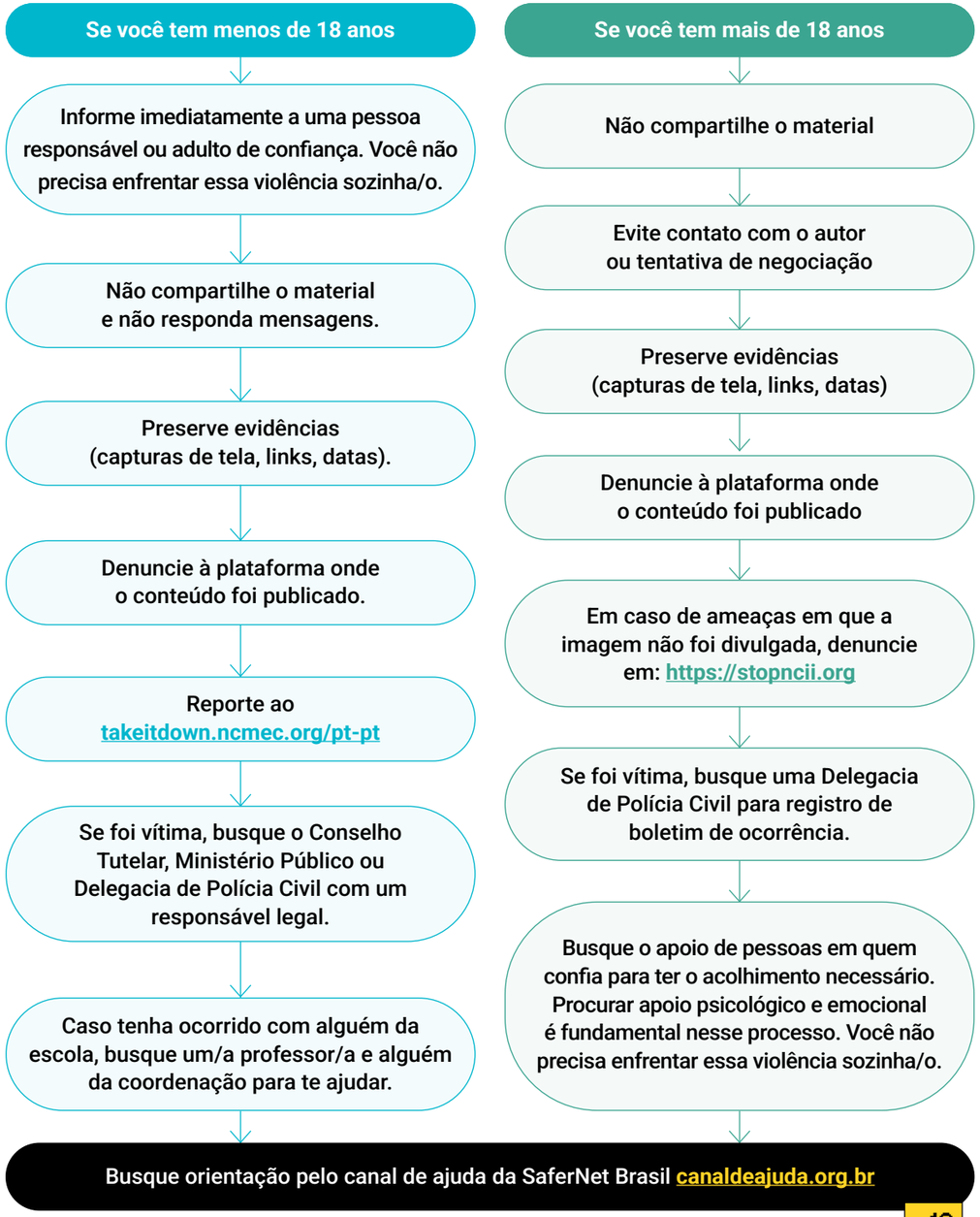
5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Muitas vítimas ainda enfrentam desinformação, insegurança e falta de orientação sobre como agir. A ausência de respostas institucionais claras, aliada à dificuldade de acesso a apoio especializado, torna o enfrentamento desse tipo de abuso ainda mais desafiador. Por isso, é fundamental reunir orientações práticas e acessíveis que ajudem vítimas e seus familiares a proteger seus direitos, buscar apoio e reduzir os danos causados por esse tipo de violência.



ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

Lista de siglas

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

AIGCSAM	AI-Generated Child Sexual Abuse Material (Material ou Conteúdo de Violência Sexual Contra Crianças e Adolescentes Gerado por IA, tradução livre)
ANPD	Autoridade Nacional de Proteção de Dados
CSAM	Child Sexual Abuse Material (Material ou Conteúdo de Violência Sexual Contra Crianças e Adolescentes)
ECA	Estatuto da Criança e do Adolescente
GANS	Generative Adversarial Networks (Redes Adversariais Generativas, tradução livre)
IA	Inteligência Artificial
LGPD	Lei Geral de Proteção de Dados
LORAS	Low-Rank Adaptations (Adaptações de Baixo Posto, tradução livre)
MCI	Marco Civil da Internet
MEC	Ministério da Educação
MJSP	Ministério da Justiça e Segurança Pública
ML	Machine Learning (Aprendizado de Máquina)
OCDE	Organização para a Cooperação e Desenvolvimento Econômico
ONU	Organização das Nações Unidas
PF	Polícia Federal
RLHF	Reinforcement Learning from Human Feedback (Aprendizado por Reforço com Feedback Humano)
SENACON	Secretaria Nacional do Consumidor
STF	Supremo Tribunal Federal
UNICEF	United Nations Children's Fund (Fundo das Nações Unidas para a Infância)
VAES	Variational Autoencoders (Autocodificadores Variacionais, tradução livre)

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

CAPÍTULO UM

Definições & tipologias

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

NESTE CAPÍTULO VOCÊ LERÁ UMA EXPLICAÇÃO SUCINTA DE COMO A IA GENERATIVA FUNCIONA, COMO ELA PODE GERAR MATERIAL DE EXPLORAÇÃO SEXUAL INFANTIL, OS DESAFIOS PARA MODERAR ISTO, QUAIS TIPOS DE VIOLÊNCIA MEDIADA POR IA A SAFERNET BRASIL MAPEOU NESTA PESQUISA E O QUE A LEGISLAÇÃO BRASILEIRA ABRANGE SOBRE ESTE PROBLEMA.

A inteligência artificial está transformando de forma profunda a natureza, o alcance e a escala da violência sexual on-line contra crianças e adolescentes. Práticas que antes exigiam conhecimento técnico especializado, equipamentos sofisticados e acesso a redes criminosas organizadas passaram a ser realizadas por qualquer pessoa com acesso à internet. À medida que os **modelos de IA** se tornam mais acessíveis, potentes e capazes de gerar imagens hiper-realistas, os danos decorrentes de seu uso indevido também se ampliam, afetando um número cada vez maior de vítimas.

Ferramentas que produzem imagens realistas a partir de simples descrições textuais, aplicativos que transformam fotografias comuns em conteúdo sexual sintético em poucos cliques, e comunidades on-line que ensinam, refinam e normalizam essas práticas formaram um ecossistema de violência digital sem precedentes históricos. Esse ambiente favorece a produção, a disseminação e a legitimação de conteúdos que violam gravemente os direitos de crianças e adolescentes.

Este capítulo se dedica a analisar o funcionamento da geração de imagens por IA, com foco no uso de tecnologias generativas para a criação de conteúdos de abuso e exploração sexual de crianças e adolescentes. São examinadas a natureza desses materiais artificiais, suas dinâmicas de circulação e o atual panorama jurídico que os envolve no Brasil.

Embora façamos um breve resumo do funcionamento da IA, é fundamental destacar que este documento não se destina à promoção de literacia digital aprofundada sobre o tema — a SaferNet Brasil dispõe de outros materiais específicos para essa finalidade. Nosso propósito, nesta seção, é estabelecer e definir os termos, tecnologias e as diversas tipologias associadas à IA no contexto da exploração sexual infantojuvenil.

MODELOS DE IA SÃO SISTEMAS MATEMÁTICOS TREINADOS COM GRANDES VOLUMES DE DADOS PARA IDENTIFICAR PADRÕES. EM VEZ DE “ENTENDER” A REALIDADE, ELES APRENDEM REGULARIDADES ESTATÍSTICAS E AS USAM PARA GERAR PREVISÕES, CLASSIFICAÇÕES OU CONTEÚDOS COMO TEXTO E IMAGENS. ELES SÃO CHAMADOS DE “MODELOS” PORQUE FUNCIONAM COMO REPRESENTAÇÕES SIMPLIFICADAS DO MUNDO: NÃO REPRODUZEM A REALIDADE, MAS UMA APROXIMAÇÃO BASEADA NOS PADRÕES APRENDIDOS. NO CASO DE MODELOS DE IMAGEM, POR EXEMPLO, ELES GERAM NOVAS IMAGENS A PARTIR DE PADRÕES OBSERVADOS EM MILHÕES DE EXEMPLOS, SEM COMPREENSÃO REAL DO CONTEÚDO PRODUZIDO.

1.1. O QUE É A INTELIGÊNCIA ARTIFICIAL?

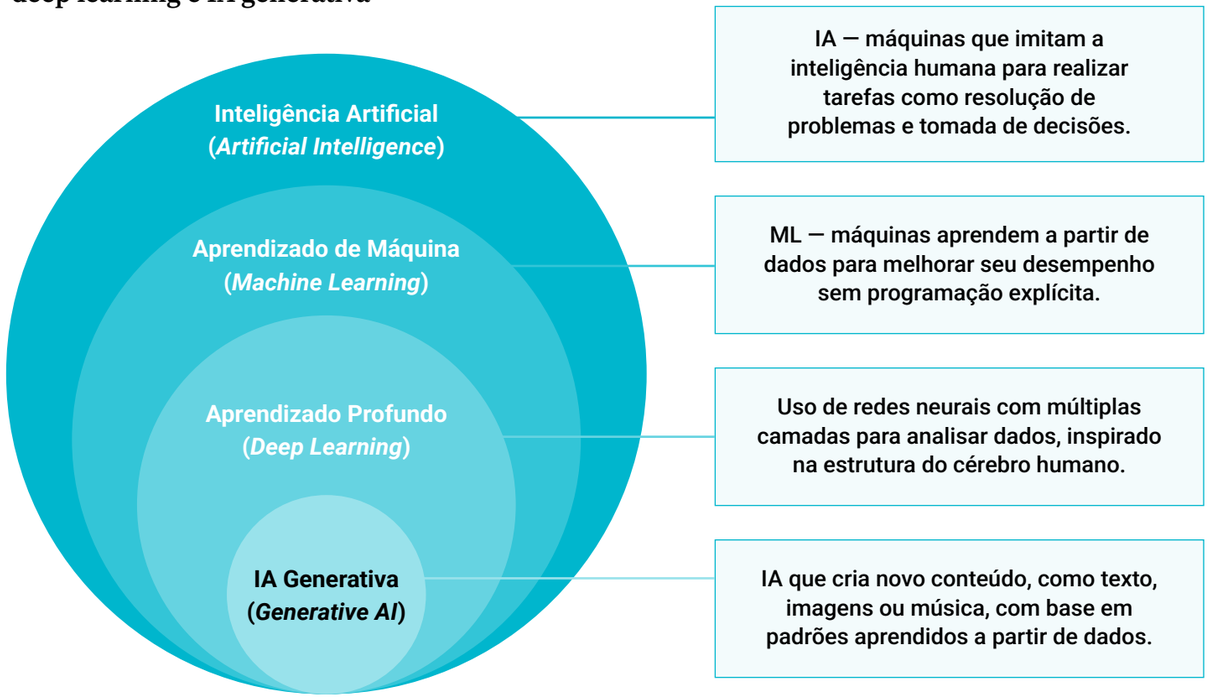
A IA já faz parte do cotidiano de milhões de pessoas — muitas vezes sem que percebam. Ela está presente nas músicas sugeridas em aplicativos de *streaming*, nos algoritmos que decidem o que aparece no *feed* das redes sociais e nos sistemas que organizam os resultados de buscas na internet. Dentro desse vasto campo tecnológico, destaca-se a IA generativa, um subconjunto de sistemas automatizados e especializados. Sua função é criar conteúdo sintético — como textos, imagens, vídeos e áudios — por meio da análise e do processamento de grandes volumes de dados.

Entre os inúmeros sistemas que usam a IA generativa, por sua vez, existe um subconjunto ainda mais específico que é o foco central deste relatório: os sistemas de geração de imagens, que são tecnologias capazes de criar representações visuais sintéticas e altamente realistas de pessoas, corpos e situações que nunca existiram.

A transição da IA como campo geral para a IA generativa e, em seguida, para a geração de imagens, representa mais do que uma mera classificação. Essa progressão é crucial porque cada etapa significa um salto no potencial de uso indevido: enquanto algoritmos de recomendação se limitam a organizar o conteúdo existente, os sistemas de geração de imagens têm a capacidade de criar conteúdo totalmente novo. Isso inclui a produção de representações realistas de abuso sexual que nunca ocorreram. Contudo, para entender como esses sistemas podem ser empregados em casos de violência sexual envolvendo crianças e adolescentes, é fundamental primeiro compreender seu funcionamento. O ponto de partida é, quase sempre, o mesmo: os dados.

FIGURA 1

Processo de IA, machine learning, deep learning e IA generativa



ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

1.1.1. COMO OS SISTEMAS DE IA APRENDEM?

A qualidade, a finalidade e a segurança de qualquer sistema de IA são determinadas pelos dados usados no seu treinamento. Isso significa que vieses, lacunas ou conteúdos problemáticos nas bases de treinamento inevitavelmente se refletem — e podem ser amplificados — nos resultados gerados pelo modelo. Um sistema aprende com aquilo que lhe é mostrado. Se o que lhe é mostrado contém violência, essa violência se torna parte do que ele é capaz de produzir.

No caso específico da geração de imagens, esse aprendizado funciona da seguinte forma: o modelo analisa milhões de imagens, identificando padrões visuais — disposição de elementos faciais, proporções corporais, comportamento da luz, texturas. O modelo não memoriza as imagens que viu. Ele extrai os princípios subjacentes que governam a aparência visual e, a partir deles, é capaz de gerar imagens completamente inéditas — sintéticas, mas visualmente convincentes.

Durante o treinamento, o sistema aprende a relacionar descrições (quando presentes) com características visuais correspondentes, construindo um mapeamento entre linguagem e imagem. Esse processo permite que, ao receber um comando textual, o modelo “projete” uma imagem coerente com os padrões que assimilou. Em muitos casos, isso envolve modelos de difusão ou arquiteturas generativas que começam com ruído aleatório e o refinam iterativamente até formar uma imagem estruturada.

É importante destacar que o modelo não reproduz imagens específicas do conjunto de treino. O que ele faz é internalizar os princípios que governam a aparência visual — geometria, cor, iluminação, contexto — e recombiná-los de maneiras novas. O resultado são imagens inéditas: sintéticas na origem, mas frequentemente convincentes do ponto de vista perceptivo, justamente porque seguem as mesmas regras estatísticas do mundo visual real.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

1.1.2. COMO OS SISTEMAS DE IA APRENDEM A REPRODUZIR CENAS DE EXPLORAÇÃO E ABUSO SEXUAL INFANTOJUVENIL?

Pesquisadores e especialistas em segurança digital convergem em um ponto central: sistemas de IA generativa produzem material ou conteúdo de violência sexual contra crianças e adolescentes por duas vias principais.[2]

A **primeira via** é a exposição a dados contaminados durante o seu treinamento. Quando um sistema de IA incorpora material ilegal em seu treinamento, o problema assume caráter estrutural. As informações passam a integrar os parâmetros internos do modelo e não podem ser removidas integralmente por simples filtragem posterior. Grandes bases de dados usadas no treinamento de modelos de larga escala são frequentemente construídas por raspagem automatizada da internet — sem filtros adequados, validação humana ou moderação prévia. Esse processo indiscriminado torna essas bases vetores críticos de contaminação.

Em dezembro de 2023, pesquisadores do Stanford Internet Observatory publicaram uma investigação identificando mais de 3.200 instâncias suspeitas de material ou conteúdo de violência sexual contra crianças e adolescentes no LAION-5B — base de dados utilizada no treinamento de algumas versões do modelo de geração de imagens Stable Diffusion (Stability AI) —, das quais 1.008 foram externamente validadas.[3] As imagens criminosas foram originalmente hospedadas em plataformas como Reddit, Twitter, Blogspot, WordPress e sites de conteúdo pornográfico adulto, que chegaram até a base de dados através de um raspador usado pela **LAION** para encontrar conteúdos que possam ser incorporados no conjunto de dados.[4]

Essa característica inerente ao código aberto, que prioriza a transparência e o acesso livre, inadvertidamente perpetua o risco. Uma vez que o código ou o modelo treinado é disponibilizado publicamente, é virtualmente impossível remover todas as cópias ou impedir seu uso e redistribuição em diversas plataformas e repositórios. Consequentemente, mesmo após uma ação corretiva inicial, como a remoção ou ajuste do conjunto de dados original (assumindo que o processo tenha sido realizado), o potencial danoso persiste, pois as versões pré-existentes e replicadas do modelo permanecem ativas, permitindo que indivíduos mal-intencionados continuem a fabricar *deepfakes* sexuais e outros conteúdos ilegais.

A **segunda via** decorre da exposição de modelos generativos a grandes volumes de pornografia adulta em seus conjuntos de dados de treinamento. Estudos indicam que essa exposição aumenta de forma mensurável a probabilidade de produção de representações sexualizadas de crianças e adolescentes.[5] Isso ocorre porque esses sistemas operam por correlação estatística entre padrões visuais e textuais — e não por critérios normativos, éticos ou jurídicos. Assim, conteúdo criminoso pode emergir como efeito direto da forma como os modelos organizam, recombinam e extrapolam os dados aprendidos, mesmo sem exemplos explícitos de abuso infantil nos dados originais.[6]

EMBORA O LAION-5B NÃO ARMAZENASSE IMAGENS DIRETAMENTE, MAS SIM URLS E METADADOS, SUA COLETA E USO GERAM RESPONSABILIDADE ÉTICA E JURÍDICA, ESPECIALMENTE AO APONTAR PARA CONTEÚDO ILEGAL. O RELATÓRIO DE STANFORD RECOMENDOU A DESCONTINUAÇÃO DE MODELOS TREINADOS COM ELE. O PROBLEMA SE AGRAVA PORQUE ESSES MODELOS DE CÓDIGO ABERTO SÃO DE DIFÍCIL CONTROLE E RETIRADA APÓS DISTRIBUIÇÃO, AO CONTRÁRIO DE BANCOS DE DADOS CENTRALIZADOS.

ÍNDICE

1.
DEFINIÇÕES & TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3.
PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4.
ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5.
RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6.
RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

1.1.3. AS ARQUITETURAS TECNOLÓGICAS QUE VIABILIZAM A GERAÇÃO DE IMAGENS

Compreender por que esses sistemas produzem conteúdo ilícito é necessário, mas insuficiente. É preciso também entender como funcionam as tecnologias que tornam isso possível — porque diferentes arquiteturas apresentam diferentes perfis de risco e diferentes possibilidades de intervenção. Atualmente, existem três arquiteturas principais que são utilizadas para geração de imagens por IA.

TECNOLOGIA	COMO FUNCIONA
GANs **(Redes Adversariais Generativas)**	Baseiam-se na competição entre dois sistemas: um “gerador” que cria imagens e um “discriminador” que avalia se essas imagens são autênticas ou falsas. O ciclo se repete até que as imagens produzidas se tornem indistinguíveis de fotos reais.
Modelos de difusão (*Diffusion Models*)	Aprendem a reconstituir imagens a partir de um estado de ruído. Durante o treinamento, as imagens originais são progressivamente corrompidas, e o modelo é treinado para reverter esse processo passo a passo, muitas vezes com base em comandos textuais.
VAEs **(Autoencoders Variacionais)**	Compactam as imagens em representações mais abstratas, conhecidas como “espaço latente”, e depois as reconstroem. Essa abordagem permite a manipulação de características ou atributos específicos da imagem sem afetar o restante.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS
2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS
3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO
4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA
5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA
6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES
7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

1.1.4. OS DESAFIOS DOS MODELOS DE CÓDIGO ABERTO: DISTRIBUIÇÃO, CONTROLE E ESCALA DO DANO

A forma como os sistemas de IA são disponibilizados determina as estratégias possíveis de prevenção, mitigação e resposta ao seu uso indevido. Atualmente, o ecossistema tecnológico divide-se em duas abordagens principais de distribuição: os modelos proprietários e os modelos de código aberto.

Modelos proprietários, como GPT (OpenAI) ou Gemini (Google), operam como sistemas controlados: os usuários interagem por interfaces restritas, sem acesso aos pesos do modelo ou à arquitetura interna. Isso permite que a empresa implemente salvaguardas centralizadas, como filtros de conteúdo, moderação automatizada e monitoramento em tempo real. Essas barreiras técnicas possibilitam atualizações instantâneas e o bloqueio de comportamentos abusivos assim que novas vulnerabilidades são identificadas.

Os modelos de código aberto como Stable Diffusion (Stability AI), FLUX (Black Forest Labs) ou Qwen (Alibaba) podem ser livremente baixados, modificados e redistribuídos, o que impede a adoção de mecanismos de controle posteriores. Essa abertura é um motor de inovação, mas também permite que o modelo seja baixado e hospedado na infraestrutura privada do próprio usuário. Sem a necessidade de uma conexão com os servidores do desenvolvedor original, o sistema opera sem qualquer visibilidade externa, o que inviabiliza o rastreamento, o monitoramento ou a intervenção remota para interromper usos nocivos.

No entanto, a diferença técnica impõe um desafio estrutural para a segurança infantil. Quando um modelo aberto é liberado com falhas no treinamento ou sem filtros nativos, ele permanece operacional indefinidamente: não há possibilidade de desligamento centralizado. Mesmo que versões corrigidas sejam lançadas, cópias já distribuídas continuam acessíveis em servidores privados, fóruns especializados e ambientes da *dark web*.

Os trechos a seguir foram extraídos de um fórum dedicado à criação de materiais artificiais de abuso sexual infantil, analisados na seção 4.2.1 deste relatório. Eles ilustram como esses usuários elogiam, comparam e recomendam abertamente diferentes modelos — destacando precisamente a ausência de filtros como vantagem operacional:

> Já tenho experiência na criação de arte com inteligência artificial, utilizando ferramentas e modelos da Stable Diffusion e da Flux, e também criei LoRAs para uso pessoal.

> Novo modelo lançado! Da equipe Alibaba. Supostamente pequeno, rápido, sem censura, de alta qualidade.

> Eu diria que o WAN 2.2 torna isso muito mais fácil do que trabalhar com imagens usando o SDXL. Com o SDXL, a taxa de sucesso para seguir o *prompt* com esses temas gira em torno de 10%. Com o WAN 2.2, eu diria que ultrapassa facilmente 60%, fornecendo claramente a aparência que você procura.

> Estou curioso, é muito difícil usar o conjunto de dados de um modelo da Flux e criar um SDXL? Eles podem ser convertidos?

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Eu uso alguns modelos SDXL/Pony que misturei a partir de fontes disponíveis gratuitamente para o meu modelo principal.

Então, o que você acha dos novos modelos Qwen-Image e Flux Kera [sic]? Existem maneiras melhores de gerar CP [*child porn* ou material ou conteúdo de violência sexual contra crianças e adolescentes] usando esses dois modelos sem sacrificar a excelente qualidade de imagem dos modelos originais?

A facilidade e o custo praticamente nulo da criação de material ou conteúdo de violência sexual contra crianças e adolescentes artificial e individualmente direcionado se tornaram uma consequência direta da existência de modelos fundacionais abertos e aplicações derivadas. Isso resultou em um aumento significativo tanto na produção quanto na distribuição desse tipo de conteúdo. Enquanto modelos proprietários oferecem a possibilidade de atualização centralizada ao detectar vulnerabilidades (embora sejam suscetíveis a *jailbreaking*), os modelos de código aberto, uma vez lançados, perdem qualquer tipo de controle posterior.[7]

Uma possível resposta a essa dinâmica seria o desaprendizado de máquina (*machine unlearning*), método que busca remover seletivamente informações já incorporadas aos parâmetros de um modelo treinado. A literatura científica recente, porém, demonstra que essa remoção é tecnicamente complexa e pouco confiável. Muitos métodos baseados em ajuste fino suprimem comportamentos de forma superficial, sem erradicar os padrões subjacentes. Avaliações de técnicas de última geração demonstraram que abordagens adversariais conseguiram recuperar conhecimento supostamente removido em mais de 40% dos casos testados.[8]

RECOMENDAÇÕES ESPECÍFICAS PARA AS EMPRESAS DE IA GENERATIVA SOBRE COMO SOLUCIONAR ESTE PROBLEMA ESTÃO DISPONÍVEIS NA **SEÇÃO 6.3** DESTE RELATÓRIO.

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

1.2. O QUE SÃO MATERIAIS ARTIFICIAIS DE ABUSO OU EXPLORAÇÃO SEXUAL INFANTOJUVENIL?

Com esse panorama tecnológico estabelecido, é possível agora definir com precisão o objeto central deste relatório.

O que em português denominamos "material ou conteúdo de violência sexual contra crianças e adolescentes" corresponde, em inglês, à expressão *child sexual abuse material*, na sigla *CSAM*. Quando esse conteúdo é produzido por IA generativa, utiliza-se a sigla *AIGCSAM (AI-generated child sexual abuse material)*. Entretanto, não há padronização consolidada desses termos ou de siglas em português, o que reforça a importância de definições claras no contexto nacional.

É fundamental ressaltar que o critério de artificialidade inclui representações que tenham sido integralmente geradas, substancialmente alteradas ou combinadas por meio de tecnologias de IA generativa. Assim, mesmo conteúdos que não envolvam registros reais podem configurar grave violação de direitos quando simulam ou induzem à sexualização de crianças e adolescentes.

Alguns exemplos dos conteúdos que estamos falando são, mas não se limitam a:

- **Montagens que inserem rostos infantis, reais ou sintéticos, em corpos ou cenários sexualizados;**
- **Nudez sintética criada a partir de imagens cotidianas de crianças vestidas, por meio de aplicativos de nudificação;**
- **Vídeos *deepfakes* que manipulam imagens faciais de crianças reais para simular situações de abuso inexistentes;**
- **Imagens híbridas que combinam características faciais de múltiplas crianças para gerar rostos sintéticos inseridos em contextos sexualizados.**

A SAFERNET BRASIL ADOTA UMA DEFINIÇÃO ABRANGENTE: TODO E QUALQUER CONTEÚDO VISUAL — IMAGENS ESTÁTICAS, VÍDEOS OU OUTROS FORMATOS MULTIMÍDIA — QUE, DE MANEIRA REALISTA OU NÃO, REPRESENTE CRIANÇAS OU ADOLESCENTES EM SITUAÇÕES DE ABUSO, EXPLORAÇÃO OU CUNHO SEXUAL, E QUE TENHA SIDO INTEGRALMENTE GERADO, SUBSTANCIALMENTE ALTERADO OU COMBINADO POR MEIO DE TECNOLOGIAS DE IA GENERATIVA.

ÍNDICE

1.
DEFINIÇÕES & TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3.
PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4.
ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5.
RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6.
RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Um ponto crucial desta definição adotada pela SaferNet Brasil é que a classificação de material ou conteúdo de violência sexual contra crianças e adolescentes se mantém mesmo quando as representações artificiais não correspondem a pessoas reais identificáveis. O foco central está na natureza do conteúdo produzido e no impacto de sua circulação para a perpetuação da cultura de abuso, objetificação e sexualização dos corpos de crianças e adolescentes.

Ainda que não haja uma vítima diretamente identificável em determinados casos (por exemplo, quando uma criança que não existe no mundo real é fabricada por uma IA), a produção e a disseminação desse material contribuem para a normalização da violência sexual e ampliam ambientes digitais permissivos a práticas abusivas. Além disso, o consumo desse tipo de conteúdo potencializa o risco de cometimento de abusos sexuais contra crianças e adolescentes fora da internet.

1.2.1. TIPOS DE VIOLÊNCIA SEXUAL INFANTOJUVENIL COM IA

Até a publicação desta pesquisa, foram identificados cinco principais tipos de violência sexual mediada por IA contra crianças ou adolescentes. Essas práticas envolvem diferentes formatos de produção e manipulação de conteúdo, utilizando tecnologias automatizadas para criar, modificar ou simular materiais de cunho sexual.

TEXTO EM CONVERSA

IA Generativa são utilizados para gerar interações conversacionais que simulam diálogos de natureza sexual envolvendo crianças ou adolescentes, criando a aparência de comunicação real entre indivíduos.

TEXTO EM TEXTO

Produção de narrativas, descrições ou roteiros de caráter sexual que têm crianças ou adolescentes como foco principal, explorando linguisticamente situações de abuso ou exploração.

TEXTO EM IMAGEM

Descrições textuais são transformadas em representações visuais que simulam situações de abuso ou exploração sexual infantil, por meio de ferramentas de geração automática de imagens.

IMAGEM EXPLÍCITA EM IMAGEM

Manipulação, alteração ou recombinação de imagens preexistentes que já contêm nudez ou cenas sexualizadas, com o objetivo de gerar novos conteúdos ilícitos.

IMAGEM NÃO EXPLÍCITA EM IMAGEM

Imagens originalmente não sexualizadas são modificadas por sistemas de IA para criar representações falsas de abuso ou exploração sexual infantil, inserindo elementos que não estavam presentes no material original.

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

1.2.2. A DIMENSÃO DE GÊNERO NA VIOLÊNCIA SEXUAL MEDIADA POR IA CONTRA CRIANÇAS E ADOLESCENTES

A violência sexual mediada por IA não ocorre em um vácuo tecnológico; ela é a expressão de um problema estrutural que reflete as disparidades históricas nas relações de gênero. O mapeamento realizado pela SaferNet Brasil entre agosto de 2025 e fevereiro de 2026 é categórico: nos 17 episódios identificados em escolas brasileiras, 100% das vítimas eram do gênero feminino, alunas e professoras; o mapeamento será discutido no próximo capítulo em detalhes. Em contrapartida, a totalidade dos perpetradores mapeados pela pesquisa eram meninos.

Essa assimetria revela que as ferramentas de IA generativa e de nudificação estão sendo instrumentalizadas para reforçar padrões de objetificação e controle que atingem mulheres e meninas de forma desproporcional. Trata-se de um ecossistema que se aproveita da facilidade técnica para produzir danos reais. Estudos e análises mostram que conteúdos íntimos não consensuais — incluindo *deepfake*s destinados a sexualizar pessoas — contribuem para um ambiente no qual a violência sexual é trivializada e normalizada.

A criação e disseminação de material artificial de abuso e exploração sexual infantil não são apenas falhas técnicas de segurança, mas graves violações de direitos, ancoradas numa estrutura que perpetua relações desiguais de poder. Ao simular a nudez de uma colega ou educadora, o agressor busca o silenciamento e a humilhação da vítima, utilizando a escala e a velocidade da rede para garantir que o dano seja persistente e devastador. Reconhecer este fenômeno como uma modalidade de violência de gênero é o primeiro passo para desenvolver estratégias de enfrentamento que protejam meninas e, consequentemente, elevem o padrão de segurança digital para toda a sociedade.

100%

DAS VÍTIMAS ERAM ALUNAS E PROFESSORAS NOS 17 EPISÓDIOS IDENTIFICADOS EM ESCOLAS BRASILEIRAS

A distinção entre *deepfakes* envolvendo adultos e aqueles que retratam crianças e adolescentes é essencial tanto do ponto de vista legal quanto ético.

1.3.1. NO CASO DE VÍTIMAS COM MAIS DE 18 ANOS

Até recentemente, a base para punir a divulgação de conteúdos íntimos não consensuais de adultos era o Artigo 218-C do Código Penal, introduzido em 2018. No entanto, este artigo apresentava uma limitação técnica: ele punia a divulgação de cenas de sexo ou nudez existentes, mas não tipificava especificamente os casos em que a cena era simulada, modificada ou fabricada.

Esse vácuo jurídico permitia que agressores alegassem a inexistência de crime por não se tratar de uma imagem "real", ignorando o impacto da violação. O avanço legislativo veio com a Lei n.° 15.123/2025, que atualizou o Código Penal para enfrentar diretamente o uso ilícito de tecnologias de IA em dois eixos:

Violência Psicológica (Art. 147-B)	Estabelecida uma causa de aumento de pena (majorada) de metade se o crime de violência psicológica contra a mulher for praticado com o uso de inteligência artificial ou qualquer recurso tecnológico que altere a imagem ou o som da vítima.
Divulgação de Cena de Sexo (Art. 218-C)	O texto foi reformulado para incluir explicitamente as condutas de "realizar montagem ou modificação" por qualquer meio, inclusive mediante o uso de IA. Além de ampliar o tipo penal para abranger o conteúdo sintético, a lei aumentou a pena combinada para 2 a 6 anos de reclusão.

Essa mudança é fundamental porque retira a discussão sobre a "veracidade" da imagem criada por IA generativa e foca no ato criminoso de fabricar e expor a intimidade de alguém. A partir disso, a legislação brasileira passa a reconhecer que a escala e a facilidade de criação proporcionadas pela IA não são apenas novos meios de cometer crimes antigos, mas elementos que exigem uma resposta penal mais severa e específica para proteger a integridade das vítimas.

Por se tratar de uma mudança legislativa muito recente, ainda não é possível mapear evidências concretas sobre como esses novos tipos penais estão sendo interpretados pelos tribunais, o que torna o monitoramento dos primeiros casos que chegam ao Judiciário um passo essencial para avaliar a eficácia da nova lei, sendo esta uma recomendação para estudos futuros.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

1.3.2. NO CASO DE VÍTIMAS COM MENOS DE 18 ANOS

Quando a vítima tem menos de 18 anos, o conteúdo ultrapassa o âmbito de mera violação de direitos de imagem e constitui material de violência sexual contra criança ou adolescente, crime tipificado no Estatuto da Criança e do Adolescente (ECA). Mesmo que a imagem ou vídeo seja integralmente artificial, sem envolver uma criança real durante sua produção, a legislação brasileira e os tratados internacionais reconhecem que a produção, o armazenamento e a distribuição desses materiais configuram formas de violência sexual contra crianças e adolescentes. A legislação passou a considerar como violência sexual qualquer representação, real ou fictícia, por qualquer meio — inclusive produzida, manipulada ou gerada por IA generativa — que retrate atividade sexual, nudez com finalidade sexual ou pose de conotação sexual envolvendo criança ou adolescente.

Com a sanção da Lei n.º 15.487, de 6 de agosto de 2026, que alterou o Estatuto da Criança e do Adolescente, esse tratamento foi significativamente reforçado. A legislação passou a adotar o termo "violência sexual contra criança ou adolescente" em substituição à antiga expressão "cena de sexo explícito ou pornográfica", ampliando o alcance da proteção. O novo conceito abrange qualquer representação, real ou fictícia, por qualquer meio — inclusive imagens produzidas, manipuladas ou geradas por IA generativa —, dispensando a existência de uma criança real, o contato físico ou a exposição explícita para a caracterização do crime. Além disso, houve aumento das penas, a classificação de vários desses crimes como hediondos e a criação de agravantes específicos para o uso de IA, *deepfake*, perfis falsos e técnicas de anonimização digital, adequando a norma penal à realidade dos crimes praticados no ambiente digital.

Tipificação (ECA)	Conduta	Pena (reclusão) + multa	Alcance sobre conteúdo fictício/IA e agravamento
Art. 241-E	Base conceitual — o que é "material de violência sexual"	—	"Criança ou adolescente real ou fictícia, inclusive gerada/manipulada por IA generativa, em atividade sexual, nudez com finalidade sexual ou pose de conotação sexual.
Art. 240	Produzir, reproduzir, dirigir, fotografar, filmar ou registrar o material; financiar, recrutar, facilitar ou intermediar; exibir ou transmitir ao vivo	**4 a 10 anos**	Aplica-se a material real ou fictício (IA) por força do art. 241-E; crime hediondo.
Art. 241	Vender ou expor à venda	**4 a 10 anos**	Perda de bens/valores ao Fundo dos Direitos da Criança e do Adolescente; +1/3 se por internet/redes/TIC; hediondo; alcança material gerado por IA.
Art. 241-A	Distribuir, publicar, divulgar, compartilhar ou transmitir; criar, administrar, hospedar ou moderar ambiente digital para o material	**4 a 10 anos**	+1/3 se difundido em mais de uma plataforma; hediondo; alcança material gerado por IA.
Art. 241-B	Adquirir, possuir, armazenar ou solicitar; acessar ou visualizar com finalidade sexual	**3 a 6 anos**	Redução de 1/6 a 1/3 (pequena quantidade); hediondo; alcança material gerado por IA.
Art. 241-C	Criar, montar, adulterar ou modificar imagem/vídeo por qualquer recurso tecnológico (montagem, deepfake, IA) para simular participação de criança/adolescente	**3 a 5 anos**	"Tipo penal específico para simulação / deepfake / IA.
Art. 241-D	Aliciar, assediar, instigar ou constranger menor de 14 anos, por qualquer meio, para ato libidinoso ou obtenção de imagens sexuais	**3 a 5 anos**	Agravante de 1/3 a 2/3: IA/deepfake/filtros, perfis falsos/anonimização, apps/redes/jogos, promessa de vantagem, relação de confiança/autoridade. Hediondo.
Art. 226-A	Uso de anonimização, mascaramento ou alteração de IP e de identificadores (ex: modulador de proxy) para dificultar a identificação	**Agravante**	"Aumento de 1/3 a 2/3 nos crimes do Capítulo (art. 240 a 244-A)

ÍNDICE

1.
DEFINIÇÕES & TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3.
PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4.
ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5.
RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6.
RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

1.3.3. RESPONSABILIDADE DE PLATAFORMAS, PROVEDORES DE CONTEÚDO E EMPRESAS DE HOSPEDAGEM

O Marco Civil da Internet (Lei n.º 12.965/2014) também se aplica. Seu artigo 21 estabelece que provedores de aplicação têm a obrigação de remover conteúdos de nudez ou atos sexuais privados e não consentidos, mediante notificação da vítima ou de seu representante legal, independentemente de ordem judicial, sob pena de responsabilização.

Importante destacar também que, em 2025, o Supremo Tribunal Federal (STF) conferiu uma interpretação mais abrangente ao regime de responsabilidade das plataformas, estabelecendo que o dever de remoção imediata — independentemente de ordem judicial e mediante apenas notificação extrajudicial — não se limita à nudez não consentida prevista no artigo 21 no Marco Civil. A Corte fixou um rol taxativo (lista fechada de itens previstos em lei, sem possibilidade de ampliação por interpretação) de outras violações que impõem a responsabilização civil da empresa caso o conteúdo não seja indisponibilizado prontamente. Esse rol inclui a divulgação de material ou conteúdo de violência sexual contra crianças e adolescentes, conteúdos que configurem crimes contra o Estado Democrático de Direito, discursos de ódio e atos de terrorismo. Com isso, o judiciário consolidou a tese de que a proteção da dignidade humana e a segurança de grupos vulneráveis prevalecem sobre a necessidade de intervenção judicial prévia em casos de flagrante ilegalidade.

Mais recentemente, a Lei n.º 15.211/2025, conhecida popularmente como ECA Digital, reforça a responsabilidade das empresas ao estabelecerem o dever de cuidado. Esse arcabouço legal exige que as plataformas adotem medidas proativas e preventivas para monitorar e impedir a disseminação de conteúdos que violem direitos de crianças e adolescentes. Ao contrário do modelo reativo do Marco Civil, no ECA Digital o dever de cuidado impõe a implementação de medidas de segurança e mecanismos de detecção que mitiguem riscos antes mesmo de qualquer notificação, abrindo espaço para a responsabilização das empresas pela omissão na proteção e garantia dos direitos das crianças e adolescentes com base na legislação brasileira.

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

CAPÍTULO DOIS

Violência sexual mediada por IA em escolas brasileiras

ÍNDICE

1.
DEFINIÇÕES & TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3.
PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4.
ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5.
RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6.
RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

e a seção anterior descreveu os mecanismos técnicos pelos quais sistemas de IA produzem conteúdo de abuso sexual infantojuvenil, esta seção documenta onde esse fenômeno aterrissa na realidade brasileira: dentro das escolas, entre adolescentes, com rostos conhecidos e consequências devastadoras. Os *deepfake*s sexuais não são uma ameaça abstrata ou futura — são uma forma de violência sexual já em curso, praticada por alunos contra colegas e professoras, com ferramentas gratuitas, acessíveis e disponíveis a qualquer indivíduo conectado à internet.

Essa dinâmica é agravada pelo contexto de exposição digital cada vez mais precoce. Dados recentes indicam que cerca de 7 em cada 10 estudantes do ensino médio já utilizam ferramentas de IA generativa em atividades escolares.[9] A isso se soma uma oferta crescente de aplicativos de nudificação e *bots* em plataformas de mensagens — muitos gratuitos ou de baixíssimo custo —, que eliminam qualquer barreira técnica de entrada.

O cenário de 17 casos, com quatro vítimas para cada agressor, ilustra a materialização de um problema que já não é mais hipotético no ambiente escolar brasileiro. Entre agosto de 2025 e fevereiro de 2026, a SaferNet Brasil realizou o monitoramento e a catalogação de incidentes de *deepfake*s sexuais e outras formas de uso de IA generativa para a prática de violência sexual. O foco da coleta de dados se concentrou em ocorrências protagonizadas por estudantes em instituições de ensino, tanto públicas quanto privadas.

A ausência de barreiras técnicas, associadas a outros fatores, como interações de risco sem supervisão, torna cada vez mais difusa a fronteira percebida pelos adolescentes entre brincadeira, violência simbólica e conduta criminosa.

NESSE CONTEXTO, A FALTA DE UM LETRAMENTO DIGITAL ADEQUADO NAS ESCOLAS AMPLIFICA A VULNERABILIDADE DOS JOVENS, ESPECIALMENTE NO QUE DIZ RESPEITO À CRIAÇÃO E AO COMPARTILHAMENTO DE DEEPFAKES SEXUAIS.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA



COMO FIZEMOS ESTE MAPEAMENTO

A metodologia desta análise está descrita logo após as Considerações Finais, na seção Sobre esta pesquisa. Em síntese, a SaferNet Brasil coletou informações por meio de um formulário permanente para recebimento de relatos e casos e de monitoramento ativo de notícias sobre casos de *deepfake*s sexuais em escolas.



CORREÇÃO SOBRE NÚMERO DE ESTADOS

Em prévias da divulgação do mapeamento, a SaferNet Brasil informou que os casos haviam ocorrido em 10 estados brasileiros, incluindo o Distrito Federal. Em comunicações posteriores, a organização atualizou o número para 11 estados. O número está correto, mas houve confusão sobre quais unidades da federação registraram as ocorrências. Os estados corretos são: AL, BA, CE, MG, MT, PA, PB, PE, RJ, RS e SP.



PRINCIPAIS RESULTADOS

Em sete meses (ago/2025 a fev/2026), o mapeamento da SaferNet Brasil identificou 17 casos com 61 infratores (todos homens, da mesma faixa etária) e 222 vítimas (todas mulheres, estudantes de 12 a 17 anos e, por vezes, professoras). A consistência desses dados revela a natureza estrutural do gênero nesta forma de violência.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

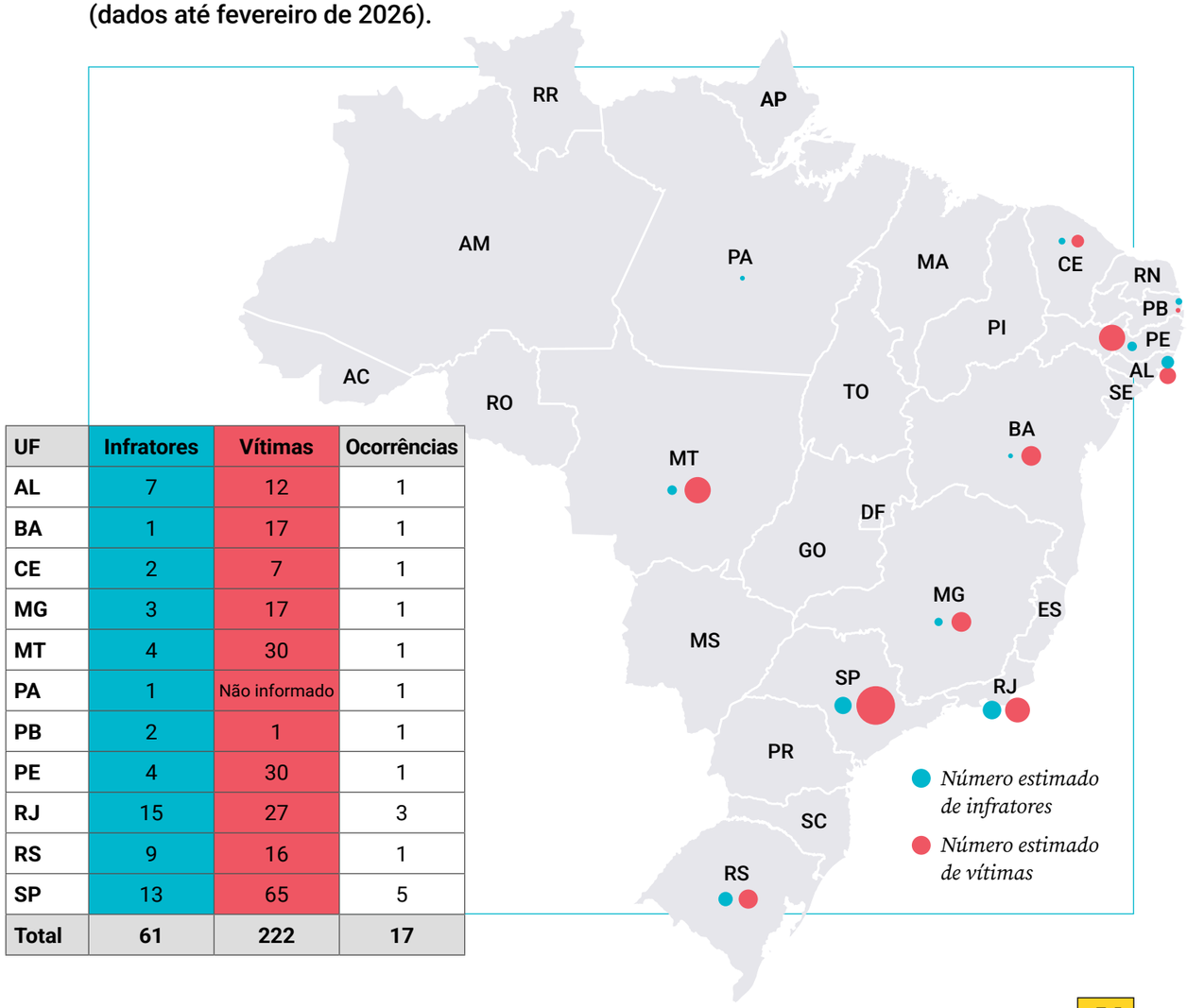
Ao longo de sete meses, entre agosto de 2025 e fevereiro de 2026, o mapeamento realizado pela SaferNet Brasil identificou 17 episódios envolvendo 61 infratores e 222 vítimas. As vítimas eram, sem exceção, do gênero feminino — estudantes na faixa dos 12 aos 17 anos e, em alguns casos, professoras. Os infratores eram, na totalidade, do gênero masculino, da mesma faixa etária. A consistência desses dois traços ao longo de todos os casos registrados não é coincidência: ela revela a natureza estruturalmente de gênero dessa violência.

Os casos foram registrados em onze estados brasileiros: Alagoas, Bahia, Ceará, Minas Gerais, Mato Grosso, Pará, Paraíba, Pernambuco, Rio de Janeiro, Rio Grande do Sul e São Paulo. Essa dispersão geográfica, por si só, evidencia que o fenômeno não é regional nem circunstancial — é nacional e sistêmico.

FIGURA 2

Mapeamento da SaferNet Brasil sobre casos de *deepfakes* sexuais em escolas brasileiras (dados até fevereiro de 2026).

UF	Infratores	Vítimas	Ocorrências
AL	7	12	1
BA	1	17	1
CE	2	7	1
MG	3	17	1
MT	4	30	1
PA	1	Não informado	1
PB	2	1	1
PE	4	30	1
RJ	15	27	3
RS	9	16	1
SP	13	65	5
Total	61	222	17



ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

O mapeamento encontrou ocorrências distribuídas entre escolas públicas e particulares, em capitais e no interior. Esse dado desfaz um equívoco comum: o de que esse tipo de violência seria um problema restrito a determinado perfil de instituição ou região. Quando o fenômeno atravessa redes de ensino, faixas de renda e contextos urbanos de forma tão abrangente, o que se evidencia é que seus fatores determinantes não são a precariedade ou o privilégio, mas elementos que estão presentes em praticamente todos os ambientes escolares — acesso a smartphones, dinâmicas de grupo entre pares adolescentes e ausência de mediação adulta qualificada sobre o uso de tecnologia.

Rio de Janeiro e São Paulo concentram um número maior de casos, o que pode ser atribuído ao porte populacional dessas regiões e, possivelmente, à maior capilaridade da cobertura jornalística local. Em termos absolutos, São Paulo lidera com 65 vítimas estimadas, seguido por Mato Grosso e Pernambuco, com 30 cada, e Rio de Janeiro, com 27.

FIGURA 3

Proporção de vítimas por infrator nos casos de *deepfakes* mapeados pela SaferNet Brasil.



ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Todas as vítimas identificadas são do gênero feminino, com idades entre 12 e 17 anos. Em alguns casos, professoras também foram vitimizadas — o que acrescenta uma dimensão de gravidade ao demonstrar que a violência pode extrapolar as relações entre pares e comprometer a própria estrutura de autoridade e segurança do ambiente escolar.

ESSA TENDÊNCIA ESPECÍFICA SERÁ APROFUNDADA NA **SEÇÃO 2.3**.

Os números de envolvidos nos incidentes revelam uma assimetria alarmante: para cada infrator identificado, há em média quase quatro vítimas. Esse dado sugere que a produção e o compartilhamento de conteúdos não seguem uma lógica individualizada, mas práticas coletivas ou seriais, nas quais um mesmo grupo de estudantes cria e circula imagens envolvendo múltiplas colegas. Esse padrão se reproduz frequentemente em canais fechados de mensagens, onde a violência é normalizada, celebrada ou reenquadrada como uma "brincadeira entre jovens" — uma minimização que, como discutido nas seções anteriores, integra os próprios mecanismos de perpetuação do abuso.

Os infratores identificados são todos do gênero masculino, com idades estimadas entre 12 e 17 anos — embora essa informação seja divulgada de forma inconsistente pela imprensa, que frequentemente omite ou generaliza dados sobre os perpetradores. No contexto de escolas privadas, essa omissão pode ser explicada, ao menos em parte, pela tentativa institucional de preservar a reputação do estabelecimento.

O perfil socioeconômico dos infratores identificados destoa do imaginário historicamente associado aos adolescentes em conflito com a lei no Brasil. Enquanto infrações como roubo, tráfico de drogas e crimes patrimoniais estiveram, por décadas, associadas a jovens em situação de vulnerabilidade socioeconômica, promotores e juízes da infância têm observado uma mudança de perfil nos crimes cometidos por meio da internet — incluindo abuso sexual digital e discurso de ódio. Nesses casos, os infratores frequentemente pertencem a estratos de maior renda e escolaridade, o que desafia tanto as representações sociais sobre o cometimento de atos infracionais quanto os mecanismos de responsabilização disponíveis.

A percepção de impunidade, nesse contexto, não emerge apenas da distância tecnológica entre agressor e vítima, mas é reforçada por privilégios concretos de classe e raça que os próprios infratores reconhecem e instrumentalizam. A tecnologia de IA, nesses casos, é mobilizada como ferramenta de exercício de poder e controle sobre o corpo e a imagem de mulheres e meninas — e a crença na impunidade integra, desde o início, o cálculo da ação. Esse cenário demonstra que o combate a esse tipo de violência exige uma abordagem que vá além da resposta tecnológica ou jurídica pontual. É necessário enfrentar a cultura de normalização do abuso e fortalecer os mecanismos de responsabilização de forma equânime, independentemente do perfil socioeconômico e racial do infrator.

O conjunto de casos mapeados, analisado em sua totalidade, permite identificar três tendências que se repetem de forma consistente e que merecem atenção específica: a vitimização exclusivamente feminina, a extensão da violência às professoras e a emergência de um modelo de comercialização das imagens. Cada uma dessas tendências aponta para dimensões distintas do problema — e todas elas agravam o diagnóstico inicial.

2.2.1. VITIMIZAÇÃO EXCLUSIVAMENTE FEMININA: VIOLÊNCIA DE GÊNERO COMO ESTRUTURA, NÃO COMO COINCIDÊNCIA

Em nenhum dos 17 casos mapeados a vítima era do gênero masculino. Essa unanimidade não é um dado neutro — é, em si mesma, uma das descobertas centrais do mapeamento. Ela indica que a produção de *deepfake*s sexuais em ambiente escolar não é uma forma genérica de *bullying* tecnológico, mas uma expressão específica de violência de gênero: uma prática direcionada sistematicamente contra mulheres e meninas, exercida predominantemente por meninos e jovens do sexo masculino.

Esse padrão é consistente com o que a literatura internacional descreve sobre abuso sexual baseado em imagem de forma mais ampla: a grande maioria das vítimas é feminina, e a motivação central não é simplesmente causar dano, mas exercer controle sobre o corpo e a sexualidade de mulheres por meio de sua imagem. No contexto escolar, esse controle é exercido dentro de relações de proximidade — entre colegas que se conhecem, que compartilham os mesmos espaços — o que torna a violência simultaneamente mais íntima e mais difícil de denunciar. A vítima não apenas tem sua imagem violada: ela precisa continuar convivendo com os perpetradores no dia seguinte.

2.2.2. PROFESSORAS MULHERES TORNANDO-SE VÍTIMAS DE ALUNOS MENINOS

Quando educadoras são sexualizadas e expostas por seus próprios alunos, algo mais amplo do que uma agressão individual está em curso. A autoridade pedagógica é corroída, o ambiente escolar torna-se hostil para as mulheres que o sustentam e a hierarquia que deveria proteger estudantes é invertida em uma dinâmica de abuso. A escola deixa de ser apenas o cenário onde a violência ocorre e passa a ser, ela mesma, desestabilizada por essa forma de violência tecnológica.

Os casos registrados ilustram essa dinâmica com precisão. Em um caso na Paraíba, um adolescente criou *deepfake*s sexuais de uma professora e as distribuiu a colegas afirmando que as imagens eram autênticas. Em outro caso registrado no Ceará, um jovem adulto foi detido acusado de perseguir uma ex-professora e utilizar IA para criar imagens pornográficas falsas da docente. Em ambos os casos, as ferramentas de IA foram mobilizadas não por curiosidade tecnológica, mas como instrumentos deliberados de humilhação dentro do espaço profissional da vítima. Trata-se de uma estratégia de poder que se aproveita do anonimato e da verossimilhança dos *deepfake*s para perpetuar a misoginia e tornar o exercício docente um terreno de vulnerabilidade extrema para as mulheres.

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

2.2.3. COMERCIALIZAÇÃO E EXPLORAÇÃO ECONÔMICA

A tendência mais alarmante identificada pelo mapeamento foi a comercialização das imagens: alguns jovens venderam fotografias manipuladas de colegas em grupos privados em aplicativos de mensagens ou em plataformas de pornografia adulta. Esse comportamento acrescenta ao fenômeno uma dimensão econômica que o aproxima de modelos de exploração sexual comercial — com ou sem que os autores tenham plena consciência dessa qualificação jurídica.

Em um caso no estado de São Paulo, ao menos 14 estudantes de um colégio particular tiveram imagens manipuladas por IA publicadas em sites de conteúdo pornográfico. De acordo com a mãe de uma das vítimas, as montagens sexualizadas de sua filha alcançaram até 9 mil visualizações.

O mesmo padrão se repetiu em Minas Gerais, onde meninas de 12 a 17 anos tiveram fotografias manipuladas transformadas em imagens sexuais e posteriormente vendidas em canais do Telegram. O caso revelou ainda uma segunda camada de violência institucional: de acordo com o relato do pai de uma das vítimas, a escola teria minimizado a gravidade dos fatos, tratando a conduta dos alunos como uma “brincadeira” e desencorajando as famílias de prosseguirem com denúncias formais. Quando a instituição que deveria acolher as vítimas passa a atuar, mesmo que involuntariamente, como obstáculo à responsabilização, o dano se aprofunda — e a impunidade deixa de ser apenas percebida para se tornar estruturalmente produzida.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

UMA ANÁLISE E REVISÃO DE LITERATURA MAIS DETALHADA SOBRE AS POSSÍVEIS MOTIVAÇÕES DE AGRESSORES ESTÁ DISPONÍVEL NO CAPÍTULO 3, **SEÇÃO 3.3**.

Quando a produção de *deepfake*s sexuais migra para dentro das escolas — quando o infrator e a vítima se conhecem, estão nas mesmas salas, dividem os mesmos corredores —, a dinâmica motivacional ganha contornos específicos que precisam ser compreendidos à parte. A teoria do espaço sem lei continua operando como estrutura de fundo, mas as motivações que movem adolescentes a criar *deepfake*s de colegas são moldadas, também, por fatores que são exclusivos dessa fase da vida e desse ambiente.

O primeiro desses fatores é a banalização por proximidade tecnológica. Até poucos anos atrás, a criação de *deepfake*s realistas exigia algum conhecimento técnico especializado. Hoje, é possível fazê-lo por meio de um aplicativo, baixado diretamente nas redes sociais, sem nenhuma necessidade de habilidade técnica extraordinária. Esse rebaixamento da barreira técnica tem consequências diretas no perfil dos infratores: estudantes que nunca teriam cogitado montar uma imagem manipulada manualmente agora o fazem com o mesmo esforço que levaria a enviar uma mensagem. Uma pesquisa com mais de 27 mil estudantes do ensino fundamental e médio na República Tcheca, realizada em 2024, constatou que 2,77% dos alunos já haviam criado *deepfake*s de nudez — e que meninos tinham 3,56 vezes mais probabilidade de fazê-lo do que meninas.[10]

A facilidade de acesso, porém, não explica a motivação. Ela apenas remove o obstáculo que, antes, impedia que a motivação se convertesse em ação. Para compreender por que adolescentes criam *deepfake*s sexuais de colegas, é preciso observar o que os estudos sobre abuso sexual baseado em imagem revelam sobre os mecanismos internos desse comportamento. Uma pesquisa da Universidade de Monash sobre perpetradores adultos identificou o reforço entre pares como uma das motivações centrais. Esse mesmo mecanismo aparece com força ainda maior no ambiente escolar, onde as hierarquias de status social são mais frágeis e mais disputadas. Aos pesquisadores, uma vítima entrevistada descreveu a lógica do agressor de forma precisa: ele simplesmente queria "superar e arrancar risadas dos amigos", demonstrando que há um aspecto de querer agradar socialmente outros homens ao se engajar nesse tipo de comportamento.[11] **A criação do *deepfake* de nudez funciona, nesse contexto, menos como fim e mais como moeda: é um ato que compra status social entre um grupo, atenção e pertencimento dentro de um grupo.**

ESTE TIPO DE JUSTIFICATIVA TAMBÉM FOI OBSERVADA POR JOVENS BRASILEIROS QUE PARTICIPARAM DE UMA SURVEY DA SAFERNET BRASIL SOBRE DEEPFAKES EM ESCOLAS. PARA LER MAIS DETALHES, VÁ ATÉ A **SEÇÃO 4.1**.

ÍNDICE

1.
DEFINIÇÕES & TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3.
PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4.
ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5.
RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6.
RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Essa dimensão de performance social é central para entender por que o ambiente escolar é particularmente propício. Adolescentes estão em etapas críticas de formação de identidade e de busca por pertencimento social, o que os torna mais vulneráveis — tanto como vítimas quanto como perpetradores — a dinâmicas que exploram essas necessidades. Dentro desse quadro, a criação de um *deepfake* sexual de uma colega pode representar, para o perpetrador, uma forma de demonstrar poder, de ganhar aprovação de pares ou de punir alguém com quem tem um conflito — motivações que não são novas no universo do *bullying*, mas que encontram na tecnologia um instrumento com escala e persistência sem precedentes.

Outro fator determinante é a percepção de que o ato é inofensivo — ou ao menos, menos grave do que seria se as imagens fossem reais. A criação e o compartilhamento de *deepfake*s sexuais frequentemente começa como uma brincadeira, impulsionada por cérebros adolescentes ainda imaturos, com baixa autorregulação e suscetíveis à pressão dos pares, sem que os envolvidos reflitam sobre as consequências. Essa minimização não é apenas ingenuidade: ela é, em parte, uma racionalização coletivamente construída e reforçada dentro do grupo.

O que emerge desse conjunto de evidências é um retrato de motivações superpostas: curiosidade, busca por status entre pares, desejo de humilhar ou punir, e a sensação de que a distância entre imagem sintética e dano real é maior do que de fato é. A tecnologia não criou nenhuma dessas motivações. O que ela fez foi torná-las possíveis de aplicação com um custo quase zero em três dimensões. No plano técnico, qualquer adolescente com um smartphone pode produzir o conteúdo sem nenhum conhecimento especializado. No plano social, o grupo valida e recompensa a prática em vez de inibi-la. E no plano jurídico, a legislação específica sobre *deepfake*s ainda é recente, fragmentada e de aplicação irregular — o que significa que, na prática, poucos perpetradores enfrentam consequências reais. Essa tríade — facilidade de produção, aprovação do grupo e impunidade jurídica — é o que transforma uma motivação latente em ação concreta.

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

2.4. SUBNOTIFICAÇÃO PREOCUPANTE SOBRE OS CASOS MAPEADOS

Se esses números já são preocupantes, é preciso considerar que eles subestimam significativamente a realidade. Os casos identificados representam apenas uma fração da incidência real, por duas razões estruturais que comprometem qualquer tentativa de mensuração precisa do fenômeno.

A primeira é a dependência de casos amplamente noticiados pela mídia, o que limita drasticamente a visibilidade de ocorrências em regiões com cobertura jornalística precária ou inexistente — os chamados desertos de notícias.[12] Áreas do Norte e Nordeste que concentram milhões de habitantes frequentemente carecem de veículos de imprensa estruturados, fazendo com que episódios permaneçam circunscritos ao ambiente escolar e comunitário, sem alcançar visibilidade pública. Essa lacuna cria o risco de uma distorção analítica relevante: regiões com maior cobertura midiática podem parecer mais afetadas não porque concentram mais casos, mas porque possuem maior capacidade de registro e difusão.

A segunda razão é a inexistência de um protocolo unificado de registro, classificação e compartilhamento de dados. Em algumas delegacias, certos casos são registrados como compartilhamento de imagens íntimas (artigo 218-C no Código Penal), e não como simular a participação de criança ou adolescente em cena de sexo explícito ou pornográfica por meio de adulteração, montagem ou modificação de fotografia, vídeo ou qualquer outra forma de representação visual (artigo 241-C do ECA). Por se tratarem de artigos e tipos penais distintos, a consolidação de informações por meio de instrumentos como pedidos de acesso à informação resulta em dados fragmentados. Embora existam tipos penais aplicáveis à manipulação e divulgação não consentida de imagens, a ausência de diretrizes específicas para casos envolvendo IA generativa em ambiente escolar gera respostas fragmentadas e desarticuladas entre instituições.

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

CAPÍTULO TRÊS

Pontos de vista de diferentes atores envolvidos no fenômeno

ÍNDICE

1.
DEFINIÇÕES & TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3.
PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4.
ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5.
RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6.
RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Um mesmo fenômeno pode ser descrito de formas radicalmente diferentes dependendo de onde se está ao vivenciá-lo. Para a criança que tem sua imagem manipulada, o *deepfake* sexual é uma violação da dignidade sexual — ainda que o corpo nunca tenha sido tocado. Para o adolescente que o produz, pode ser um gesto de status entre pares, esvaziado de peso moral pelo ambiente que o envolve. Para o pai que desconhece o que acontece, é uma ameaça invisível. Para o criminoso que monetiza esse mercado, é uma oportunidade de baixo risco e alto retorno.

Essas quatro leituras não são contraditórias — são simultâneas. E é exatamente por isso que nenhuma resposta eficaz ao problema pode se concentrar em apenas uma delas. Este capítulo examina o fenômeno dos materiais artificiais de abuso e exploração sexual infantojuvenil a partir dessas quatro perspectivas: as vítimas, os jovens como grupo exposto, os adultos responsáveis pela proteção e os criminosos e infratores. O objetivo não é apenas descrever cada grupo, mas mostrar como as lacunas entre eles — de percepção, de informação, de poder — são, elas mesmas, parte do problema.

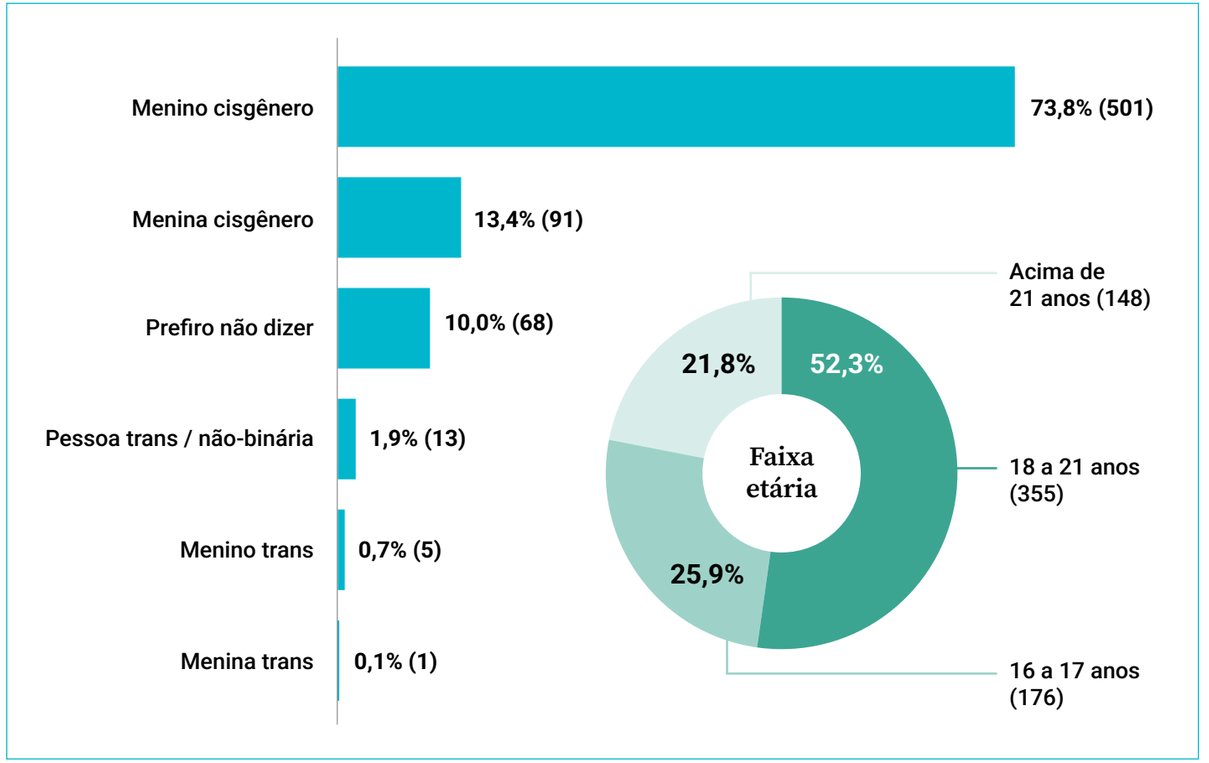
3.1. QUAL A OPINIÃO DOS JOVENS BRASILEIROS SOBRE ESTE PROBLEMA?

Entre adolescentes, a produção de *deepfake*s sexuais não pode ser compreendida apenas como desvio individual ou curiosidade tecnológica. Ela emerge de uma vulnerabilidade específica: a combinação entre imaturidade emocional, necessidade de pertencimento, busca por status social e dificuldade de antecipar consequências legais e psicológicas. Esse conjunto de fatores cria um terreno propício para que tecnologias de manipulação de imagem sejam apropriadas como instrumentos de experimentação, afirmação de poder e regulação informal das relações entre pares.

Para compreender a realidade brasileira, entre janeiro e março de 2026 a SaferNet Brasil conduziu uma pesquisa de opinião com 679 jovens, predominantemente entre 16 e 21 anos, sobre o problema dos *deepfake*s sexuais nas escolas. A faixa de 18 a 21 anos concentrou 52,3% da amostra (355 respondentes), seguida pela faixa de 16 a 17 anos (25,9%, ou 176 respondentes) e pelo grupo acima de 21 anos (21,8%, ou 148 respondentes). Em termos de gênero, 73,8% identificaram-se como meninos cisgênero (501 respondentes), 13,4% como meninas cisgênero (91 respondentes) e os 12,8% restantes distribuíram-se entre pessoas trans, não-binárias e respondentes que preferiram não declarar.[13]

FIGURA 4

Distribuição da amostra por gênero e faixa etária (N = 679)



De acordo com os resultados, a maioria dos jovens brasileiros possui algum grau de conhecimento sobre o que são *deepfake*s sexuais. Cerca de 75% dos participantes (510 respondentes) declararam ter ouvido falar do termo: 53,3% (362) afirmaram saber exatamente do que se trata e outros 21,8% (148) reconheceram já ter ouvido falar sobre *deepfake*s, ainda que sem conseguir explicar com precisão do que se trata. Os 24,9% restantes (169 respondentes)

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

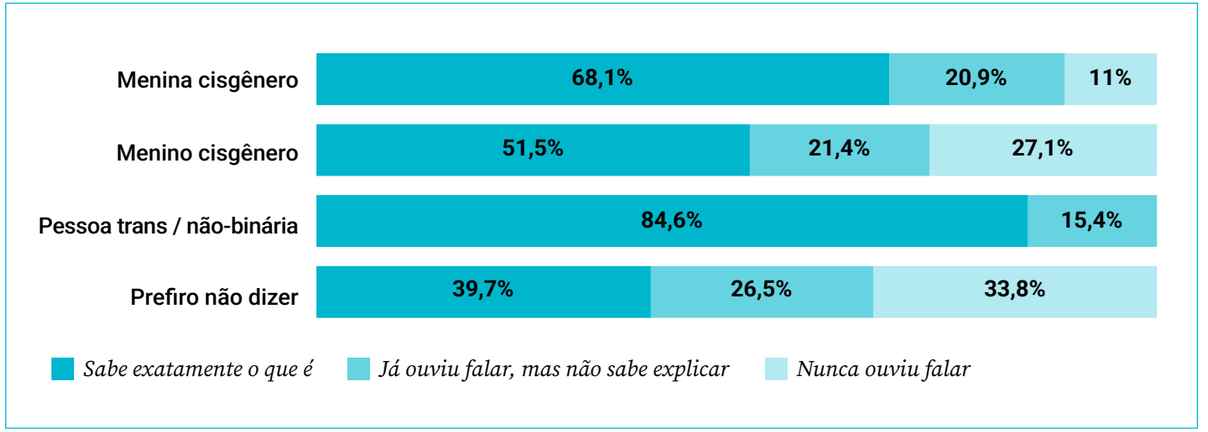
declararam nunca ter ouvido falar do termo — número expressivo, considerando o caráter cotidiano que a tecnologia já assumiu no debate público sobre violência de gênero digital.

A leitura desagregada por gênero revela, no entanto, um padrão menos óbvio do que à primeira vista poderia parecer. Entre os meninos cisgênero, 51,5% (258 de 501) afirmaram saber exatamente o que é um *deepfake* sexual, 21,4% (107) já tinham ouvido falar sem saber explicar e 27,1% (136) nunca tinham ouvido falar.

Entre as meninas cisgênero, 68,1% (62 de 91) declararam saber exatamente o que é, 20,9% (19) já tinham ouvido falar sem saber explicar e apenas 11,0% (10) nunca tinham ouvido falar. Em termos proporcionais, portanto, são as meninas que demonstram maior familiaridade com o tema — um achado que nos faz perguntar se as meninas, por serem as principais vítimas, conhecem melhor o fenômeno justamente porque vivem, ou veem seus pares vivê-lo, com mais frequência. Já entre os meninos, quase metade (48,5%) admite domínio incompleto sobre o tema — defasagem que, à luz dos dados sobre produção e atitude apresentados adiante, ajuda a iluminar o terreno cultural em que esse tipo de violência se reproduz.

FIGURAS 5 E 6

Nível de conhecimento sobre *deepfake*s sexuais, por gênero



Gênero	Sabe exatamente o que é	Já ouviu falar, mas não sabe explicar	Nunca ouviu falar	Total
Menina cisgênero	62 (68,1%)	19 (20,9%)	10 (11,0%)	91
Menino cisgênero	258 (51,5%)	107 (21,4%)	136 (27,1%)	501
Pessoa trans / não-binária	11 (84,6%)	2 (15,4%)	0 (0,0%)	13
Menino trans	3 (60,0%)	2 (40,0%)	0 (0,0%)	5
Menina trans	1 (100,0%)	0 (0,0%)	0 (0,0%)	1
Prefiro não dizer	27 (39,7%)	18 (26,5%)	23 (33,8%)	68
Total	**362 (53,3%)**	**148 (21,8%)**	**169 (24,9%)**	**679**

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

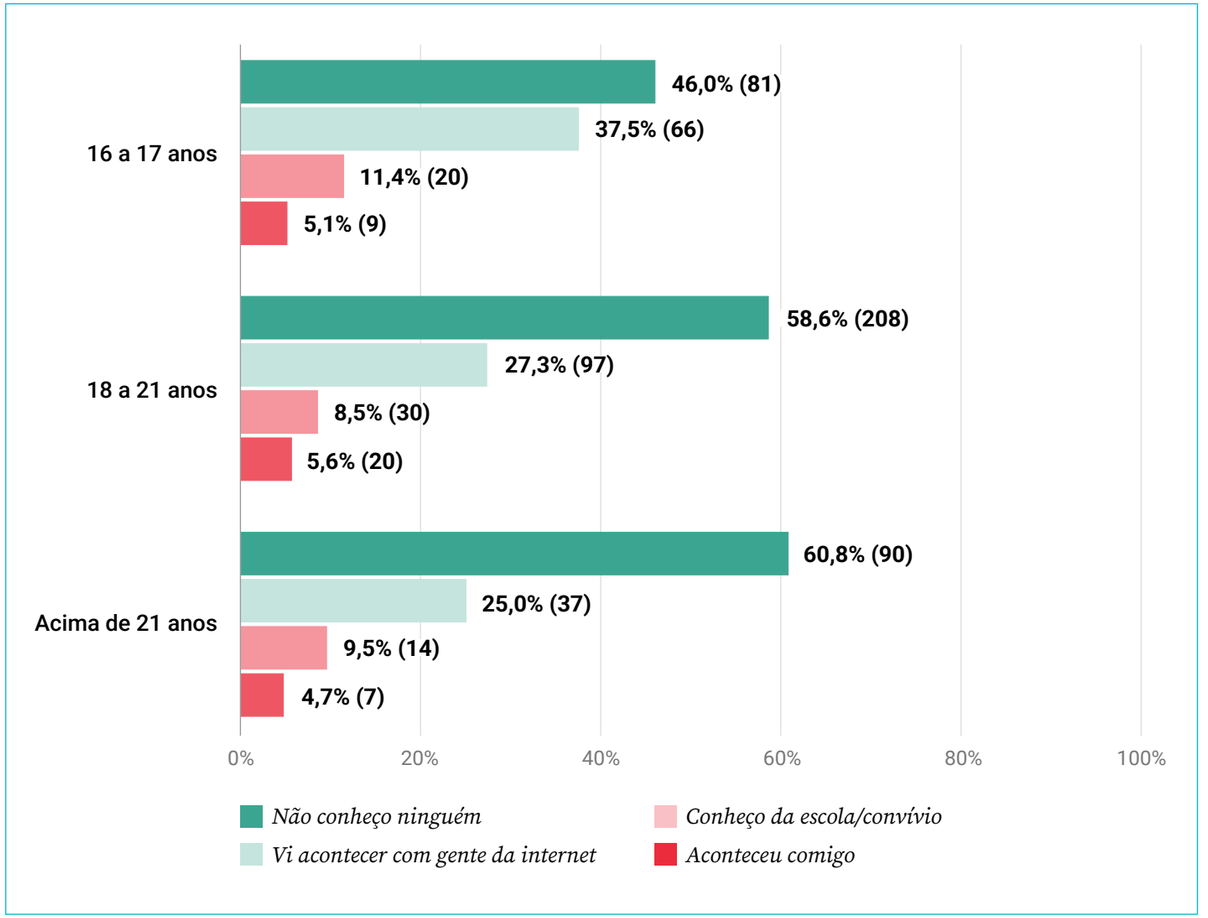
7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

A faixa etária de 18 a 21 anos foi também a que registrou o maior número absoluto de vitimização direta: 20 participantes declararam ter sido pessoalmente vítimas de um *deepfake* sexual. Nas faixas de 16 a 17 anos e acima de 21 anos, esses números foram de nove e sete respondentes, respectivamente, totalizando 36 vítimas diretas — 5,3% da amostra. O contato indireto — casos vistos na internet envolvendo terceiros ou celebridades — foi a forma de exposição mais comum em todos os grupos, mencionado por 29,5% dos respondentes (200 jovens). Já o conhecimento de vítimas do convívio pessoal — colegas de escola, amigos, conhecidos — manteve-se relativamente estável entre as faixas, oscilando entre 8% e 11%, sugerindo que a proximidade com casos reais não varia de forma expressiva com a idade dentro desta amostra.

FIGURA 7

Contato com vítimas de *deepfake* sexual, por faixa etária



ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

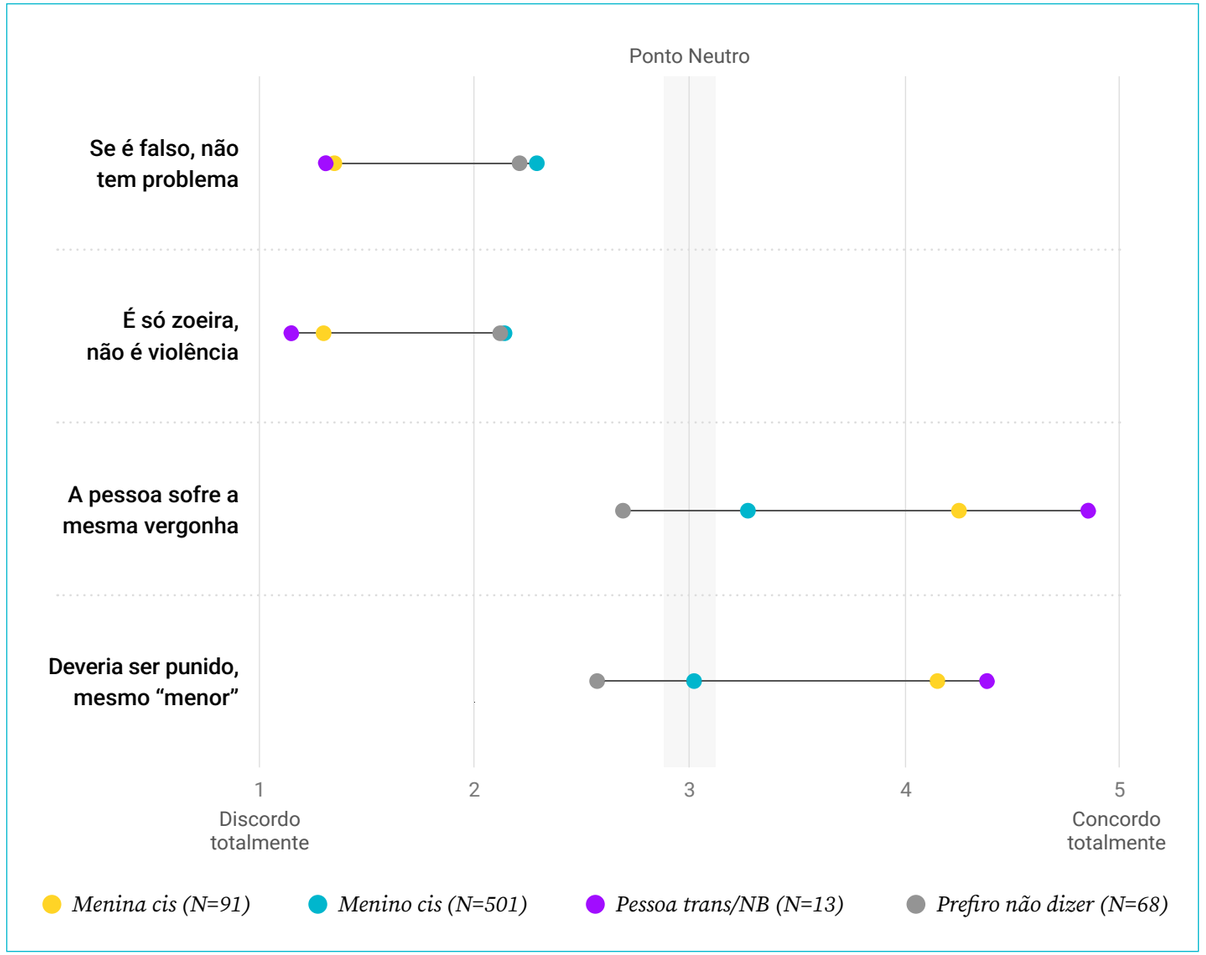
SOBRE ESTA PESQUISA

Esse mapa de conhecimento, no entanto, conta apenas a primeira camada do problema. Saber o que é um *deepfake* sexual não é o mesmo que reconhecer nele uma forma de violência — e é precisamente nesse hiato que reside a fratura mais reveladora dos dados. Quando solicitados a se posicionar diante da afirmação "se a imagem é falsa, não tem problema", numa escala de 1 (discordo totalmente) a 5 (concordo totalmente), as meninas cisgênero registraram média 1,35 e mediana 1, enquanto os meninos cisgênero apresentaram média 2,29 e mediana 2.

O mesmo padrão se repete na frase "é só zoeira, não é violência": meninas com média 1,30 contra meninos com 2,14. Inversamente, ao avaliarem a afirmação de que "a pessoa sofre a mesma vergonha que sofreria com uma foto verdadeira", as meninas cisgênero pontuaram 4,25 (mediana 5), enquanto os meninos ficaram em 3,27 (mediana 3). E quanto à proposição de que o agressor "deveria ser punido, mesmo sendo 'menor de idade'", a distância se mantém: 4,15 entre as meninas, 3,02 entre os meninos.

FIGURA 8

Atitudes diante de afirmações sobre *deepfakes* sexuais, por gênero (escala 1–5; 1 = discordo totalmente, 5 = concordo totalmente)



ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

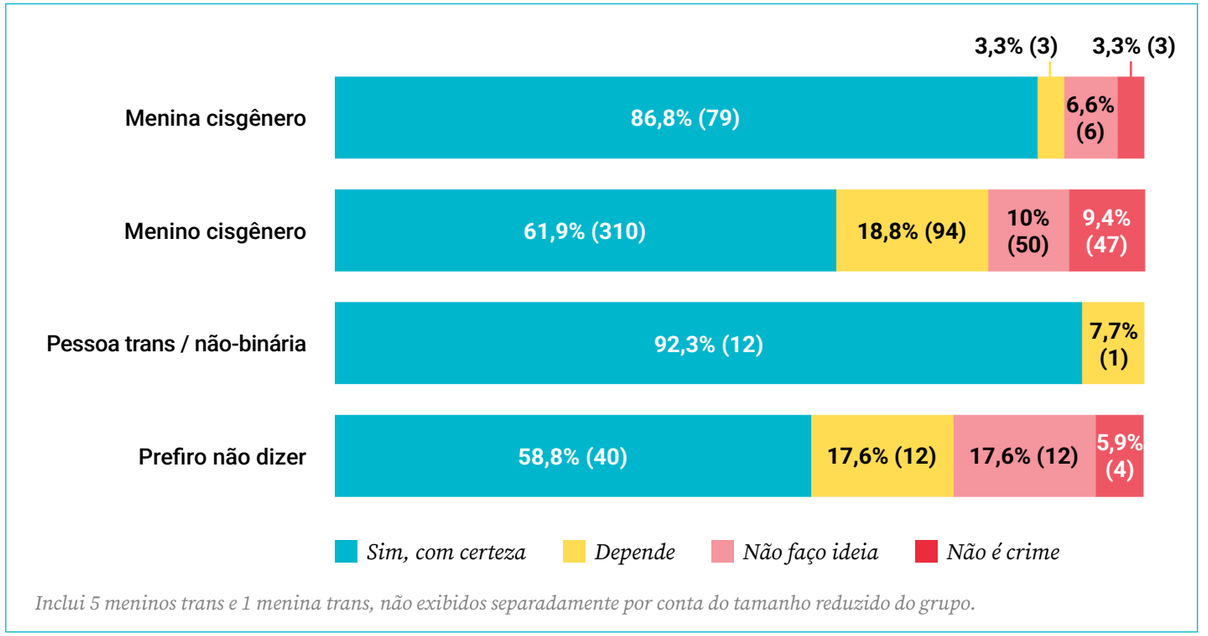
Esses números traduzem, em escala numérica, aquilo que já era observado por especialistas no terreno: o fenômeno opera, entre meninos brasileiros, em uma zona de moralidade ambígua, em que a falsidade da imagem é mobilizada como argumento de redução de gravidade — uma justificativa que não encontra eco entre as meninas, que tendem a perceber, com clareza, que o sofrimento da vítima não depende da autenticidade técnica do material. A literatura internacional sobre violência de gênero digital tem destacado que essa "lógica do simulacro" é, em si mesma, um marcador cultural: ela reflete a ideia de que apenas o corpo material importa, ignorando que a dignidade sexual é violada justamente no plano simbólico — pelo olhar de terceiros, pela circulação não consentida e pelo deslocamento da vítima de sujeito a objeto.

A diferença de gênero também aparece — embora de forma menos acentuada — no entendimento jurídico da conduta. Quando perguntados se a criação de um *deepfake* sexual de uma colega da escola sem o consentimento dela é crime, 86,8% das meninas cisgênero responderam "sim, com certeza", contra 61,9% dos meninos. Mais relevante é que 18,8% dos meninos cisgênero (94 respondentes) declararam que "depende", e outros 9,4% (47 respondentes) afirmaram categoricamente que "não é crime".

FIGURA 9

Percepção de criminalidade da conduta, por gênero

Pergunta: criar um *deepfake* sexual de uma colega da escola sem o consentimento dela é crime?



Inclui 5 meninos trans e 1 menina trans, não exibidos separadamente por conta do tamanho reduzido do grupo.

Essa zona de tolerância moral e jurídica não é abstrata: ela se materializa, antes de mais nada, na produção. Entre os 501 meninos cisgênero da amostra, 62 (12,4%) admitiram já ter criado um *deepfake* sexual. Entre as meninas cisgênero, esse número foi de zero. Entre os respondentes que preferiram não declarar gênero, sete (10,3%) também afirmaram já ter produzido esse tipo de conteúdo.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

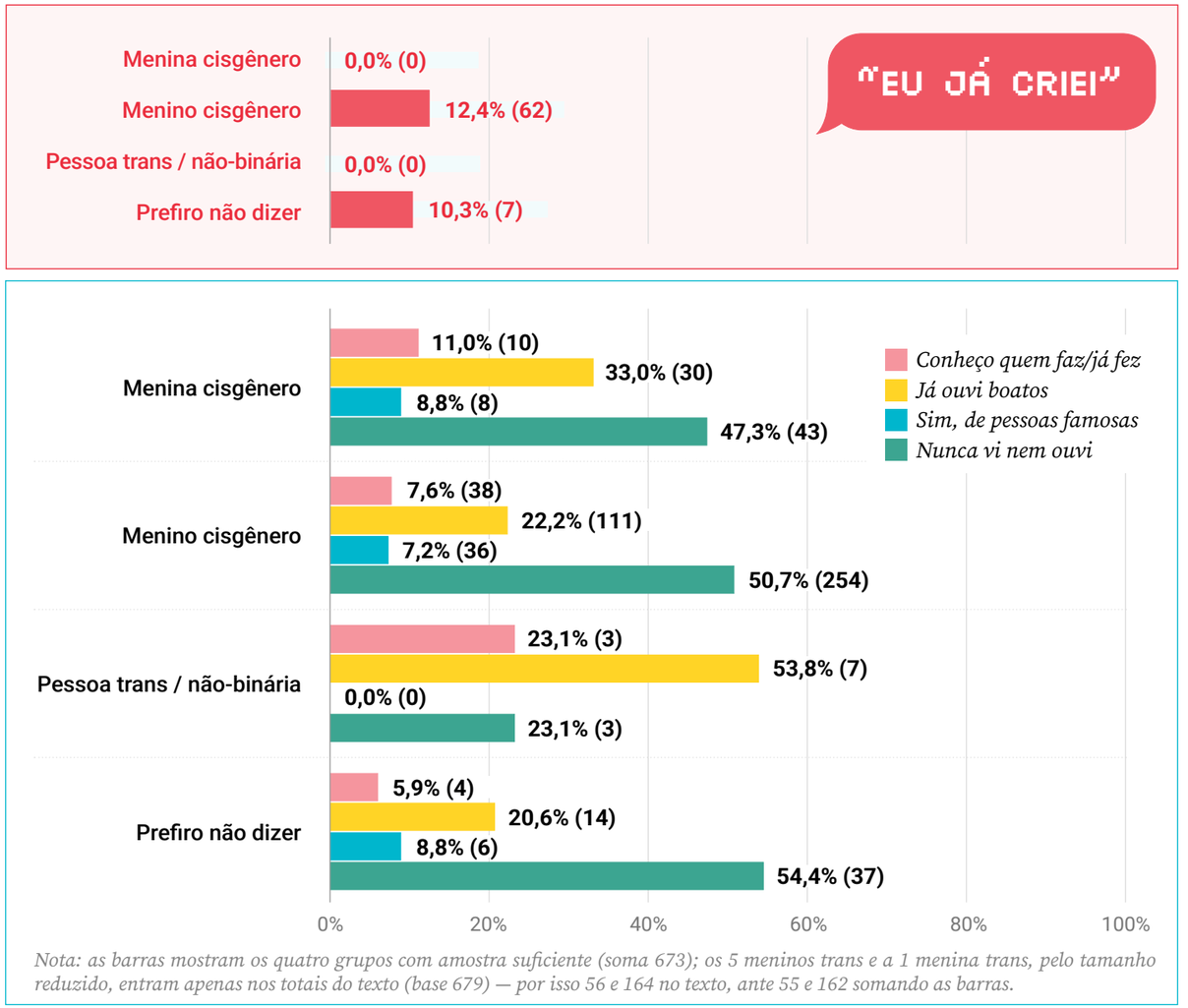
No total, 69 dos 679 jovens entrevistados — 10,2% — assumiram autoria, número que muito provavelmente subestima a realidade dada a natureza socialmente estigmatizada da pergunta. Somam-se a esse grupo os 56 jovens (8,2%) que afirmam conhecer pessoalmente alguém que produz e os 164 (24,2%) que dizem ter ouvido boatos sobre colegas que o fazem. Quando reunidas, essas três posições — produzir, conhecer quem produz, ouvir falar — abarcam 42,6% da amostra. O *deepfake* sexual, na vida desses jovens, não é uma abstração distante: é algo que está, no mínimo, na conversa do dia a dia.

Para entender como essa produção ocorre na prática, a survey perguntou onde, segundo a percepção dos jovens, esses conteúdos são mais comumente fabricados (Figura 11).

A resposta mais frequente foi o Telegram, citado por 42,6% dos respondentes (289 de 679), seguido por buscas no Google ou em sites de busca (33,9%), por outros sites (29,5%), por aplicativos de nudificação disponíveis em lojas para celular (27,1%) e por servidores no Discord ou fóruns (21,9%). Apenas 1,6% mencionou o Twitter/X (Grok) e um único respondente citou o ChatGPT/OpenAI. Entre os jovens de 16 e 17 anos — a faixa etária mais nova da amostra —, o Te-

FIGURA 10

Autoria e proximidade com criadores de *deepfakes* sexuais, por gênero



Nota: as barras mostram os quatro grupos com amostra suficiente (soma 673); os 5 meninos trans e a 1 menina trans, pelo tamanho reduzido, entram apenas nos totais do texto (base 679) — por isso 56 e 164 no texto, ante 55 e 162 somando as barras.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

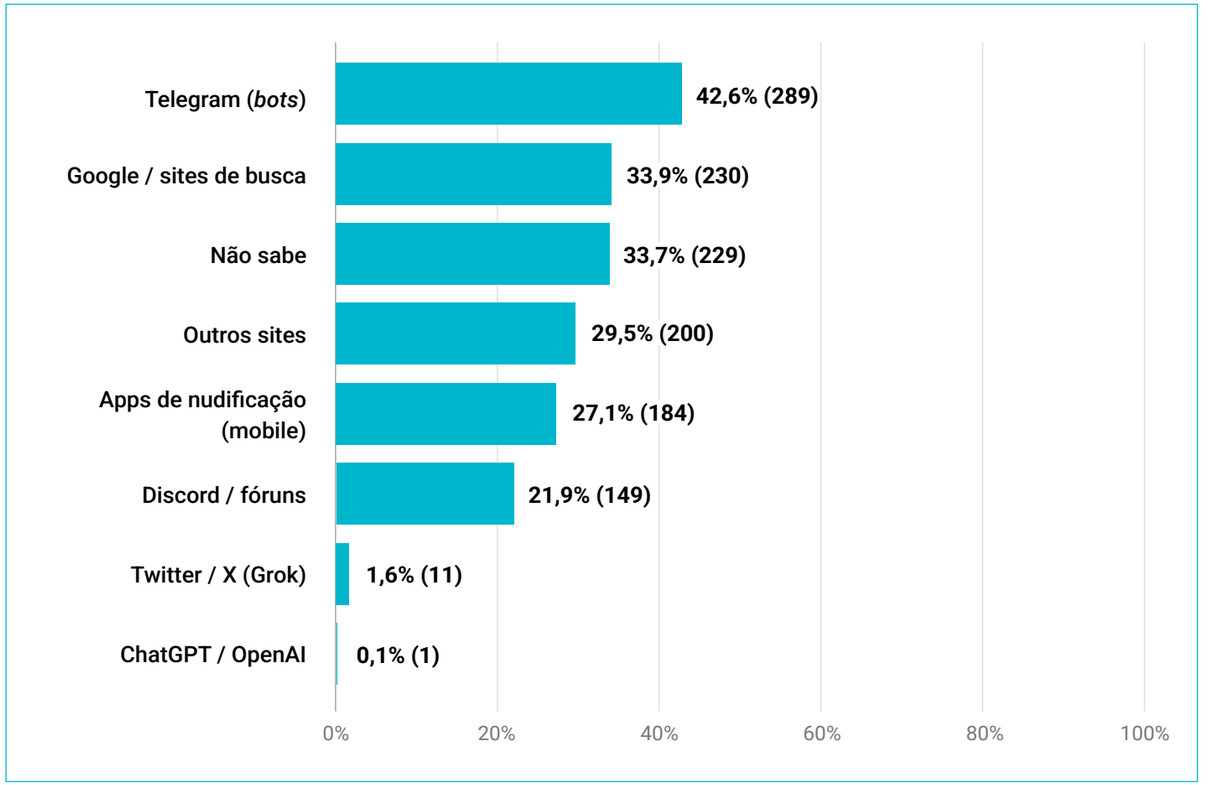
SOBRE ESTA PESQUISA

legram aparece com força ainda maior: 51,1% (90 de 176) o apontam como o principal canal, seguido pelo Google (48,9%). O dado é convergente com o que vem sendo documentado pela imprensa investigativa e por organizações da sociedade civil: uma investigação do Núcleo Jornalismo, em parceria com o *Pulitzer Center*, identificou no início de 2025 ao menos 23 *bots* ativos na plataforma capazes de gerar material ou conteúdo de violência sexual contra crianças e adolescentes, em diversos idiomas, incluindo o português.[14] O Brasil é hoje o terceiro maior mercado do Telegram, atrás apenas de Rússia e Índia, com cerca de 38% da população usando o aplicativo.[15] Os jovens não estão imaginando um problema: estão descrevendo, com precisão, a infraestrutura à qual têm acesso direto.

A escola, lugar onde grande parte dessas violências encontra solo fértil, aparece nos dados como um espaço de exposição parcial mas não desprezível. Quando perguntados com que frequência viam o compartilhamento de *deepfake*s sexuais ou nudes falsos na própria escola, 7,7% dos respondentes disseram que era "muito comum (todo mundo sabe/comenta/vê)", 13,1% relataram que "acontece de vez em quando", e 19,6% classificaram como "raro". Apenas 59,6% afirmaram nunca ter presenciado.

FIGURA 11

Plataformas mais citadas como locais de produção de *deepfakes* sexuais (resposta múltipla, N = 679)



ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

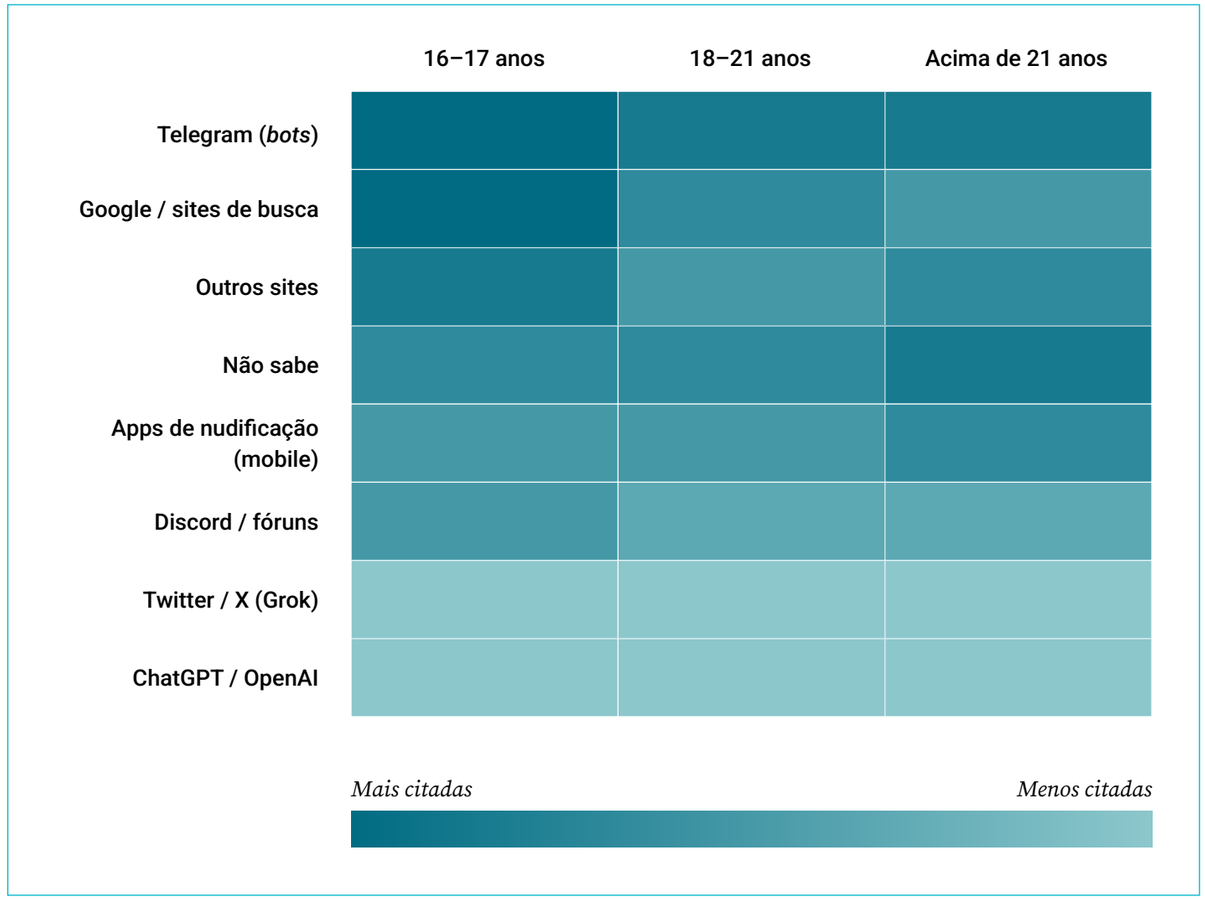
6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

FIGURA 12

Plataformas mais citadas, por faixa etária (% da faixa)



ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

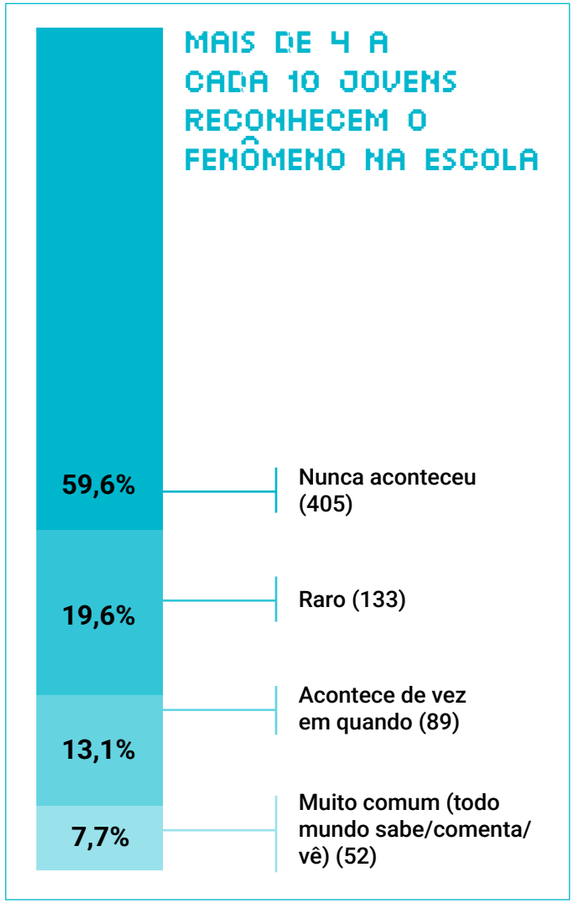
7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Em outras palavras, mais de quatro em cada dez jovens reconhecem alguma presença do fenômeno em seu próprio ambiente escolar — número compatível com o levantamento da SaferNet Brasil discutido no capítulo anterior. Casos como o do Colégio Santo Agostinho, no Rio de Janeiro (2023), o de uma escola tradicional em Porto Alegre (2024), o do Colégio Sagrada Família em Blumenau (2026) e o de Campo Grande (2026) — em que adolescentes chegaram a comercializar imagens falsas de colegas por R$ 50 dentro da própria escola — ilustram a continuidade entre o ambiente digital e o ambiente escolar, não como mundos paralelos, mas como uma única superfície social atravessada pela mesma violência.

FIGURA 13

Frequência com que jovens veem *deepfakes* sexuais sendo compartilhadas na própria escola (N = 679)



Diante desse cenário, a percepção dos próprios jovens sobre a capacidade de resposta da escola é eloquente: apenas 19,6% (128 de 653) acreditam que a instituição de ensino está preparada, sabendo o que fazer e acolhendo a vítima. A ampla maioria — 80,4% (525 de 653) — não confia nessa preparação: 40,1% (262) afirmam categoricamente que a escola não está preparada, supondo que os responsáveis culpariam a vítima ou não compreenderiam a tecnologia envolvida, e outros 40,3% (263) avaliam a preparação apenas como “mais ou menos”.

Em nenhuma faixa etária a confiança na escola chega a um quarto dos jovens: acreditam que a escola saberia agir 24,1% dos que têm 16 e 17 anos, 17,4% dos que têm de 18 a 21 e 19,6% dos maiores de 21.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

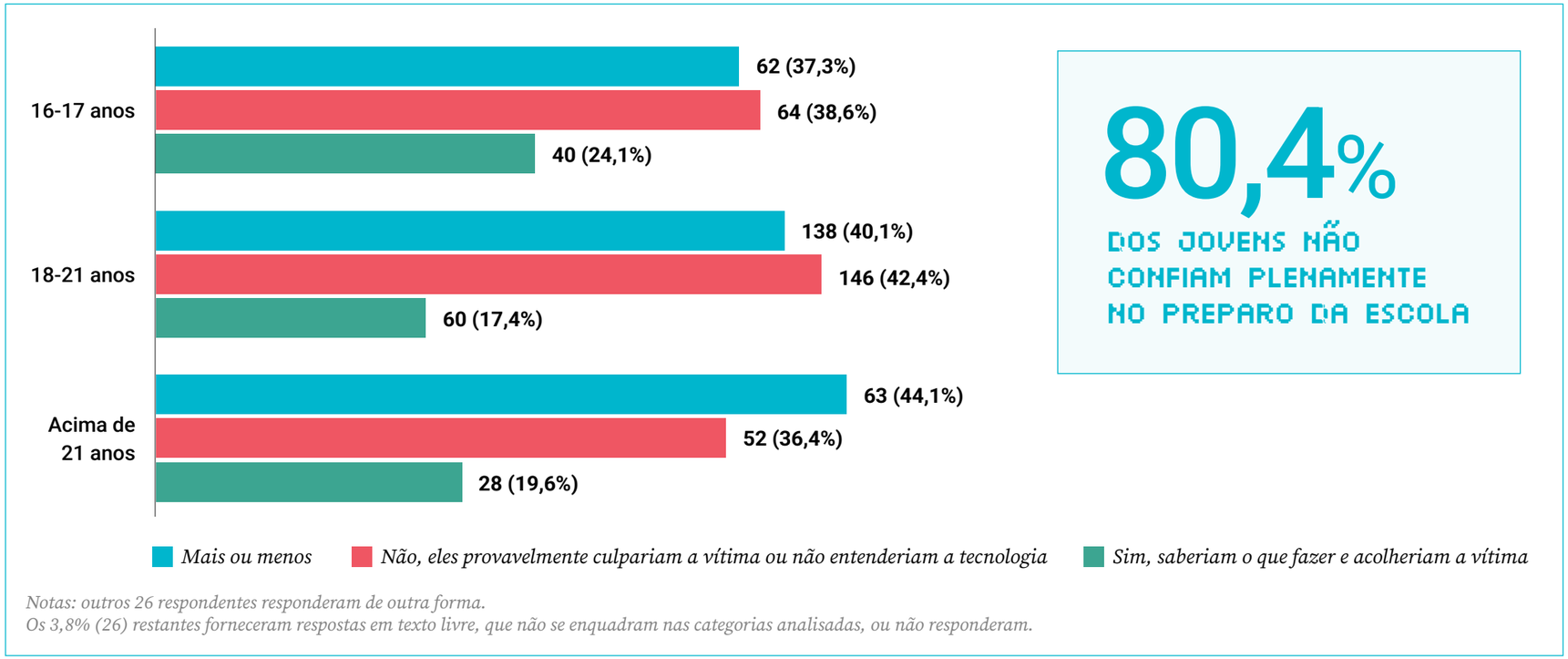
7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Por fim, perguntados sobre o que diriam às empresas de IA que desenvolvem produtos de IA para resolver esse problema — uma pergunta aberta respondida por 409 dos 679 jovens —, a distribuição temática das respostas oferece um retrato significativo das prioridades dessa geração.

FIGURA 14

Percepção dos jovens sobre o preparo da escola para lidar com *deepfakes* sexuais, por faixa etária.
Pergunta: Você acha que a sua escola está preparada para lidar com um caso desses hoje?



80,4%
DOS JOVENS NÃO CONFIAM PLENAMENTE NO PREPARO DA ESCOLA

Notas: outros 26 respondentes responderam de outra forma.
Os 3,8% (26) restantes forneceram respostas em texto livre, que não se enquadram nas categorias analisadas, ou não responderam.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

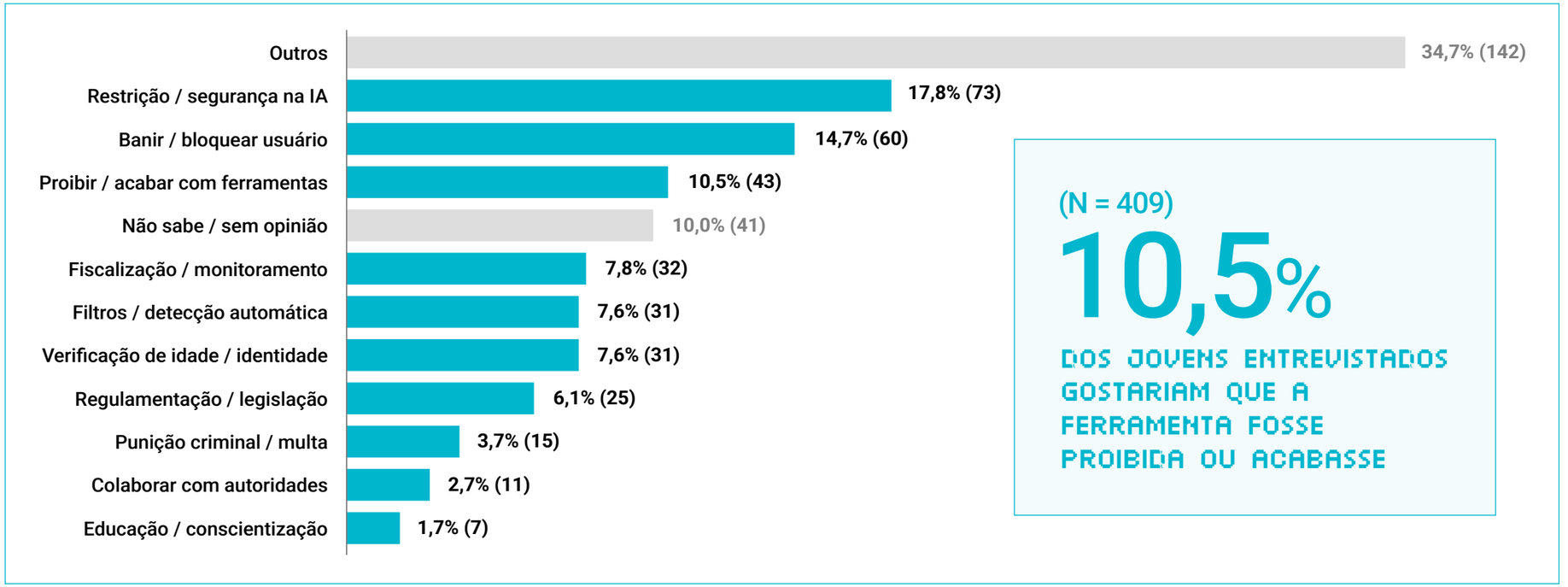
7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Na codificação das respostas, os pedidos mais frequentes foram, em ordem: maior restrição e segurança nos próprios sistemas de IA (17,8%); banimento ou bloqueio de usuários que criam esse tipo de conteúdo (14,7%); proibição ou eliminação completa das ferramentas que permitem a fabricação de *deepfake*s sexuais (10,5%);

FIGURA 15

Temas das sugestões dos jovens às empresas de IA (Q17, pergunta aberta, N = 409 respostas codificadas)



ÍNDICE

1.
DEFINIÇÕES & TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3.
PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4.
ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5.
RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6.
RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

fiscalização e monitoramento mais rigorosos (7,8%); filtros e detecção automática (7,6%); verificação obrigatória de idade ou identidade dos usuários (7,6%); e regulamentação ou legislação específica (6,1%).

A leitura conjunta desses indicadores revela um descompasso essencial. De um lado, os jovens brasileiros entrevistados mostram-se majoritariamente conscientes de que o problema existe, sabem onde ele acontece, conhecem suas vítimas — direta ou indiretamente — e têm opiniões nítidas sobre o que esperam das plataformas e do Estado. De outro, os mesmos jovens convivem com uma fratura interna entre o que reconhecem como crime e o que praticam, toleram ou silenciam — fratura que se sobrepõe, com precisão preocupante, às linhas de gênero.

Não se trata de uma característica brasileira isolada: é, antes, a tradução local de um padrão global em que, segundo o UNICEF, ao menos 1,2 milhão de crianças tiveram suas imagens manipuladas em *deepfake*s sexualmente explícitos no último ano.[16] No Brasil, porém, esse padrão ganha contornos próprios — opera sobre uma juventude profundamente conectada, em um ambiente digital onde o Telegram tem penetração massiva, em que a escola chega ao centro do fenômeno sem treinamento e sem protocolo, e em que a violência de gênero ainda é, com frequência, desqualificada socialmente como brincadeira entre pares.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

3.2. QUAL A OPINIÃO DOS EDUCADORES SOBRE ESTE PROBLEMA?

É complexo unificar as opiniões de um grupo inteiro de pessoas, sobretudo enquanto não houver pesquisas nacionais com educadores sobre o tema. Para captar percepções qualitativas, a SaferNet Brasil organizou, em fevereiro de 2026, um workshop em São Paulo (SP) durante o pré-lançamento do *Guia de Literacia em IA*, desenvolvido em parceria com o *Pulitzer Center for Crisis Reporting*.[17] O encontro reuniu 30 educadores de áreas diversas — diretores, professores do ensino fundamental e médio, de escolas públicas e particulares — que debateram, em quatro grupos, um cenário hipotético envolvendo *deepfake*s sexuais produzidos e comercializados por estudantes de um colégio paulista. A partir das discussões, foi possível identificar percepções recorrentes sobre a natureza do problema e os obstáculos atuais para enfrentá-lo.

Um sentimento generalizado é o de desorientação institucional. Os educadores relatam que, diante de crimes cibernéticos envolvendo IA, as escolas frequentemente não sabem se devem tratar o caso internamente ou acionar instâncias externas, e a ausência de diretrizes uniformes faz com que decisões de grande peso recaiam isoladamente sobre a direção escolar. Essa fragilidade de orientação é apontada como uma das causas centrais da paralisia institucional observada em casos concretos.

Outra preocupação manifestada é a tendência à omissão por receio reputacional. Na percepção dos participantes, especialmente em escolas particulares, há pressão para abafar casos a fim de proteger o nome da instituição ou evitar atritos com famílias influentes. Nas escolas públicas, o problema assume outros contornos: a sobrecarga de questões estruturais faz com que a violência digital seja tratada como demanda secundária, e os educadores consideram que os danos gerados por IA precisam ser encarados com a mesma seriedade dispensada às violências físicas. Há, ainda, crítica explícita às respostas atualmente disponíveis na rede pública, em que a transferência compulsória do aluno infrator para outra unidade é vista como uma medida que apenas desloca o problema, sem garantir acompanhamento ou reabilitação.

Os educadores também demonstraram entender o fenômeno de forma ampla e relacional. Para eles, tanto vítimas quanto autores do compartilhamento ou produção de *deepfake*s são jovens em fase de desenvolvimento e demandam acolhimento especializado — não se trata, na visão do grupo, de um problema que se resolve apenas pela punição. Há, ainda, um diagnóstico compartilhado de que o uso nocivo da IA não atua de forma isolada, mas se entrelaça com questões sociais mais amplas, como misoginia, racismo e dinâmicas identitárias entre adolescentes. Essa leitura leva os participantes a enxergar a prevenção como algo que exige observação atenta do bem-estar dos alunos como um todo, e não apenas resposta pontual a incidentes on-line.

Por fim, há consenso de que a responsabilidade pelo enfrentamento do problema não pode recair apenas sobre as escolas. Na visão dos educadores, o combate aos danos da IA generativa precisa ser descentralizado e compartilhado entre escolas, autoridades públicas, sociedade civil e empresas de IA generativa, com participação ativa dos próprios jovens. A percepção de que o problema é estrutural — e não estritamente escolar — atravessa todas as discussões registradas no workshop.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

3.2.1. PERCEPÇÕES SOBRE AS INSTITUIÇÕES DE ENSINO

Os educadores avaliaram que as escolas atuam sem padrão definido diante de crimes cibernéticos e *deepfake*s sexuais, frequentemente sem saber como proceder. Essa ausência de referência interna deixa direção e professores em posição vulnerável, obrigados a tomar decisões isoladas em situações de alta complexidade jurídica e emocional.

Houve também uma leitura crítica sobre o comportamento institucional diante da exposição pública. Na percepção dos participantes, muitas escolas — sobretudo as particulares — tendem a abafar casos para proteger a reputação da instituição ou evitar desgaste com famílias influentes, o que os educadores classificam como uma forma de omissão.

Nas escolas públicas, o diagnóstico é distinto: a sobrecarga de problemas estruturais faz com que a violência digital seja despriorizada. Os educadores consideram que os danos provocados por IA não vêm sendo tratados com a mesma seriedade dispensada às violências físicas, e apontam que as respostas atualmente disponíveis são insuficientes — a penalidade máxima costuma ser a transferência compulsória do aluno infrator, medida vista como mero deslocamento do problema, já que não há acompanhamento contínuo da reabilitação nem do histórico do estudante.

3.2.2. PERCEPÇÕES SOBRE AS AUTORIDADES PÚBLICAS

Em relação ao poder público, os educadores identificaram uma lacuna de articulação entre as Secretarias de Educação estaduais e o Ministério Público. Para o grupo, essa desarticulação faz com que as escolas não saibam exatamente para onde encaminhar denúncias, sobrecarregando a direção escolar com decisões que deveriam ter um fluxo institucional definido. Há também a percepção de que falta atendimento multidisciplinar dentro das escolas, especialmente apoio psicológico e social. Os participantes entendem que tanto vítimas quanto autores da disseminação ou produção de *deepfake*s são sujeitos em desenvolvimento e demandam acolhimento especializado — uma necessidade que, segundo eles, o poder público ainda não supre adequadamente.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

3.2.3. PERCEPÇÕES SOBRE A SOCIEDADE CIVIL

No âmbito da sociedade civil — organizações, ONGs e fundações —, os educadores reconhecem um papel relevante de governança e produção de conhecimento. Por atuarem em parceria com autoridades e especialistas, essas entidades são vistas como produtoras de material de alta confiabilidade, capaz de oferecer respaldo técnico aos gestores escolares. O grupo também enxergou nessas organizações um potencial articulador para a observação de sinais comportamentais precoces nos alunos, como isolamento ou queda súbita de desempenho. Essa percepção parte de um diagnóstico mais amplo: o uso nocivo da IA não atua de forma isolada, mas está profundamente interligado a problemas sociais, identitários, de misoginia e racismo, exigindo observação contínua do bem-estar dos estudantes antes que as crises se instalem.

3.2.4. PERCEPÇÕES SOBRE AS EMPRESAS DE IA GENERATIVA

As empresas de IA generativa e plataformas de IA concentram boa parte das críticas dos educadores. Há a percepção de que os atuais mecanismos de cadastro e identificação de usuários são frágeis, permitindo a criação de contas com pouca verificação e dificultando a responsabilização em casos de uso malicioso. Vale notar, contudo, que medidas como exigência de documentos de identidade e rastreamento de localização — discutidas pelo grupo — esbarram em princípios fundamentais da Lei Geral de Proteção de Dados (LGPD), como a minimização de dados.

Os participantes também avaliam que falta diferenciação clara entre ferramentas seguras e ferramentas usadas para fins prejudiciais, o que dificulta tanto a orientação a famílias e escolas quanto a proteção de crianças e adolescentes. Nessa mesma linha, há a percepção de que as plataformas não se preocupam ativamente com o conteúdo gerado, incluindo *deepfake*s e materiais artificiais de exploração sexual infantil.

Por fim, os educadores apontam uma deficiência de transparência e rastreabilidade nos sistemas atuais. Na visão do grupo, não há mecanismos consolidados para sinalizar a origem sintética de textos e imagens, nem ferramentas que permitam auditar o que está sendo gerado nas plataformas — o que torna especialmente difícil identificar quem solicitou a criação de uma imagem falsa envolvendo, por exemplo, uma vítima escolar.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

ESTA SEÇÃO FALA DA PERSPECTIVA DE INFRATORES E AGRESSORES NO ÂMBITO DA PRÁTICA DE EXPLORAÇÃO SEXUAL INFANTIL. PARA A PERSPECTIVA NOS CASOS DE BULLYING, COMO A MAIORIA DOS CASOS DE DEEPFAKES EM ESCOLAS, VÁ ATÉ A **SEÇÃO 2.3**.

3.3. O CRIME NA PERSPECTIVA DOS INFRATORES E AGRESSORES

Quando falamos sobre crimes digitais, é preciso reconhecer como a percepção da internet como um espaço separado do mundo real ainda afeta o comportamento de quem comete atos ilícitos. O fato de que diferentes condições influenciam o cometimento de diferentes tipos de crime — contra a vida, contra a dignidade sexual, contra o patrimônio — é estudado há décadas. O que muda no ambiente digital é a escala e a velocidade com que essas condições se estabelecem.

Ferramentas como os *deepfake*s sexuais tornam a produção e a circulação de material sintético mais fáceis, baratas e rápidas — e, ao mesmo tempo, quebram o vínculo direto entre agressor e vítima, ampliando a sensação de impunidade. Essa distância psicológica é um fator determinante: quando o agressor não precisa estar fisicamente presente no abuso, o custo emocional da ação diminui. O ambiente socialmente validado pelo grupo reforça, adicionalmente, a sensação de que o espaço é seguro. A tecnologia, portanto, não cria o "espaço sem lei" — mas atua como seu multiplicador, tornando a prática mais frequente, mais acessível e mais tolerada dentro desses grupos.

A primeira pesquisa a entrevistar diretamente perpetradores de abuso *deepfake* sexual, conduzida pela Universidade de Monash (Austrália) e publicada no *Journal of Interpersonal Violence*, oferece evidências sobre o funcionamento dessas motivações.[18] O estudo identificou quatro motivações centrais: ganho financeiro via sextorsão, curiosidade, intenção de causar dano e reforço entre pares.

A dimensão do reforço social aparece de forma explícita nas falas dos entrevistados. Um perpetrador descreveu o desejo de obter afirmação do grupo: "Eles dizem 'Uau, isso é incrível'. Você sabe, eles ficavam boquiabertos". A validação pelos pares não era um subproduto da prática — era uma das suas motivações centrais.

Para além da validação social, emerge uma motivação ainda mais específica nestas comunidades, a 'cristalização de corpos infantis de pessoas reais' por meio da IA. Observa-se em fóruns e páginas, uma obsessão direcionada a celebridades e influenciadoras mirins, onde a tecnologia é empregada para gerar conteúdos sexuais que mantêm essas pessoas eternamente crianças, mesmo após terem atingido a vida adulta. Trata-se de um esforço deliberado para perpetuar o abuso, contornando a idade real das vítimas para sustentar uma fantasia de exploração sexual permanente. Nesses casos, a

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

IA atua como uma ferramenta que busca perpetuar a fantasia de abuso sexual, mantendo-as como um objeto de exploração sexual infantojuvenil, ainda que as vítimas não sejam mais crianças na realidade.

O estudo também documentou que perpetradores e espectadores normalizam esses comportamentos por meio de mecanismos cognitivos que minimizam os danos causados. A percepção de que o abuso por *deepfake* é "menos grave" do que outras formas de abuso baseado em imagem porque "a imagem é falsa" foi um elemento recorrente — observado tanto como técnica de justificação pelos perpetradores quanto como mecanismo de enfrentamento pelas próprias vítimas. Essa minimização é, em si, um sinal de como o ambiente molda a percepção moral dos envolvidos.

Significativamente, nenhum dos dez perpetradores entrevistados havia enfrentado consequências legais — ainda que todos tivessem criado e compartilhado as imagens. As vítimas, por sua vez, relataram ausência de mecanismos de proteção, mesmo quando os casos foram reportados à polícia. A impunidade, portanto, não era apenas percebida: era real.

Um levantamento publicado no *Journal of Cybersecurity* com mais de 16 mil respondentes em 10 países aprofunda esse panorama: mais de 1 em cada 10 relatou ao menos um comportamento de perpetração de abuso sexual baseado em imagem, sendo a poliperpetração — prática de múltiplas formas de abuso — mais comum do que a perpetração isolada.[19] A acessibilidade crescente das ferramentas é central para compreender essa dinâmica: a geração de *deepfake*s deixou de exigir conhecimento técnico especializado e passou a estar ao alcance de qualquer usuário com acesso a um smartphone.

A PERCEPÇÃO DE IMPUNIDADE NÃO É O ÚNICO FATOR QUE SUSTENTA ESSE ECOSSISTEMA. O ASPECTO FINANCEIRO EXERCE PAPEL IGUALMENTE CENTRAL — E A EXISTÊNCIA DE UM MERCADO ESTRUTURADO TRANSFORMA O QUE PODERIA SER UMA PRÁTICA ISOLADA EM UMA CADEIA ECONÔMICA COM INCENTIVOS PRÓPRIOS.

Um estudo publicado em outubro de 2024 analisou a economia interna da plataforma *MrDeepFakes*, mapeando 43 mil vídeos envolvendo 3.800 indivíduos, com mais de 1,5 bilhão de visualizações.[20] As vítimas eram majoritariamente mulheres, representando 95,3% dos indivíduos e 88% dos vídeos. Os valores cobrados por *deepfake* variaram de US$ 1 a US$ 1.500, influenciados por fatores como a quantidade de imagens disponíveis da vítima, a qualidade desejada e a complexidade técnica — uma estrutura de precificação que trata a imagem da vítima como mercadoria.

O fenômeno da extorsão sexual mediada por *deepfake*s é um desdobramento direto desse mercado. Dados da Thorn apontam que 13% das vítimas de sextorsão relataram ter sido extorquidas por alguém usando *deepfake*s gerados por IA.[21] Em quase três quartos dos casos, o perpetrador usou a ameaça de compartilhar imagens sexuais ou informações pessoais para pressionar a vítima — frequentemente ameaçando enviar o material a familiares, amigos ou colegas. É uma abordagem que mantém a vítima presa pelo medo e pela vergonha, sem que o abuso precise sequer ter ocorrido de fato. Um em cada cinco adolescentes pesquisados relatou ter experienciado sextorsão.

ÍNDICE

1.
DEFINIÇÕES & TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3.
PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4.
ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5.
RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6.
RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

O caso sul-coreano de 2024 ilustra com nitidez como todos esses fatores — tecnologia acessível, validação entre pares, mercado ilícito e impunidade percebida — se combinam em escala. Estudantes do sexo masculino de centenas de escolas e universidades criaram canais no Telegram para produzir e compartilhar vídeos explícitos sintéticos de professoras, colegas e, em alguns casos, familiares.[22] Um dos maiores canais tinha mais de 220 mil inscritos. Muitos perpetradores e vítimas tinham menos de 18 anos.

A análise desse caso, publicada no periódico *Gender and Education* em janeiro de 2026, conclui que a violência digital de gênero nesse contexto não é obra de indivíduos isolados, mas de um ecossistema multidirecional: comunidades online masculinas que produzem e monetizam narrativas de ressentimento, plataformas que permitem a escala sem fricção, e ferramentas tecnológicas que reduzem o custo de entrada a quase zero.

Ferramentas de nudificação e manipulação de imagem cada vez mais acessíveis reduzem barreiras de entrada, enquanto a percepção de anonimato diminui o risco percebido. Isso cria um ciclo contínuo: quanto mais fácil e socialmente tolerado se torna o acesso a essas tecnologias, mais frequente e disseminada é a prática.

3.4. PERSPECTIVA DAS VÍTIMAS E IMPACTOS PSICOLÓGICOS

Por ser um fenômeno relativamente novo ainda há poucos estudos sobre os impactos dessa violência do ponto de vista das vítimas. Esta prática se insere na categoria de abuso sexual baseado em imagem (IBSA, na sigla em inglês), cujos impactos já são amplamente documentados pela literatura científica. No contexto da IA generativa, a vitimização apresenta características distintas e complexas devido à natureza do material produzido: a imagem retrata o rosto e a identidade de uma pessoa real, mas a cena de violência é uma fabricação ou montagem.

A perda de controle sobre a própria identidade amplifica o dano emocional — o *deepfake* reduz a imagem da vítima a um objeto manipulável, dissociado de sua vontade e de sua história, gerando uma dissociação profunda — um sofrimento real experimentado diante de uma imagem que ela sabe que não produziu, mas que o mundo consome como se fosse ela.

Essa dualidade cria uma barreira significativa para o reconhecimento social do dano envolvido. O fato de a cena da violência ser uma montagem ou fabricação gerada por IA muitas vezes alimenta a percepção equivocada — tanto de agressores quanto de parte do público — de que não se trata de uma violência "real", essa racionalização ignora que o dano não deriva da autenticidade da imagem, mas da violação da autonomia, da privacidade e da dignidade da vítima — e da experiência real e concreta de humilhação, medo e perda de controle que ela produz.

Um dos aspectos mais críticos documentados é o estigma social gerado por essa prática. A vítima passa a ser lembrada e definida pelo episódio, enfrentando o que pesquisadores chamam de "morte social" no ambiente digital. Em muitos casos mapeados, o impacto é tão profundo que obriga a criança ou adolescente a mudanças drásticas: o abandono de redes sociais, a alteração da aparência física e até a mudança de escola ou endereço para tentar escapar da estigmatização, como nos casos mapeados pela SaferNet Brasil nas escolas brasileiras.

Essa realidade é particularmente cruel durante a adolescência, fase do desenvolvimento em que o pertencimento ao grupo e a opinião dos pares são pilares decisivos para a saúde mental. O isolamento provocado pela humilhação pública interrompe processos vitais de sociabilização, e o medo constante de ser reconhecida ou de que outros acreditem na veracidade da imagem gera um estado de hipervigilância persistente. Do ponto de vista da saúde mental, o sofrimento prolongado frequentemente desencadeia quadros de transtornos mentais. Estudos recentes apontam que a revitimização contínua pode desencadear o Transtorno de Estresse Pós-Traumático (*TEPT*, na sigla em português), além de comportamentos de autolesão e ideação suicida, que representam os desfechos mais graves deste ecossistema de abuso.[23]

Esse conjunto de sintomas é agravado por dois fatores estruturais que diferenciam o *deepfake* de outras formas de abuso de imagem. O primeiro é a permanência: diferentemente de uma agressão física, o conteúdo digital pode permanecer acessível indefinidamente, aparecendo em diferentes plataformas mesmo após tentativas de remoção, que faz com que a vitimização não seja um evento pontual, mas um processo contínuo e cíclico, em que o trauma é reativado a cada novo compartilhamento ou visualização. O segundo é a verossimilhança: imagens geradas por IA são frequentemente indistinguíveis de fotografias reais, o que impede a vítima de simplesmente afirmar que o conteúdo é falso e ser acreditada. Mesmo quando o *deepfake* é desmascarado ou nunca foi acreditado, ele inflige danos duradouros ao alterar negativamente o discurso em torno da

ÍNDICE

1.
DEFINIÇÕES & TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3.
PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4.
ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5.
RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6.
RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

vítima — um efeito que pesquisadores denominam "dano ilocucionário".[24]

Um estudo publicado em setembro de 2025 no *Journal of Interpersonal Violence* — um dos primeiros trabalhos revisados por pares a entrevistar tanto vítimas quanto infratores — registrou que três vítimas tentaram reportar seus casos de *deepfake*s sexuais não consentidos à polícia sem que qualquer medida fosse tomada, recebendo orientações que minimizam o dano, como "troque o número do celular" ou "apenas apague a foto". Essa resposta institucional inadequada não apenas deixa a vítima desamparada, mas reforça a percepção de impunidade dos agressores.

Grupos já em situação de maior vulnerabilidade enfrentam riscos adicionais. Pessoas LGBTQIAPN+, grupos étnicos minoritários, pessoas com deficiência e comunidades linguisticamente diversas estão expostos a taxas mais elevadas de violência sexual facilitada por tecnologia e enfrentam barreiras adicionais para buscar apoio. Mulheres negras, em particular, são afetadas pela sobreposição de estereótipos racistas e sexistas que tendem a aumentar a atribuição de culpa às vítimas e reduzir a percepção de dano por parte de terceiros.[25]

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

CAPÍTULO QUATRO

Ecossistema digital que facilita a violência sexual mediada por IA

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

A criação, manutenção e exploração comercial de ferramentas de IA generativa capazes de reproduzir violência sexual contra crianças e adolescentes não decorrem de ações isoladas, mas de um ecossistema criminoso estruturado. Tal ecossistema opera sob a lógica de uma organização criminosa descentralizada, numa cadeia que abrange o desenvolvimento técnico de modelos generativos sem salvaguarda, bem como sua distribuição que extrai lucro financeiro dessas práticas. Por isso, a pesquisa da SaferNet Brasil buscou compreender o ecossistema cibercriminoso que sustenta e facilita essa violência — o que envolve analisar os fatores econômicos, culturais e sociais que estruturam as comunidades voltadas à produção desses crimes e ao desenvolvimento das ferramentas que os viabilizem.

Este capítulo trata dos principais vetores de risco associados a esse ecossistema, entendidos como canais e plataformas digitais cujas características facilitam a criação ou a propagação dos conteúdos abusivos. Os vetores primários — de maior impacto e recorrência — são descritos nos subtópicos seguintes, acompanhados de cinco estudos de caso.* Esta seção da pesquisa contém descrições textuais de conversas entre criminosos que podem causar desconforto, mas não há imagens explícitas de violência ou exploração sexual infantojuvenil (real ou gerada por IA).

Os dados disponíveis evidenciam uma escalada sem precedentes. Segundo o *AI Incident Monitor* da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), ao longo de 2020 os incidentes de segurança envolvendo IA e crianças mal ultrapassavam uma ocorrência mensal. Janeiro de 2026 rompe esse padrão: foram 169 incidentes registrados em um único mês — mais do que a soma de tudo o que havia sido contabilizado nos dois primeiros anos de monitoramento.

Embora há décadas exista um receio social sobre a possibilidade de criação de conteúdos artísticos, materiais escritos ou peças de pro-

*__Nota metodológica.__ *Os dados apresentados neste capítulo não decorrem de amostragem estatística, mas de seleção qualitativa. A análise priorizou domínios recorrentes e ambientes de maior incidência identificados pela equipe técnica, sem a pretensão de cobrir de forma exaustiva todos os espaços onde tais crimes ocorrem.*

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

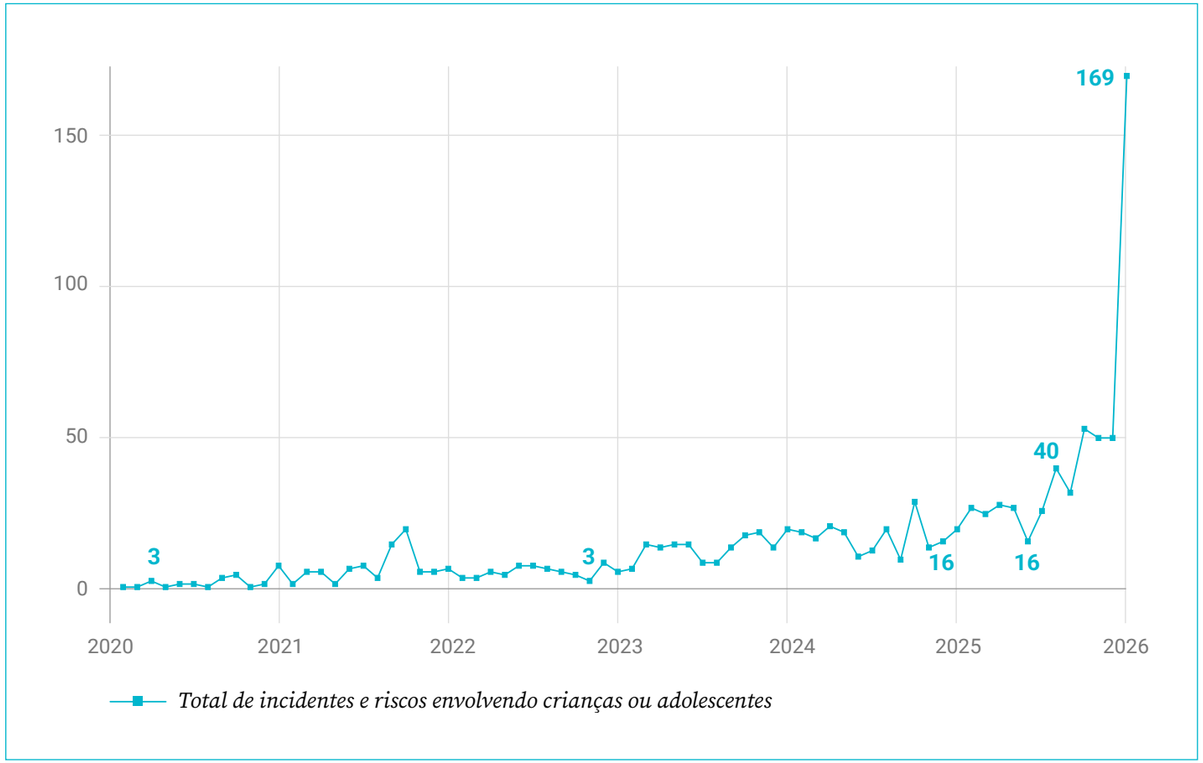
6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

FIGURA 16

Casos de incidentes e riscos classificados como "dano psicológico" e "violação de direitos humanos" envolvendo IA generativa e crianças ou adolescentes, com base nos registros da Organização para a Cooperação e Desenvolvimento Econômico (OCDE). Dados de janeiro de 2020 até janeiro de 2026 .



paganda que enalteçam a exploração sexual infantil — tanto que esse risco já estava previsto no ECA em 2008 —, a tecnologia de IA generativa transforma esse receio em um dano real, não mais em uma mera possibilidade sem consequências. O monitoramento conduzido pela SaferNet Brasil para esta seção da pesquisa revela um cenário grave, que se estende para além da ausência de regulamentação da IA generativa entre adolescentes e do desenvolvimento descontrolado dessa mesma tecnologia em aplicações associadas à pornografia, à nudificação não consensual e à violência sexual.

Em 2026, **as ferramentas necessárias para a criação desses conteúdos estão presentes tanto na *surface web* quanto na *dark web***, em

A DARK WEB É PARTE DA INTERNET ACESSÍVEL APENAS COM SOFTWARES COMO TOR BROWSER, QUE OCULTAM IDENTIDADE E LOCALIZAÇÃO. É USADA PARA PRIVACIDADE E ANONIMATO, MAS TAMBÉM ABRIGA ATIVIDADES ILEGAIS, MERCADOS CLANDESTINOS E COMUNICAÇÃO DIFÍCIL DE RASTREAR.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

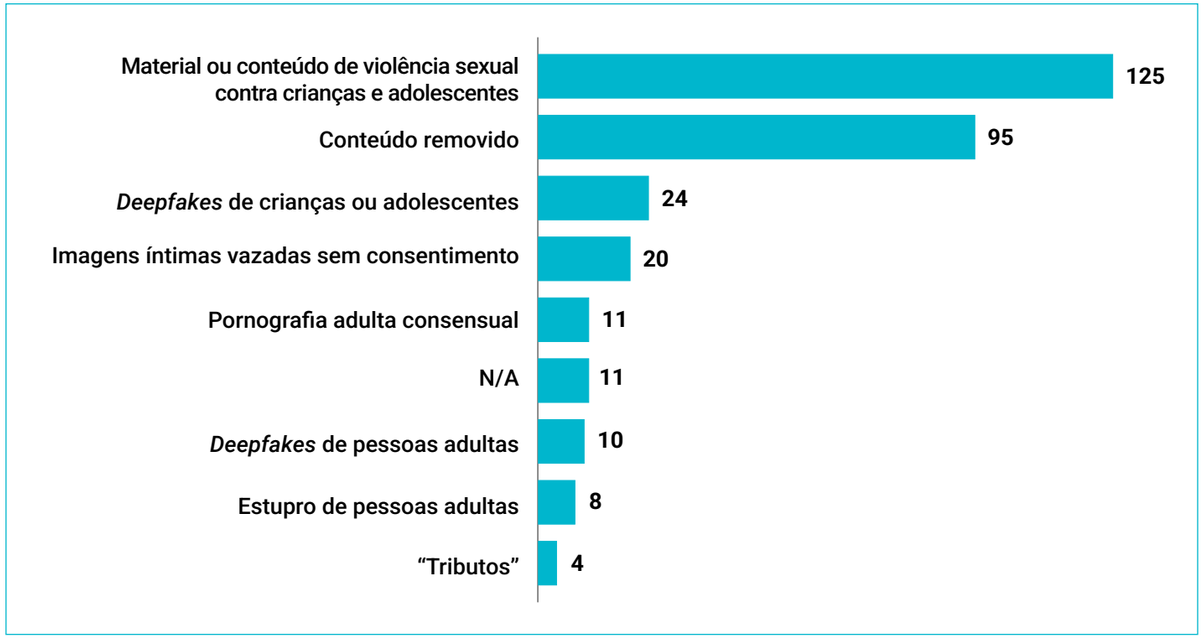
modelos gratuitos e pagos, e podem ser localizadas com relativa facilidade por mecanismos de busca comuns — sobretudo quando os filtros de navegação segura estão desativados. Esse achado evidencia a fragilidade dos controles atualmente existentes e coloca em xeque a eficácia das salvaguardas voluntárias adotadas pelas plataformas. Soma-se a isso um padrão estrutural relevante: plataformas de pornografia adulta, historicamente caracterizadas pela ausência de moderação ativa, tornaram-se também espaços de hospedagem de materiais artificiais de exploração sexual infantojuvenil — uma continuidade entre os dois fenômenos que não pode ser tratada de forma dissociada.

Para esta análise, foram selecionadas 264 URLs submetidas ao canal de denúncias da SaferNet Brasil, relacionadas a suspeitas de abuso ou exploração sexual infantil. Esse número cresceu ao longo dos estudos de caso; a seleção inicial serviu para delimitar o escopo e constituir uma amostra quantitativa de partida. Os critérios de escolha foram qualitativos — como a presença de palavras-chave nos relatos dos denunciantes, entre elas "*deepfake*", "nudificação" e "IA" — além de evidências de pesquisas anteriores que já apontavam plataformas e aplicativos reconhecidos pela SaferNet Brasil como vetores de risco. As denúncias foram extraídas em setembro de 2025, embora cubram registros acumulados sobretudo entre 2018 e 2025.

Das URLs analisadas, 149 (56,4%) apresentaram ao menos um tipo de conteúdo ilícito, ao passo que 111 (42%) não continham nenhuma das categorias investigadas. A categoria mais prevalente foi a de conteúdos com crimes contra crianças — conteúdos artificiais de abuso sexual infantojuvenil e outros —, identificada em 149 URLs ou 56,4% de toda a amostra.

FIGURA 17

Total de análises das 264 URLs da amostra inicial. Há 308 menções porque uma mesma URL contém mais de uma análise (ex.: um site que hospeda tanto *deepfakes* sexuais de adultos como material artificial de abuso sexual infantil)



ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Em outras 34 URLs (12,9%), havia conteúdos com crimes contra adultos. Em 20 URLs, havia crimes tanto contra pessoas adultas como crianças e adolescentes — por exemplo, um site que hospeda tanto material de exploração sexual infantojuvenil quanto filmagens de estupro de pessoas adultas.

No que diz respeito a *deepfake*s sexuais, 27 URLs (10,2%) continham ao menos um tipo de conteúdo sintético. *Deepfake*s envolvendo crianças foram detectados em 24 URLs (9,1%), ao passo que *deepfake*s envolvendo adultos apareceram em 10 (3,8%). Em sete URLs, foram identificadas todas as categorias analisadas simultaneamente — *deepfake*s de adultos e de crianças, além de crimes contra adultos e crianças —, configurando os casos de maior gravidade e sobreposição no conjunto de dados.

A distribuição por tipo de plataforma aprofunda essa leitura. As 264 URLs concentram-se em dois grandes grupos: *dark web*, com 127 URLs (48,1%), e plataformas de pornografia ou conteúdo erótico adulto, com 123 (46,6%). Os demais tipos — aplicativos de nudificação (cinco URLs), redes sociais (cinco) e sites diversos (quatro) — somam apenas 14 registros, mas sua análise qualitativa revela padrões que não podem ser ignorados.

A *dark web* apresentou a maior taxa de conteúdo ilícito: 109 de suas 127 URLs (85,8%) continham ao menos um tipo de crime ou material abusivo. A totalidade desses ilícitos dizia respeito a crimes contra crianças, enquanto crimes contra adultos apareceram em apenas 10 URLs (7,9%), sempre em sobreposição a material infantil. *Deepfakes* na *dark web* foram encontrados em 13 URLs e, dado relevante, todos os 13 casos envolviam exclusivamente crianças — nenhum registro continha *deepfake* de pessoa adulta. Esses números confirmam que a *dark web* opera, nesta amostra, como vetor predominantemente voltado à exploração infantil, com uma concentração temática que evidencia redes estruturadas de produção e distribuição.

Nas plataformas de pornografia, a taxa de conteúdo ilícito foi substancialmente menor — 37 de 123 URLs (30,1%) —, motivada pela maior taxa de remoção de conteúdo por serem sites hospedados na internet aberta (77 das 123 URLs, ou 62,6%, já haviam sido removidas no momento da análise), porém com perfil distinto: 23 URLs (18,7%) envolviam crimes contra crianças e 19 (15,4%), crimes contra adultos, estes últimos concentrados sobretudo em imagens íntimas vazadas sem consentimento. Como esperado, os aplicativos de nudificação, ainda que apresentem poucas URLs da amostra original (inicialmente 5, com URLs adicionais analisadas durante o estudo de caso sobre o Telegram), destacam-se por conta da gravidade absoluta dos achados: a totalidade deles apresentou *deepfakes* de crianças e crimes contra crianças, e em 80% dos casos havia simultaneamente *deepfakes* de mulheres e crimes contra adultos. Trata-se do único tipo de plataforma em que todas as URLs analisadas concentraram todas ou quase todas as categorias de ilícito ao mesmo tempo — um dado que reforça o caráter intrinsecamente abusivo dessas ferramentas e torna difícil sustentar qualquer argumento de que possuam finalidade legítima predominante.

Este monitoramento identificou, ainda, a consolidação de comunidades organizadas de indivíduos que consomem, comercializam ou compartilham esse tipo de conteúdo. Muitas dessas comunidades operam em espaços de difícil acesso, onde são desenvolvidos e disseminados guias, tutoriais e sistemas de IA generativa treinados com material ou conteúdo de violência sexual contra crianças e adolescentes, frequentemente de forma gratuita, o que reduz as barreiras de entrada e amplia o alcance dessas redes.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

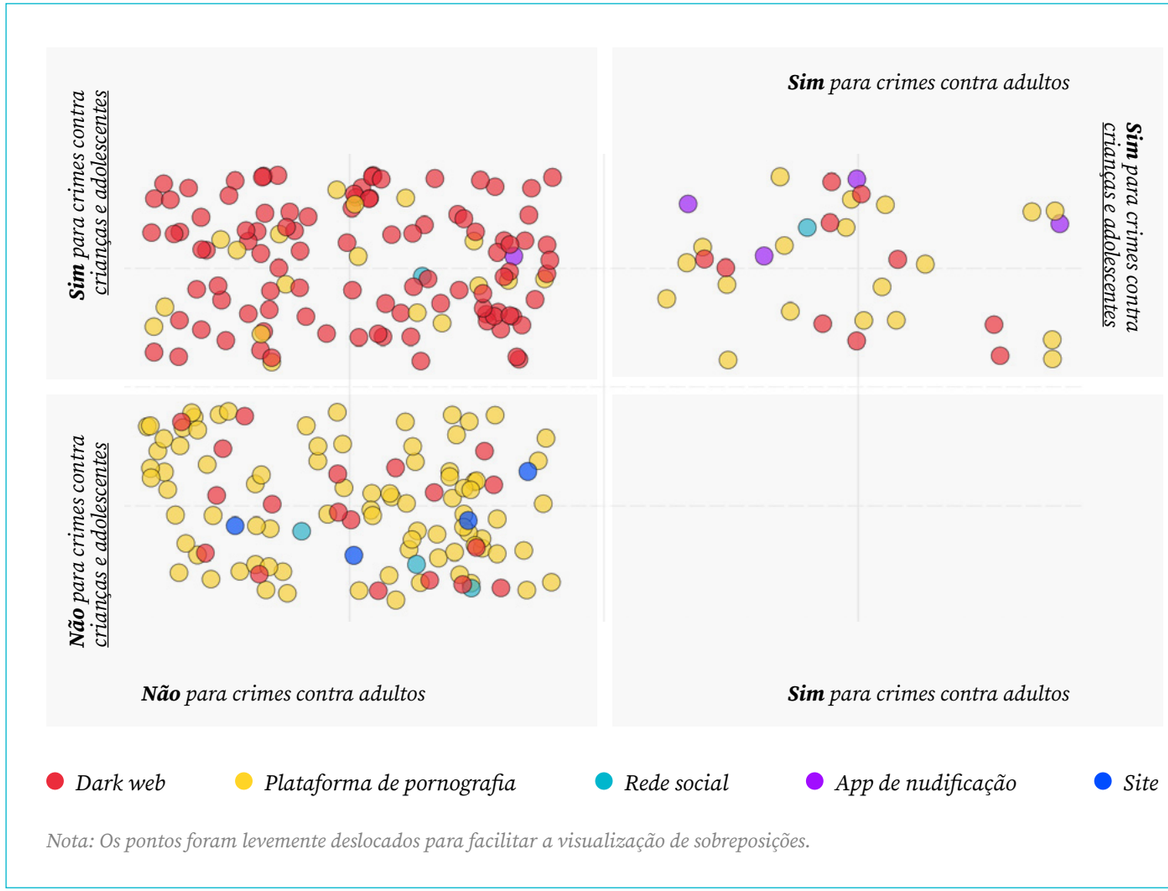
6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

FIGURA 18

Gráfico de dispersão com distribuição de análise de conteúdos vs. plataformas, crime ou não contra crianças e mulheres.



Nota: Os pontos foram levemente deslocados para facilitar a visualização de sobreposições.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

O QUE ENCONTRAMOS NESTE VETOR DE RISCO

A análise de URLs de grupos e ferramentas de nudificação conduzida pela SaferNet Brasil indica que o maior risco não reside apenas na natureza do conteúdo, mas sobretudo em sua enorme acessibilidade — tanto para a produção quanto para a disseminação, que alcança todos os perfis de usuários e praticamente todas as plataformas.

Parte expressiva e estratégica dessas ferramentas opera em plataformas e redes sociais populares, como Telegram e Discord. A menção frequente a esses ambientes pelos jovens participantes da pesquisa, detalhada na seção 3, corrobora essa observação.

A monetização constitui um fator agravante central. Enquanto o objetivo primário da nudificação é, em grande parte dos casos, o lucro financeiro, outros espaços — como os da *dark web* — priorizam a exploração sexual como finalidade em si. Os crimes nesse ecossistema, portanto, articulam-se em torno de duas motivações principais: o interesse econômico e o próprio ato de explorar sexualmente crianças e adolescentes, mesmo quando mediados por IA.

Um agravante adicional da monetização diz respeito aos meios de pagamento: os aplicativos de nudificação aceitam modalidades amplamente difundidas, como PIX, carteiras digitais vinculadas a lojas de aplicativos e criptomoedas — estas últimas impulsionadas pelo crescimento das *exchanges* durante a pandemia —, operando diretamente em plataformas e aplicativos de uso cotidiano.

Entre os casos mais preocupantes estão aqueles que envolvem as Telegram Stars, moeda criada pela própria plataforma e de uso restrito ao seu ambiente interno. As Stars são adquiridas por meio dos métodos de pagamento já mencionados e facilitam o comércio internacional de conteúdos nudificados, *deepfake*s sexuais não consentidos e materiais artificiais de exploração sexual infantojuvenil.

O risco específico das Stars se intensifica porque, após um período de 21 dias, os chamados “criadores” — usuários que gerenciam serviços como *bots* dentro do Telegram — podem convertê-las em criptomoedas pela TON Network, blockchain desenvolvida pelo irmão do fundador do Telegram. Configura-se, assim, um circuito financeiro que permite a entrada de recursos por vias convencionais e sua posterior saída em ativos de difícil rastreamento.

O perfil das vítimas é heterogêneo, justamente em função da acessibilidade dessa tecnologia. A gratuidade ou o baixo custo dessas ferramentas tem permitido, por exemplo, que adolescentes produzam conteúdos nudificados de colegas de escola. Também é frequente o uso contra pessoas do convívio do agressor — namoradas, ex-namoradas, familiares —, mas igualmente comum o emprego contra desconhecidos, incluindo celebridades. Essa variedade de alvos reflete, em última instância, o fato de que a barreira técnica para a produção desse tipo de material foi praticamente eliminada.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Aplicativos, sites e robôs de nudificação são os primeiros vetores de risco mapeados na produção de material artificial de abuso ou exploração sexual infantojuvenil identificados nesta pesquisa.

HÁ DUAS TÉCNICAS PRINCIPAIS PARA CRIAR ESSAS IMAGENS ÍNTIMAS FALSAS SEM CONSENTIMENTO. NA NUDIFICAÇÃO, A IA REMOVE AS ROUPAS DE UMA PESSOA EM UMA FOTO COMUM, MANTENDO SEU ROSTO SOB UM CORPO NU GERADO DIGITALMENTE. NA TROCA DE ROSTO (FACE SWAP), OCORRE O INVERSO: SUBSTITUI-SE O ROSTO DE ALGUÉM EM UMA CENA PORNOGRÁFICA OU ERÓTICA PELO DA VÍTIMA.

O ato de "nudificar" pode ser feito através de produtos personalizados, como, por exemplo, robôs embutidos em plataformas, aplicativos e websites, mas sempre há de envolver uma pessoa real. Ou seja, a criação da nudificação sempre se baseia em uma ou mais mídias de uma pessoa real.

Entre os produtos personalizados capazes de criar esses conteúdos, os robôs de nudificação são os mais comuns e serão o foco deste capítulo. Geralmente, esses sistemas funcionam com uma interface simples de comandos para que o usuário solicite o tipo de mídia, posições sexuais e outras funcionalidades como inserção de áudio, no caso de vídeos. Os robôs podem estar disponíveis em redes sociais, aplicativos ou geradores de imagem hospedados na *surface web*. Em muitos casos, o usuário pode criar uma ou duas imagens de teste para, posteriormente, seguir com um maior número de criações em troca de pagamentos.

A origem dessas tecnologias remonta aos primeiros sistemas de IA capazes de realizar *face swap* (em português, troca de rosto. O termo se refere à substituição do rosto de uma vítima pelo de uma pessoa da indústria pornográfica). Com o tempo, novos sistemas foram treinados especificamente para contextos sexuais ou de abuso, dispensando até mesmo o rosto da vítima: passaram a alterar posições corporais e replicar feições a partir de fotografias comuns. O resultado são ferramentas capazes de transformar uma foto de perfil, uma imagem escolar ou qualquer registro cotidiano em material sexualmente explícito — em segundos, sem qualquer conhecimento técnico.

A preocupação com essa disseminação massiva por meio das ferramentas de nudificação não é exclusiva da SaferNet Brasil. No *Dia da Internet Segura*, 10 de fevereiro de 2026, o INHOPE — rede internacional de canais de denúncias da qual a SaferNet Brasil faz parte — emitiu **um documento** conjunto com organizações de proteção infantil de todo o mundo pedindo o banimento global de aplicativos de nudificação.[26]

"Ferramentas de nudificação usam IA para gerar imagens de nudez a partir de fotos vestidas. Embora comercializadas principalmente para adultos, estas ferramentas são frequentemente utilizadas indevidamente para criar imagens nuas não consensuais de pessoas e são cada vez mais utilizadas para produzir imagens ilegais de crianças", diz a nota, da qual a SaferNet Brasil foi signatária.

O documento afirma que essas ferramentas não têm nenhum uso positivo, sendo utilizadas exclusivamente para humilhar, assediar e explorar — cada vez mais para produzir imagens ilegais de crianças, frequentemente por outras crianças, além de estarem associadas a casos de extorsão sexual. **"As empresas, desenvolvedores, e os indivíduos que os criam ou distribuem devem ser responsabilizados e enfrentar con-**

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

sequências legais e criminais. A tecnologia que permite a criação de imagens de nudez e de abuso sexual têm um impacto insidioso na vida das crianças em todo o mundo. Imagens de crianças, incluindo sobreviventes de violência sexual na infância, estão sendo manipuladas de formas cada vez mais extremas e abusivas."

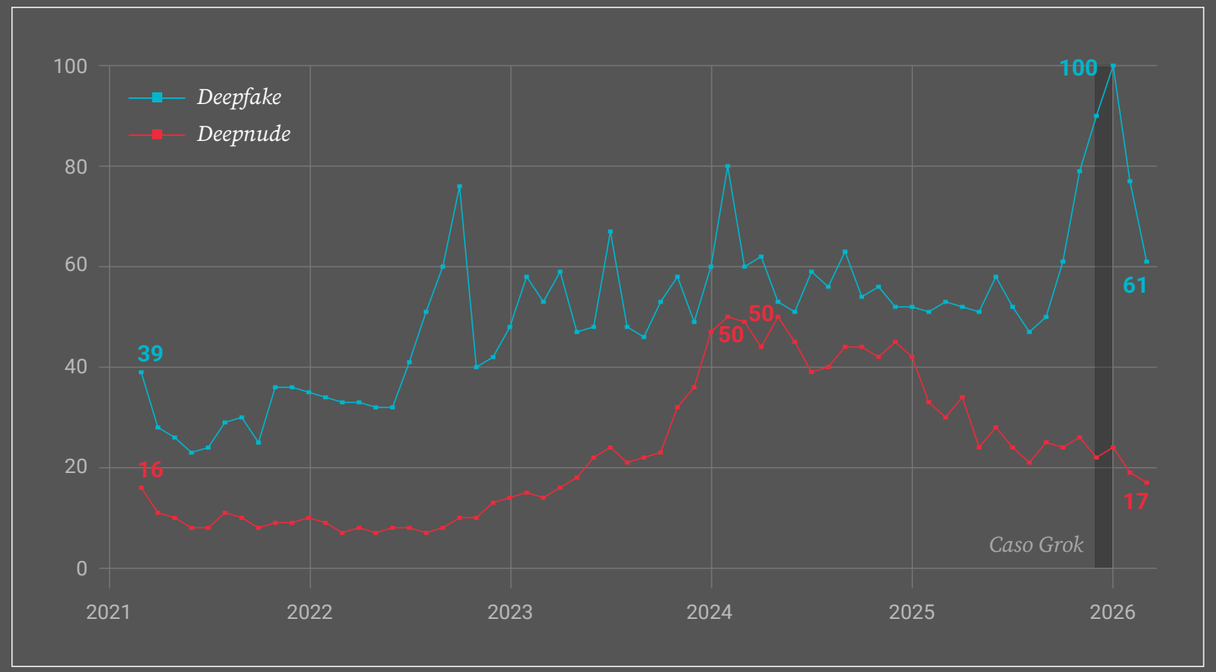
"Esta tecnologia não é apenas uma ferramenta, é um novo vetor de exploração, colocando mulheres e crianças — especialmente as meninas — em um risco sem precedentes. Esse tipo de exploração não têm lugar em nenhuma sociedade que valorize a segurança e a proteção da infância."

Embora plataformas de IA generalistas tenham investido progressivamente em salvaguardas, os aplicativos de nudificação — construídos sobre modelos de difusão de código aberto treinados em imagens pornográficas — removem esse obstáculo por completo. A barreira técnica, que antes funcionava como um limitador natural, deixou de existir: qualquer pessoa com acesso a uma fotografia e a um dispositivo conectado à internet é hoje capaz de produzir esse tipo de material. É essa combinação entre acessibilidade técnica, ausência de verificação de idade e ampla disponibilidade que torna os aplicativos de nudificação o vetor de maior preocupação no ecossistema analisado por esta pesquisa.

O ecossistema é amplo e fragmentado: plataformas na internet aberta, aplicativos móveis, robôs em serviços de mensagens e marketplaces alternativos coexistem e se retroalimentam. O modelo de negócio predominante é o freemium ou por créditos — a imagem de teste é gratuita, mas a remoção de marcas d'água, o aumento de resolução ou a produção em escala exigem assinaturas pagas ou moedas virtuais.

FIGURA 19

Gráfico com dados do Google Trends por pesquisas de brasileiros sobre termos como "*deepfakes*" e "*deepnudes*". Os dados abrangem o período de 1 de março de 2021 a 16 de março de 2026.



ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Dados do Google Trends confirmam a tendência: o volume atingiu picos em junho de 2024 e em janeiro de 2026 — este último coincidindo com o caso Grok. A média de buscas pelo termo *deepnude* no Brasil é de 21,87, e por *deepfake*, de 45,8, indicando interesse persistente ao longo de todo o período analisado.

O caso do Grok ilustra com precisão o que ocorre quando ferramentas de IA são integradas a plataformas de escala massiva sem salvaguardas adequadas.[27] O Grok é um robô de IA generativa criado pela xAI, empresa de desenvolvimento de IA e subsidiária da rede social do X (antigo Twitter). Ele possui uma versão de assinatura, mas antes poderia ser utilizado gratuitamente por usuários do X apenas marcando @grok em uma publicação; a funcionalidade foi limitada para somente usuários pagos após o caso. Em janeiro de 2026, usuários relataram que o Grok-2 produziu material ou conteúdo de violência sexual contra crianças e adolescentes a partir de *prompt*s com imagens reais de crianças publicadas no X, que conta com mais de 500 milhões de usuários mensais e um algoritmo orientado à viralidade.[28] A funcionalidade foi suspensa temporariamente pela xAI após alguns dias e representantes da xAI admitiram publicamente a existência de "falhas nos mecanismos de proteção" no Grok. A empresa reconheceu que os filtros de segurança foram insuficientes para impedir pedidos que resultaram em imagens de crianças e adolescentes em contextos sexuais ou de exploração.[29]

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

4.1.1. ESTUDO DE CASO
ROBÔS DE NUDIFICAÇÃO NO TELEGRAM

Os achados relativos aos aplicativos de nudificação — nos quais 100% das URLs analisadas apresentaram *deepfake*s de crianças e crimes contra crianças — evidenciam que essas ferramentas não são um vetor de risco entre outros, mas o elo mais diretamente associado à produção de conteúdo artificial de abuso sexual infantojuvenil. Parte significativa desse ecossistema opera por meio de robôs (*bots*) hospedados em plataformas de mensagens, que permitem a criação de imagens sintéticas de nudez a partir de fotografias reais, sem necessidade de conhecimento técnico por parte do usuário. É nesse contexto que o Telegram se impõe como estudo de caso incontornável: a arquitetura aberta de sua API (*Application Programming Interface*), a possibilidade de criação de robôs automatizados sem supervisão prévia e a postura historicamente refratária da plataforma à cooperação com autoridades criam um ambiente particularmente propício à proliferação dessas ferramentas.

Entre 2023 e 2026, a SaferNet Brasil recebeu mais de 10 mil denúncias de potenciais conteúdos de abuso ou exploração sexual infantil no Telegram — aplicativo de mensagens com mais de 950 milhões de usuários mensais. O dado, por si só, já seria suficiente para justificar atenção; mais revelador, contudo, é o sentido da curva: mesmo restrito ao primeiro trimestre de 2026, o mês de janeiro deste ano já havia superado em volume de denúncias todos os janeiros anteriores do período analisado — 249 registros, contra 203 em janeiro de 2025, 205 em 2024 e 212 em 2023.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

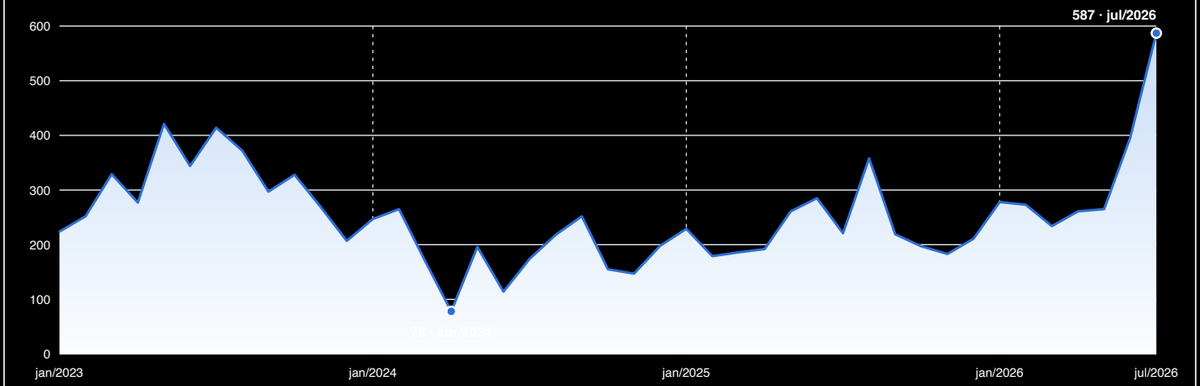
SOBRE ESTA PESQUISA

FIGURA 20

Distribuição mensal de denúncias à SaferNet Brasil sobre potenciais materiais de abuso ou exploração sexual infantil com o domínio "t.me/". Os dados abrangem o período de 1 de janeiro de 2023 a 30 de junho de 2026.

Evolução mensal das denúncias

por data da denúncia · base consolidada: URLs + entidades citadas nos comentários (14.782 registros datados até jul/2026)



Registros por mês e ano

	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez	Total
2023	224	252	329	277	421	344	414	372	297	328	269	207	**3.734**
2024	247	265	170	78	196	114	174	218	252	155	147	198	**2.214**
2025	229	179	186	192	261	285	221	358	219	197	183	211	**2.721**
2026	278	273	234	261	265	396	587	—	—	—	—	—	**2.294***

78 — 587 registros/mês — fora do recorte (ago–dez/2026)

Base = 14.782 registros datados (URLs + entidades citadas), por data da denúncia. * 2026 parcial (jan–jul); ago/2026 excluído por apuração incompleta.

Uma pesquisa conduzida pela SaferNet Brasil em outubro de 2024, no âmbito do projeto *DISCOVER*, aprofundou esse diagnóstico ao identificar centenas de canais e grupos na plataforma com conteúdo e comportamento criminoso que extrapolavam a comercialização e a distribuição de material ou conteúdo de violência sexual contra crianças e adolescentes. Entre os conteúdos encontrados estavam imagens íntimas compartilhadas sem consentimento, cenas de zoofilia e material de exploração sexual infantojuvenil gerado por IA — evidências de que o Telegram opera não apenas como repositório de conteúdo ilícito preexistente, mas como ecossistema ativo de produção e circulação de novas formas de violência digital.

Em agosto de 2024, o fundador e CEO do Telegram, Pavel Durov, foi detido no aeroporto de *Le Bourget*, em Paris, e formalmente indiciado pelas autoridades francesas por cumplicidade na distribuição de conteúdo de exploração sexual infantil, além de facilitação de crime organizado, tráfico ilícito e fraude. Um ano após sua prisão, Durov disse "preferir morrer" do

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

que autorizar terceiros a acessar mensagens privadas de usuários do seu aplicativo.[30] Ainda que, nos últimos anos, o Telegram tenha anunciado compromissos com a ampliação de seus esforços de moderação (estabelecendo parcerias com outros canais de denúncia, por exemplo, como a *Internet Watch Foundation*, o hotline do Reino Unido), o aplicativo permanece como um dos principais vetores de risco para atividade cibercriminosa, incluindo exploração sexual infantojuvenil.

A persistência e o crescimento de denúncias à SaferNet Brasil nesse âmbito indica que as medidas adotadas até o momento estão aquém da dimensão do problema. A tendência de aumento de atividade criminosa no Telegram também é corroborada pela análise das denúncias e pela literatura especializada sobre o tema há ao menos cinco anos.[31]

Em novembro de 2025, uma influenciadora narrou anonimamente à SaferNet Brasil ter sido vítima de *deepfake*s sexuais sem o seu consentimento. A jovem havia acabado de fazer 18 anos. As imagens fabricadas tiveram como base vídeos da influenciadora publicados na plataforma Twitch. Ela relatou: "Eu tinha 18 anos, mas ninguém garante que as pessoas que fizeram isso sabiam que eu tinha 18. E eu sei que se eu tivesse entrado em algum processo não teria dado em nada — tem várias situações que prejudicam a gente e essa é a situação mais grave e mesmo assim não dá em nada quando vai pra justiça."

A crescente atratividade do Telegram para atividades criminosas decorre de um conjunto de fatores estruturais que se reforçam mutuamente. A plataforma é amplamente percebida como segura e avessa à cooperação com autoridades — reputação construída ao longo de anos de resistência a pedidos governamentais de acesso a dados —, o que transmite aos agentes criminosos uma sensação de impunidade. Somam-se a isso a facilidade para criar ou ingressar em comunidades, a permissão para canais com até 200.000 usuários e a possibilidade de criar contas com números de telefone descartáveis, o que torna a rastreabilidade dos responsáveis extraordinariamente difícil.

O funcionamento desses grupos segue, no geral, um padrão deliberadamente fragmentado e com camadas de acesso progressivo. Canais públicos ou semipúblicos funcionam como vitrines e repositórios de links, nos quais usuários normalmente não compartilham mídias diretamente, mas direcionam o tráfego para comunidades privadas — frequentemente pagas — onde a troca de conteúdo ilícito de fato ocorre. Os usuários que operam como intermediários ou promotores costumam apagar suas contas com rapidez após a divulgação, aproveitando-se do fato de que o Telegram permite que um mesmo número de telefone seja utilizado para criar e recriar contas repetidas vezes, sem qualquer registro permanente.

Diferentemente de outras comunidades que compartilham material de exploração sexual infantil, como fóruns da *dark web* — tradicionalmente voltados à curadoria e distribuição de conteúdo já existente —, alguns grupos que consomem conteúdo "real" vêm incorporando uma dimensão adicional: a promoção e circulação de ferramentas de IA capazes de sinteti-

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

A PARTIR DAQUI, ALGUMAS IMAGENS FORAM TARJADAS POR CONTER CONTEÚDOS SENSÍVEIS, QUE SERÃO DISPONIBILIZADOS ÀS AUTORIDADES.

FIGURA 21

Grupo do Telegram denunciado à SaferNet Brasil por potencial conteúdo de abuso e exploração sexual infantil avaliado no âmbito desta pesquisa.



zar imagens. Nesse cenário, a dinâmica deixa de se limitar à troca e comercialização de acervos e passa a incluir a produção ativa de novos conteúdos. Um elemento distintivo, em comparação com ambientes mais restritos como a *dark web*, é a acessibilidade: plataformas como o Telegram são gratuitas e amplamente disponíveis a qualquer usuário com acesso a lojas de aplicativos ou a um computador, o que reduz barreiras de entrada e amplia o alcance dessas práticas. Essa mudança de paradigma eleva significativamente o risco, ao diminuir a dependência de acervos preexistentes e viabilizar a geração de material ilícito em escala.

No que diz respeito aos meios de pagamento, os conteúdos e as ferramentas ilícitas de nudificação no Telegram são comercializados majoritariamente por criptomoedas — com destaque crescente para o Monero, cuja arquitetura dificulta o rastreamento em comparação com o Bitcoin —, além de PayPal e, mais recentemente, pelas próprias *Telegram Stars* (Estrelas do Telegram). **Essa diversificação de meios de pagamento não é acidental: ela reflete um esforço deliberado de adaptação dos operadores a diferentes perfis de usuário, combinando a opacidade das criptomoedas com a conveniência de sistemas integrados à própria plataforma.**

As Estrelas são a moeda digital interna do Telegram, utilizadas para adquirir serviços de robôs ou jogos e para remunerar seus criadores. Somente Estrelas recebidas como pagamento por conteúdo ou serviços podem ser sacadas — Estrelas compradas por usuários para uso pessoal não são elegíveis para retirada. Para efetuar o saque, é necessário acumular um mínimo de 1.000 unidades e aguardar 21 dias para o recebimento. As Estrelas também podem ser empregadas para comprar anúncios no Telegram e itens colecionáveis ou convertidas em Toncoin — a criptomoeda da *The Open Network* (TON) — por meio da plataforma[32] Fragment. A TON é uma blockchain originalmente desenvolvida por Nikolai Durov, cofundador do Telegram e irmão de Pavel Durov, e é caracterizada por oferecer alto grau de privacidade e anonimato, com uma infraestrutura descentralizada e supervisão mí-

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

nima mesmo quando comparada aos padrões já permissivos do mercado de criptoativos.

Embora o Telegram tenha formalmente se afastado da TON em 2020, após um litígio da Comissão de Valores Mobiliários dos EUA (a SEC)[33], essa separação é cada vez mais tênue. Em maio de 2026, Pavel Durov anunciou que o Telegram substituirá a TON Foundation como principal força motriz da blockchain e se tornará o seu maior validador.[34] Na prática, isso significa que um operador de *bot* de nudificação pode receber pagamentos dentro do próprio ecossistema do Telegram, convertê-los em criptomoeda e sacá-los — tudo sem sair da infraestrutura vinculada ao aplicativo. O Telegram, aliás, retém uma comissão de 15% a 30% sobre cada saque de Estrelas, o que significa que lucra indiretamente com cada transação realizada por esses e outros operadores criminosos.

Uma avaliação dos termos de uso e de privacidade da Fragment em seu site — na versão mais recente, de abril de 2026, não deixa claro como a plataforma se diferencia do próprio Telegram além do nome empresarial Fragment Corp. como operadora, uma vez que todos os serviços oferecidos envolvem diretamente o aplicativo de mensagens. Os próprios termos de uso da Fragment confirmam essa arquitetura circular e revelam uma estrutura deliberadamente voltada ao anonimato. A empresa está registrada na República de Seychelles, jurisdição offshore conhecida por regulação financeira permissiva, sob o número de registro 221287. Seus termos de uso definem a plataforma como "não custodial", o que significa que ela não detém fundos nem controla (ou modera) carteiras de usuários, o que difere completamente de outras plataformas de compra e venda de criptoativos.

O processamento da compra de Estrelas no Brasil pode ser feito via PIX ou por meio das lojas de aplicativos (Google Play Store ou Apple App Store). Essas lojas retêm uma comissão de 30% do valor total; o Telegram aplica uma taxa administrativa que varia entre 4% e 5%; e o restante — aproximadamente 65% — é repassado ao desenvolvedor ou criador de conteúdo. Historicamente, Apple e Google impõem uma comissão de 30% sobre qualquer transação de bens e serviços digitais realizada dentro de aplicativos hospedados em suas plataformas. Essa política gerou disputas jurídicas globais de grande repercussão (como o caso *Epic Games vs Apple*).

No caso do canal exemplificado a seguir, cuja assinatura mínima é de 500 Estrelas (US$ 9,99), estima-se que o operador receba cerca de US$ 6,50 por assinante — valor que, multiplicado por centenas ou milhares de membros pagantes, gera receita recorrente significativa. Em um contexto de comercialização de material ilícito, esse modelo fragmenta o fluxo financeiro entre ao menos três camadas — loja de aplicativo, plataforma e destinatário final —, dificultando tanto o rastreamento quanto a responsabilização.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

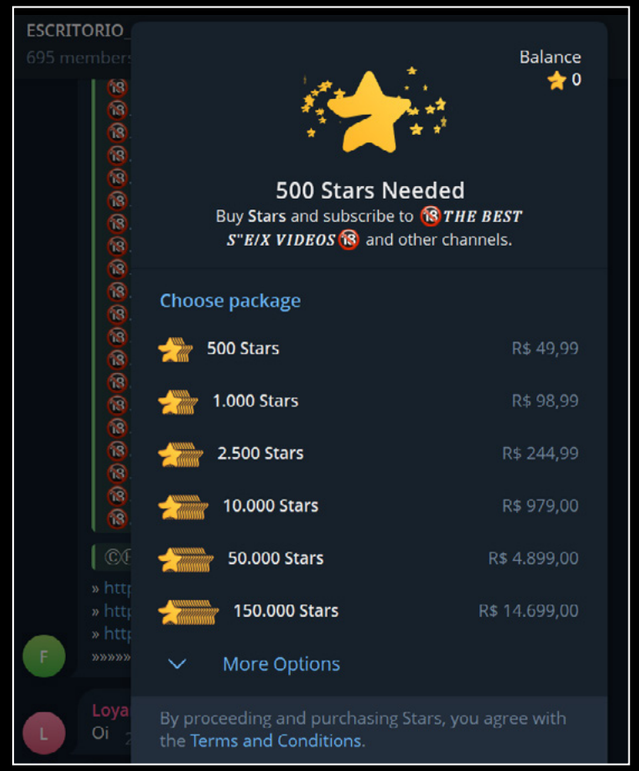
6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

FIGURAS 22 E 23

Canal de compartilhamento e venda de material real de abuso sexual infantil e conteúdos gerados por IA pede compra de *Telegram Stars* para acesso.



Star Amount	Price for Users	Apple / Google 30% Commission	Telegram Administrative Costs	Developer Proceeds
50	$0.99	$0.30	$0.04	**$0.65**
75	$1.49	$0.45	$0.07	**$0.98**
100	$1.99	$0.60	$0.09	**$1.30**
150	$2.99	$0.90	$0.14	**$1.95**
250	$4.99	$1.50	$0.24	**$3.25**
350	$6.99	$2.10	$0.34	**$4.55**
500	$9.99	$3.00	$0.49	**$6.50**
750	$14.99	$4.50	$0.74	**$9.75**
1000	$19.99	$6.00	$0.99	**$13.00**
1500	$29.99	$9.00	$1.49	**$19.50**
2500	$49.99	$15.00	$2.49	**$32.50**

Tabela retirada do FAQ do Telegram sobre as Estrelas apontando as principais taxas de comissão quando as moedas são compradas por intermédio das lojas de aplicativo.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Para lançar as *Telegram Stars* como a moeda interna oficial para o ecossistema de *Mini Apps*, robôs e conteúdos pagos, o Telegram foi obrigado a integrar os sistemas de pagamento *In-App Purchase* dessas corporações. No entanto, ciente de que a comissão de 30% limitaria sua margem de lucro e comprometeria parcela significativa da remuneração de criadores de conteúdo e desenvolvedores de robôs, o Telegram arquitetou e deliberadamente incentiva rotas alternativas de aquisição que barateiam o custo unitário da Star para o usuário final, conforme evidenciado ao lado.

FIGURA 24

Formas de Pagamento e Comissões para compra de Telegram Stars, segundo a documentação oficial da TON Wallet[35]

Forma de pagamento e origem	Intermediário principal	Comissões e taxas	Contexto e regras da documentação
Compra integrada (iOS)	Apple App Store	~30% de comissão (Taxa padrão da Apple para produtos digitais, já embutida no valor final do pacote).	A carteira TON Wallet redireciona o usuário para a interface de compra da App Store. Disponível em pacotes fixos de 100, 250 ou 500 Stars.
Compra integrada (Android)	Google Play Store	~30% de comissão (Taxa padrão da Google para produtos digitais, já embutida no valor final do pacote).	A carteira TON Wallet redireciona o usuário para a interface de compra da Google Play. Disponível em pacotes fixos de 100, 250 ou 500 Stars.
Transferência de saldo geral	Telegram Messenger	0% de taxa para a transferência interna.	Permite mover Stars que já estejam no saldo geral da conta do Telegram para o saldo da TON Wallet.
Alternativa direta (Fallback)	Rede Blockchain TON	Variável (Taxa de rede calculada dinamicamente baseada na volatilidade do Toncoin).	Caso o usuário não possua Stars na TON Wallet, a documentação aponta que o usuário pode alterar a opção de pagamento da taxa para TON (Toncoin) diretamente na tela antes de confirmar o envio ou swap.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Para compreender a engenharia financeira e regulatória que permite ao Telegram oferecer descontos significativos contornando o monopólio das lojas de aplicativos (Apple App Store e Google Play Store), é necessário analisar a arquitetura multicanal criada pela plataforma, que une soluções tradicionais de pagamento e a infraestrutura Web3 da blockchain TON (*The Open Network*).

A precificação das Telegram Stars varia de acordo com o nível de intermediação bancária e regulatória de cada canal de venda. Ao comprar Stars no cartão de crédito via app móvel, o KYC — *Know Your Customer*, prática regulatória desenhada para assegurar a identidade dos clientes, suas atividades financeiras e os riscos que eles podem representar — é herdado da conta do usuário na Apple Play Store ou Google Pay, que possuem cartões de crédito vinculados, nomes reais e CPFs/identidades rastreáveis). Para comprar Toncoin utilizando moedas como o Real (BRL) e depois gastar na Fragment, o usuário precisa transferir fundos de uma corretora (como Binance, OKX, Mercado Bitcoin). Essas corretoras aplicam KYC estrito (conferência de documentos e biometria facial) por força de regulação bancária.

Essas transações, geralmente feitas apenas no mobile, estão submetidas a níveis maiores de escrutínio e conformidade regulatória, ao contrário das transações efetuadas fora da interface mobile que se valem do ecossistema expandido do Telegram formado pela Fragment, pelo @PremiumBot e por diversas plataformas de gift cards, marketplaces de jogos (como Kinguin, LootBar, GGSel) e sites de recarga terceirizados (MooGold, Joytify) que operam por meio de um modelo paralelo conhecido como arbitragem de ecossistema.

O grande diferencial desse método é que ele não exige as credenciais de login (e-mail ou senha) do comprador, demandando exclusivamente o arroba (@username) da conta do Telegram. O funcionamento desse mercado paralelo, desde a compra até a entrega do saldo, é idêntico ao processo de compra de créditos para um jogo on-line (como Free Fire ou Roblox).

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Aproximadamente, a estrutura de custos por unidade de Star se desenha da seguinte forma:

FIGURA 25

Comparativo de custos e mecanismos de desoneração por canal de compra.

Canal de compra	Meio de pagamento	Preço estimado por Star	Mecanismo de desoneração ou de desconto
In-App (iOS / Android)	Apple Pay / Google Pay	~$0,025 (Preço Base)	Nenhum. Sofre a incidência integral da comissão de 30% das Big Techs.
@PremiumBot (Versões Direct/Web)	Cartão de Crédito / Processadores locais	~$0,021 (~15% a 20% de desconto)	Parcial. Elimina os 30% das lojas mobile, mas ainda arca com taxas de adquirentes de cartão (Visa/ Mastercard).
Fragment.com	Criptoativos Toncoin ($TON)	~$0,015 (Até 40% de desconto)	Total. Desintermediação bancária e de Big Techs. Transação direta peer-to-peer na blockchain, com possibilidade de ofuscamento da origem e pseudo animização das transações.

O funcionamento é simples: o usuário acessa o site, escolhe um pacote de Stars (de 50 a 100 mil unidades), digita apenas seu @nomedeusuário do Telegram e paga com Pix, boleto, PayPal ou cartão de crédito local. Nenhuma senha do Telegram é solicitada. Nenhuma verificação de identidade é exigida para compras de baixo valor. Por trás da tela, o que acontece é o seguinte: assim que o pagamento é aprovado, o sistema do vendedor — muitas vezes automatizado por robôs integrados à API da Fragment — acessa a plataforma, insere o @username fornecido e liquida a transação na blockchain TON. O Telegram reconhece o pagamento e credita as Stars na conta do comprador de forma quase instantânea. [detalhes disponíveis às autoridades — A-01]

Desse modo, a arquitetura híbrida do Telegram (que mistura a Web2 do aplicativo e dos *bots* com a Web3 da blockchain TON e da Fragment), é estruturada com o objetivo de promover a desintermediação bancária, facilitando assim a prática de ilícitos através de ofuscamento da origem financeira, permitindo a um criminoso utilizar Stars para comprar material ilícito ou serviços de IA

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

generativa voltados ao crime. Para fins didáticos, vejamos o seguinte exemplo: ao adquirir Stars no Kinguin pagando com Pix, o extrato bancário do usuário mostrará apenas uma transação legítima para uma empresa intermediadora de pagamentos (gateways de e-commerce). Não haverá qualquer menção ao Telegram ou a criptoativos em seus registros bancários tradicionais.

Como esses sites não exigem senhas ou KYC para compras de baixo valor, o criminoso consegue abastecer sua conta do Telegram com fundos digitais sem precisar expor cartões de crédito pessoais em lojas oficiais como Google Play ou App Store. A investigação, nesse caso, passaria necessariamente pelo marketplace (como o Kinguin) pois é esta entidade que faz a custódia dos dados cadastrais, meios de pagamento e metadados associados (IP, e-mail, timestamp, etc) do usuário que comprou aquele pacote específico para o @username sob investigação. Essa camada adicional não apenas onera forças de segurança, como de fato pode retardar ou até inviabilizar investigações de diversos crimes se o marketplace não tiver representação legal no país, o que ensejaria pedido de cooperação internacional.

As compras financeiras de Stars através das lojas de aplicativos (Apple e Google) são finais e não reembolsáveis, conforme os Termos de Serviço do Telegram. No entanto, se o usuário transferiu um pacote de Stars do seu saldo do Telegram para a TON Wallet e não as utilizou, a plataforma permite realizar o estorno desse pacote completo (de 100, 250 ou 500 Stars) de volta para a sua conta geral do Telegram por meio do histórico de transações. Não são permitidos estornos parciais. Essa feature evidencia a existência de um sistema de *billing* e registro interno das transações. Dentro da TON Wallet, as Telegram Stars funcionam apenas para mitigar/pagar taxas de transação da rede e não podem ser convertidas ou trocadas (*swapped*) por outras criptomoedas dentro da interface da carteira, criando um sistema hermeticamente fechado.

Ao permitir que usuários comprem Stars pelas vias oficiais das lojas (mesmo que mais caro), o Telegram mantém seus aplicativos ativos e em conformidade com as diretrizes da Apple e do Google. Quando um usuário gasta suas Stars dentro de um jogo ou canal do Telegram, o desenvolvedor recebe essas Stars. O Telegram permite que esses criadores saquem seus ganhos convertendo as Stars de volta para Toncoin ($TON) pela própria Fragment.com.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

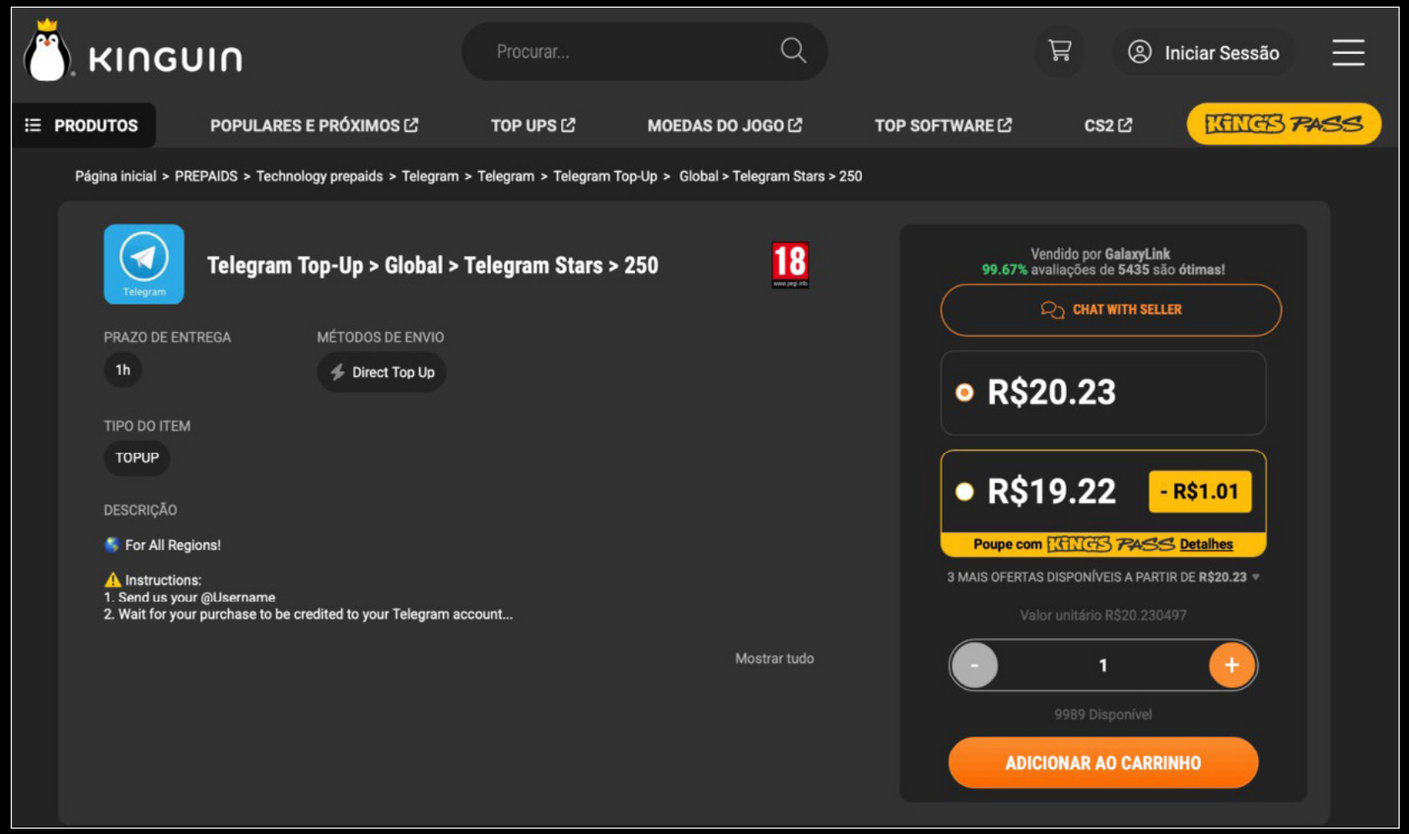
6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

FIGURA 26

Captura de tela do site Kinguin com oferta de Estrelas do Telegram a partir de cartões-presente.



Ao notar que comprar Stars na Fragment utilizando TON gera uma economia de até 40%, o usuário comum é financeiramente incentivado a aprender a usar uma carteira de criptomoedas, retroalimentando e valorizando o ecossistema da blockchain TON. Em suma: as regras das Big Techs são aplicadas apenas na "porta de entrada" convencional (*mobile*), enquanto o verdadeiro volume financeiro e de desconto é direcionado para a sua própria infraestrutura descentralizada e soberana na Web3.

Na avaliação da SaferNet Brasil, o modelo de negócio dos grupos de exploração sexual infantil vem se consolidando em torno de assinaturas recorrentes — em vez de transações isoladas —, o que gera receita previsível para os operadores. Os valores frequentemente não ultrapassam US$ 100 mensais para usuários estrangeiros ou R$ 50 para usuários brasileiros (valor máximo observado ao longo desta análise, ainda que variável), criando uma barreira de entrada baixa que amplia o alcance dessas redes. A lógica é a mesma de qualquer serviço digital por assinatura: fidelização do usuário, previsibilidade de caixa e escala.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Embora a maioria dos operadores comercialize ferramentas pagas, há também aqueles que oferecem robôs de nudificação gratuitamente. Esses *bots* operam, em geral, no modelo freemium — permitindo a criação de algumas imagens sem custo ou limitando a geração a uma mídia por dia —, enquanto produções adicionais requerem a compra de moedas ou créditos internos. A gratuidade parcial não é um gesto de abertura: é uma estratégia de conversão, desenhada para criar familiaridade com a ferramenta e conduzir o usuário ao pagamento.

Muitos dos robôs também permitem o chamado check-in diário para acúmulo de moedas virtuais — mecanismo pelo qual o usuário precisa acessar o *bot* todos os dias para receber uma bonificação em créditos. Essa dinâmica, inspirada em práticas comuns de engajamento em jogos eletrônicos e aplicativos de fidelidade, cumpre uma função estratégica no contexto criminoso: garante o retorno recorrente do usuário à plataforma, aumenta o tempo de permanência nas comunidades ilícitas, dificulta o abandono do ecossistema e, ao distribuir créditos gratuitos de forma periódica, reduz a fricção para que o usuário consuma — e eventualmente pague por — conteúdos de maior gravidade. O resultado é um ciclo de engajamento que reproduz, com precisão, as técnicas de retenção empregadas pela indústria de tecnologia, mas a serviço da exploração sexual.

No escopo principal desta pesquisa — a produção e distribuição de conteúdos criminosos em escolas brasileiras por adolescentes —, esse modelo adquire contornos particularmente preocupantes. Adolescentes, em geral, não dispõem de renda própria para adquirir ou assinar ferramentas digitais sem o conhecimento de seus responsáveis, o que os torna o público mais suscetível a ofertas gratuitas ou baseadas em mecânicas de acúmulo progressivo. A combinação entre gratuidade inicial, gamificação e acessibilidade técnica configura, na prática, uma porta de entrada desenhada sob medida para quem tem menos de 18 anos — ainda que seus criadores eventualmente aleguem desconhecer o perfil etário de seus usuários.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS
2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS
3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO
4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA
5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA
6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES
7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

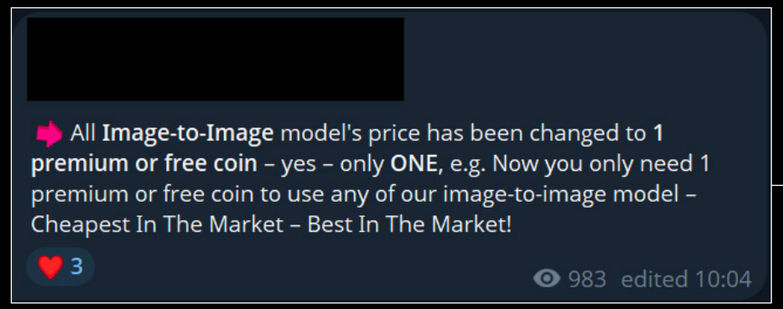


FIGURA 27

Mensagem compartilhada por robô de nudificação no Telegram promovendo novos modelos de criação de *deepfakes* sexuais.



FIGURA 28

Apresentação de *bot* de nudificação no Telegram: exemplos de *deepfakes* sexuais, compra de créditos e minijogo para obter créditos gratuitos.

Em abril de 2026, a organização europeia *AI Forensics* publicou relatório com base na análise de 2,8 milhões de mensagens em 16 grupos do Telegram ao longo de seis semanas, documentando um ecossistema de abuso em escala que alcançou aproximadamente 52 mil pessoas na Itália e na Espanha.[36] O relatório identificou que arquivos contendo imagens íntimas não consensuais — incluindo material ou conteúdo de violência sexual contra crianças e adolescentes — eram vendidos por valores entre 20 e 50 euros, e que alguns usuários atuavam como afiliados de *bots* de nudificação com comissões de até 40%. Segundo a organização, o Telegram viabiliza esse ecossistema por design: grupos fechados, uma vez removidos, eram recriados com o mesmo nome em questão de horas.

Em escala global, o relatório da *Internet Watch Foundation* de 2026 registrou 8.029 imagens e vídeos gerados por IA avaliados como material realista de abuso sexual infantil ao longo de 2025, com um aumento de 26.385% nos vídeos em relação ao ano anterior — de 13 em 2024 para 3.443.[37] Do total de vídeos identificados, 65% foram classificados como Categoria A — a classificação mais grave segundo a legislação britânica, que abrange atos sexuais penetrativos, sadismo e práticas com animais.

O CONJUNTO DESSAS EVIDÊNCIAS — PRODUZIDAS POR ORGANIZAÇÕES DE DIFERENTES PAÍSES, COM METODOLOGIAS INDEPENDENTES — CONVERGE PARA UMA MESMA CONCLUSÃO: O TELEGRAM NÃO É APENAS UMA PLATAFORMA ONDE O ABUSO ACONTECE, MAS UMA INFRAESTRUTURA QUE, POR SUA ARQUITETURA TÉCNICA, SEU MODELO DE MONETIZAÇÃO E SUA POSTURA DE MODERAÇÃO, ESTRUTURA E VIABILIZA A PRODUÇÃO, A DISTRIBUIÇÃO E A COMERCIALIZAÇÃO DE MATERIAL OU CONTEÚDO DE VIOLÊNCIA SEXUAL CONTRA CRIANÇAS E ADOLESCENTES EM ESCALA INDUSTRIAL.

O segundo vetor de risco identificado na produção de material artificial de abuso ou exploração sexual infantojuvenil é a *dark web*. Embora a internet aberta permaneça como o principal meio pelo qual usuários — com mais ou menos de 18 anos — acessam as ferramentas de IA generativa necessárias à produção de conteúdos ilegais, é indispensável compreender o papel dessa camada da internet na crise em curso.

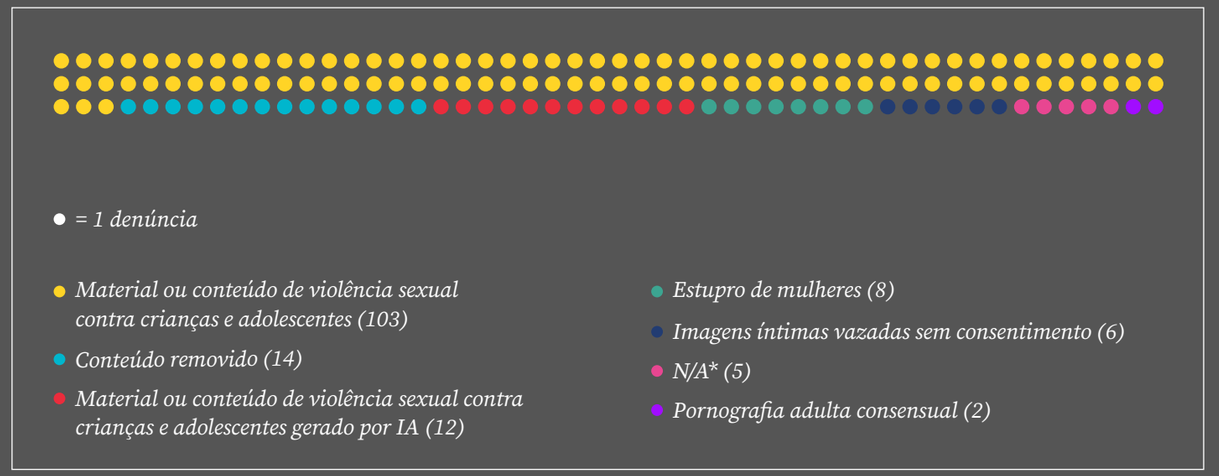
Essa relevância decorre de ao menos três fatores. Primeiro, apesar de ser menos acessada por usuários comuns, é na *dark web* que surgem muitos dos sistemas de IA generativa mais avançados, capazes de gerar imagens com maior grau de realismo e violência. Segundo, é nesse ambiente que organizações criminosas estruturadas operam com capacidade de movimentar volumes financeiros significativos oriundos de cibercrimes, atravessando múltiplas jurisdições — dinâmica já amplamente documentada historicamente. Terceiro, a convergência entre o desenvolvimento de ferramentas ilícitas e o elevado realismo dos conteúdos ali produzidos amplia, na avaliação da SaferNet Brasil, significativamente o potencial de dano dessas operações.

A dimensão do problema é evidenciada por dados recentes. Em março de 2025, uma operação coordenada pela Europol e liderada pela polícia da Baviera, na Alemanha, resultou no desmantelamento do *Kidflix*, uma das maiores plataformas de exploração sexual infantil já identificadas na *dark web*, com aproximadamente 1,8 milhão de usuários registrados em mais de 35 países. A plataforma hospedava cerca de 91 mil vídeos únicos, com uma média de 3,5 novos uploads por hora, e operava um sistema de recompensas baseado em criptomoedas que incentivava a publicação de conteúdo abusivo.

Para dimensionar esse potencial, a SaferNet Brasil selecionou todos os domínios da rede Tor denunciados à central de denúncias de conteúdos de abuso ou exploração sexual infantil no mês de setembro de 2025. Ao todo, foram analisados 127 links — 108 originalmente denunciados à central e outros 19 identificados durante a própria análise. Entre os domínios

FIGURA 29

Distribuição da análise da SaferNet Brasil sobre links da rede Tor denunciados ao canal de denúncias no mês de setembro de 2025.



ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

examinados, destacam-se fóruns de discussão frequentados por indivíduos autodeclarados pedófilos e ferramentas voltadas à criação de imagens de abuso sexual infantil. Do total de URLs ".onion" analisadas, 81,89% (104) continham material real de abuso ou exploração sexual infantil e 10,24% (13) continham material artificial da mesma natureza. As categorias não são excludentes: oito URLs apresentaram os dois tipos de material simultaneamente.

Para aprofundamento, foram selecionados dois domínios específicos como estudos de caso neste trecho da pesquisa. O primeiro é um fórum dedicado à produção e disseminação de ferramentas de IA generativa e conteúdos artificiais de abuso sexual infantil; o segundo, um gerador de imagens por IA gratuito hospedado na *dark web*, voltado exclusivamente à criação desses mesmos conteúdos criminosos.

O QUE ENCONTRAMOS NESTE VETOR DE RISCO

- Comunidades altamente estruturadas e organizadas em sites e fóruns oferecem diferentes formas de "apoio" aos seus membros.
- Isso inclui suporte psicológico e emocional entre pares para persuadir membros, novos ou antigos, de que a produção de imagens de exploração sexual infantil utilizando IA não é tão grave por ser artificial, além de fornecerem sistemas e ferramentas para esse fim.
- Diálogos entre cibercriminosos indicam que a crescente crise de materiais artificiais de abuso e exploração sexual infantil pode dificultar a atuação das autoridades de proteção à infância. Isso inclui o risco de sobrecarregar os serviços de investigação, ajuda e proteção, além de potencialmente atrapalhar a localização de vítimas reais em situações de emergência.
- Comunidades criminosas envolvidas na produção, distribuição e comercialização de material ou conteúdo de violência sexual contra crianças e adolescentes estão atualmente segmentando os websites entre aqueles que veiculam conteúdo "real" e aqueles que veiculam e/ou produzem conteúdo artificial (*deepfakes*). Observa-se uma dissonância cognitiva entre os usuários que consomem ambos os tipos. Alguns defendem que o material artificial representa o "futuro" da exploração sexual e a "solução para seus transtornos", pois acreditam que podem satisfazer seus desejos sem causar dano real a uma criança. Em contraste, os consumidores de conteúdo "real" manifestam oposição a essa visão.
- O perfil das vítimas é diverso e varia de acordo com o propósito do site (como fórum de discussão ou gerador de imagens). No entanto, é recorrente a criação de "personas" de vítimas inexistentes de abuso sexual infantil por usuários nesses ambientes da *dark web*. Nesses casos, são construídas personalidades, histórias e traços específicos para essa criança ou adolescente fictício, e em seguida é produzida uma vasta quantidade de conteúdo diferente envolvendo essa pessoa irreal. É também comum que outros usuários acompanhem, sigam e interajam com essas criações.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

4.2.1. ESTUDO DE CASO

FÓRUM DA DARK WEB DEDICADO À CRIAÇÃO DE CONTEÚDOS DE ABUSO E EXPLORAÇÃO SEXUAL DE CRIANÇAS E ADOLESCENTES

O primeiro estudo de caso neste escopo se trata de um fórum da *dark web [detalhes disponíveis às autoridades — A-02]* criado com o propósito exclusivo de incentivar e compartilhar materiais gerados por IA retratando erotização, abuso ou exploração sexual de crianças e adolescentes.

A SaferNet Brasil começou a monitorar o fórum em outubro de 2025. A primeira publicação do site data de 4 de setembro de 2025. Com base em indicadores qualitativos — como o número de publicações semanais e o ranqueamento de postagens mais votadas pela comunidade —, estima-se que, no início do monitoramento, o fórum contasse com 2 a 4 mil usuários mensais ativos. Em março de 2026, publicações individuais já ultrapassavam 20 mil visualizações, o que sugere um crescimento expressivo da base de usuários em menos de seis meses.

O fórum é estruturado de forma organizada e didática. As seções são divididas por gênero das vítimas retratadas, por modalidade de criação do conteúdo criminoso — como "txt2vid" (texto para vídeo) e "img2img" (imagem para imagem), padrões previamente documentados na seção 1 deste relatório — e por tipos de modelos e dispositivos para utilização de IA generativa. Há ainda seções dedicadas a instruções sobre como evitar a detecção por autoridades.

As **regras do fórum** são construídas em torno de uma retórica pseudocientífica que sustenta que a criação de imagens artificiais de exploração sexual infantil seria uma alternativa ética, supostamente respaldada pela ciência, para impedir abusos no mundo real. Conteúdos artificiais graficamente violentos — como representações de tortura — e o comércio de imagens ou de modelos de IA são explicitamente proibidos. Na avaliação da SaferNet Brasil, essas regras não refletem qualquer preocupação genuína com a proteção de crianças: foram desenhadas com o propósito de ajudar os administradores a evadir responsabilidade legal em jurisdições onde materiais de IA que retratam abuso, sexualização ou exploração sexual infantojuvenil não são expressamente criminalizados, além de fortalecer um senso de pertencimento baseado em crenças distorcidas de que tais práticas seriam moralmente defensáveis.

REGRAS E DIRETRIZES COMUNS

Amamos crianças, até mesmo crianças com IA.
As crianças podem não ser reais, mas ainda assim queremos tratá-las com a dignidade e o respeito que todas as crianças merecem. A IA permite-nos explorar fantasias que podem não ser possíveis ou realistas na vida real, mas isso não significa magoá-las ou degradá-las. Uma criança feliz é uma criança sorridente.

REGRAS DE CONTEÚDO

O conteúdo (imagens, vídeos, histórias) deve apresentar indivíduos com menos de 18 anos. Exceções são feitas para discussões técnicas ou demonstrações.

Este site é para arte artificial (AI/Art/CGI), conteúdo real só deve ser postado para auxiliar na criação dessa arte. *[detalhes disponíveis às autoridades — A-03]*

Gostamos de ficar fora da política, evitar quaisquer temas ou conteúdos políticos que possam ser considerados controversos.

Nenhum conteúdo envolvendo bestialidade/sexo em zoológico.

Nenhum conteúdo em que as crianças estejam claramente em perigo.

As crianças não deveriam estar em alguma situação de tortura.

Sem estupro.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

SEGURANÇA

Os modelos devem estar em formatos seguros, como tensores de segurança. Formatos menos seguros (pickle, etc.) não são permitidos.

Atividades de compra, venda ou negociação e quaisquer sites envolvidos em tais atividades não devem ser compartilhados aqui.

É proibido falar ou compartilhar links ou contas de redes sociais.

Discussões sobre aplicativos de chat também não são permitidas.

Nunca compartilhe informações sobre sua localização, idade ou qualquer outro detalhe que possa ser usado para identificá-lo.

Arquivos com material sobre o tema devem ser protegidos por senha.

Nenhuma solicitação de imagem, vídeo ou outros geradores de IA on-line.

O fórum oferece suporte apenas à geração de IA local (offline).

COMPORTAMENTO GERAL

Sites envolvidos com atividades como lesão/stresscore, chantagem, etc., não podem ser discutidos ou vinculados neste fórum. Isso envolve gateways de links que optam por direcionamento

O objetivo declarado do fórum, segundo seus administradores, é fornecer um autoproclamado "espaço seguro" para a produção e disseminação desses conteúdos. Em uma postagem comemorativa da passagem de 2025 para 2026, um dos administradores escreveu: "Atualmente estamos numa fase desafiadora, que é bastante prejudicial para pessoas como nós. Temos procurado outros grupos que compartilhem nossas ambições em relação ao amor e respeito às crianças, mas até agora não encontramos nenhum." Na mesma publicação, ao discorrer sobre planos futuros, o mesmo usuário criticou operações de autoridades internacionais contra fóruns de material ou conteúdo de violência sexual contra crianças e adolescentes e afirmou estar reforçando a infraestrutura de cibersegurança da plataforma: "Continuaremos a manter nosso próprio domínio nesta comunidade e a colaborar com diversas fontes externas para garantir que nossos usuários estejam seguros e possam viver vidas felizes com sua sexualidade."

Parte do que sustenta a retenção de usuários recorrentes e atrai novos membros é essa retórica coletiva — impulsionada pelos administradores — de que suas ações estariam contribuindo para combater a violência sexual contra crianças no mundo real. Em outra publicação, o mesmo administrador escreveu: "Nosso objetivo é minimizar o abuso originado em nosso grupo e acreditamos em relacionamentos amorosos e sustentáveis, onde todos estão seguros. [...] Trabalhar com psiquiatras profissionais é um processo lento, mas continuará. Ainda temos ambições de oferecer acesso seguro a ferramentas profissionais para usuários que sentem que precisam de ajuda." A construção discursiva é deliberada: ao empregar vocabulário terapêutico e afetivo, o fórum normaliza o consumo de material de exploração sexual infantil e cria um senso de pertencimento comunitário que busca "normalizar" as condutas de seus membros.

Ao contrário de outros mercados de conteúdo ilícito envolvendo crianças e adolescentes, o fórum não opera com base em contrapartidas financeiras entre os administradores e os usuários. Em publicação de março de 2026, o administrador

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

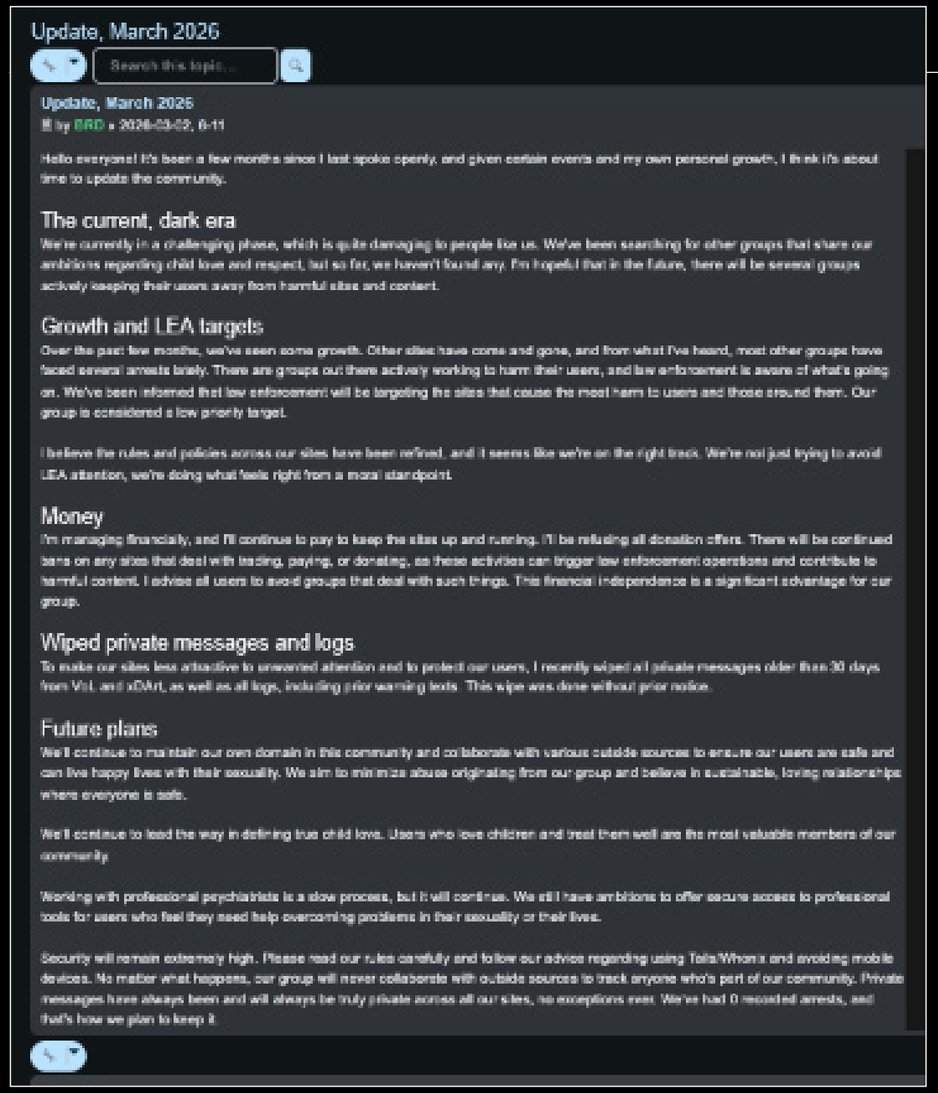
4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA



Update, March 2026

Search this topic...

Update, March 2026

Hello everyone! It's been a few months since I last spoke openly, and given certain events and my own personal growth, I think it's about time to update the community.

The current, dark era

We're currently in a challenging phase, which is quite damaging to people like us. We've been searching for other groups that share our ambitions regarding child love and respect, but so far, we haven't found any. I'm hopeful that in the future, there will be several groups actively keeping their users away from harmful sites and content.

Growth and LEA targets

Over the past few months, we've seen some growth. Other sites have come and gone, and from what I've heard, most other groups have faced several arrests lately. There are groups out there actively working to harm their users, and law enforcement is aware of what's going on. We've been informed that law enforcement will be targeting the sites that cause the most harm to users and those around them. Our group is considered a low priority target.

I believe the rules and policies across our sites have been refined, and it seems like we're on the right track. We're not just trying to avoid LEA attention, we're doing what feels right from a moral standpoint.

Money

I'm managing financially, and I'll continue to pay to keep the sites up and running. I'll be refusing all donation offers. There will be continued bans on any sites that deal with trading, paying, or donating, as these activities can trigger law enforcement operations and contribute to harmful content. I advise all users to avoid groups that deal with such things. This financial independence is a significant advantage for our group.

Wiped private messages and logs

To make our sites less attractive to unwanted attention and to protect our users, I recently wiped all private messages older than 30 days from V-L and xDArt, as well as all logs, including prior warning texts. This wipe was done without prior notice.

Future plans

We'll continue to maintain our own domain in this community and collaborate with various outside sources to ensure our users are safe and can live happy lives with their sexuality. We aim to minimize abuse originating from our group and believe in sustainable, loving relationships where everyone is safe.

We'll continue to lead the way in defining true child love. Users who love children and treat them well are the most valuable members of our community.

Working with professional psychiatrists is a slow process, but it will continue. We still have ambitions to offer secure access to professional tools for users who feel they need help overcoming problems in their sexuality or their lives.

Security will remain extremely high. Please read our rules carefully and follow our advice regarding using Tails/Whonix and avoiding mobile devices. No matter what happens, our group will never collaborate with outside sources to track anyone who's part of our community. Private messages have always been and will always be truly private across all our sites, no exceptions ever. We've had 0 recorded arrests, and that's how we plan to keep it.

FIGURA 30
Publicação de administrador(es) do fórum declarando manter toda a estrutura financeira do fórum pessoalmente.

principal declarou arcar individualmente com todos os custos da plataforma e reafirmou que o comércio não é permitido — segundo ele, para evitar atrair a atenção de autoridades.

O crescimento do conteúdo artificial de abuso sexual infantil tem, paradoxalmente, gerado conflitos dentro das próprias comunidades criminosas.

Grupos originais de exploração sexual infantojuvenil na *dark web* vêm priorizando materiais mais violentos e banindo usuários que utilizam IA generativa. Esse movimento pode ser exemplificado pelo relato de um usuário recém-registrado no fórum: “Sou criador aqui e em outros lugares. [...] Eu estava em um lugar onde eles ficaram cada vez mais restritivos com a IA, a ponto de ser praticamente só nudez [de crianças e adolescentes].” Em resposta, um moderador do fórum explicou que a principal proibição no fórum recai sobre representações de violência física explícita e orientou o usuário a consultar os guias e manuais sobre criação de imagens com IA disponíveis no fórum.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

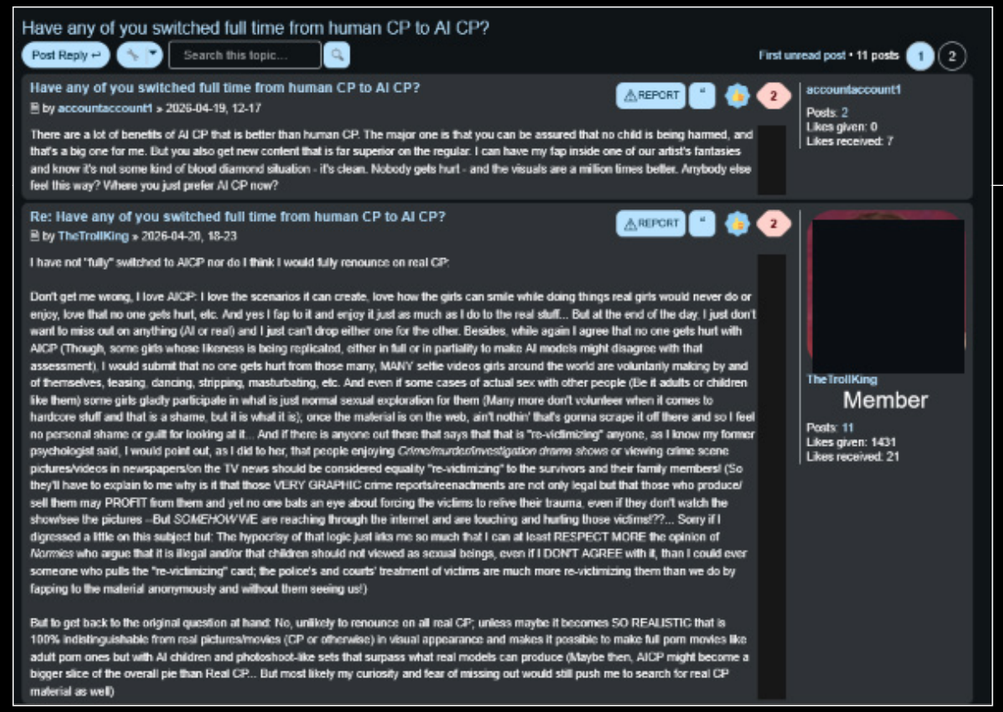
4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

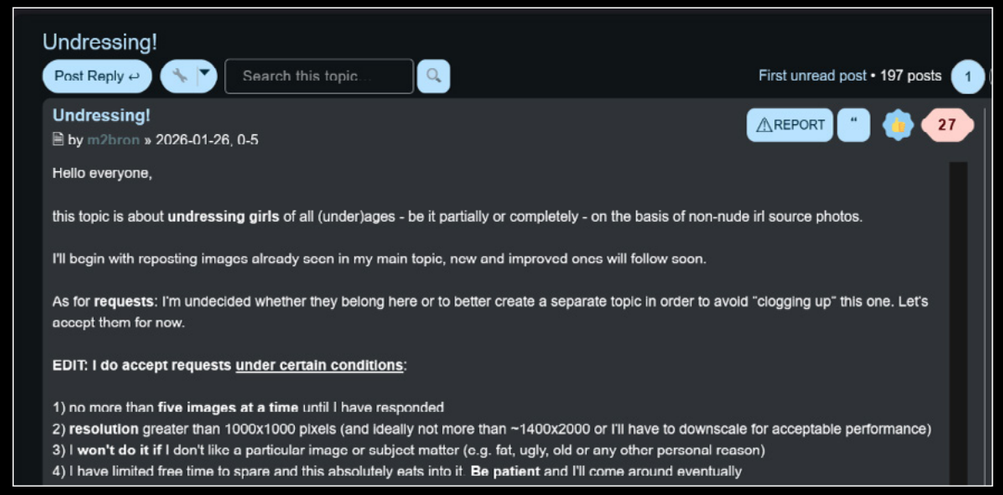
5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA





FIGURAS 31 E 32
Publicação de usuários do fórum discutindo se algum membro da comunidade substituiu o consumo de material real de abuso sexual infantil por conteúdos gerados por IA.

É precisamente nesse ponto que reside uma das principais preocupações da SaferNet Brasil: **o surgimento, nesses espaços, de comunidades dedicadas ao aprendizado e ao aperfeiçoamento sistemático de técnicas de produção de materiais artificiais de exploração sexual infantil.** O fórum mantém um repositório extenso com centenas de publicações — guias, manuais e tutoriais — que ensinam desde a geração de textos ficcionais descrevendo encontros sexuais entre adultos e adolescentes até o treinamento de modelos de IA dedicados à criação de conteúdo sob demanda retratando abuso sexual infantil. Existem múltiplas seções em que usuários experientes orientam novos membros, muitos dos quais se descrevem como curiosos ou consumidores regulares de materiais reais, mas interessados em passar a produzir conteúdo próprio.

Em uma das publicações, um usuário iniciante pediu orientação sobre como começar, e as respostas indicaram caminhos de aprendizado disponíveis na internet aberta, inclusive vídeos em plataformas populares. *[detalhes disponíveis às autoridades — A-04]* A naturalidade com que o aprendizado é incentivado e o fato de que ferramentas acessíveis na internet aberta — incluindo tutoriais no YouTube — são indicadas como ponto de partida evidenciam a porosidade entre a superfície da internet e esses ecossistemas criminosos.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

FIGURAS 33, 34 E 35

Publicação no fórum sobre como treinar e aprimorar um modelo de IA generativa para exploração sexual infantil sem gastar recursos.



A seção “txt2vid” (texto para vídeo) representa um risco particularmente crítico. Indivíduos envolvidos nessas práticas vêm aprimorando modelos capazes de gerar vídeos de abuso sexual infantil com níveis cada vez mais elevados de realismo. Embora parte dos materiais analisados ainda apresente sinais de artificialidade — como saturação e contraste excessivos, proporções anatômicas imprecisas ou transições abruptas entre cenas —, muitos já atingem alto grau de verossimilhança. Esses arquivos são frequentemente indexados e rotulados como “vídeos musicais” (na sigla em inglês, *MVs*), possivelmente como estratégia para facilitar sua disseminação em ambientes fora da *dark web*, como redes sociais e aplicativos de mensagens.

Em outra publicação, um usuário compartilhou doze vídeos gerados por IA representando abuso e exploração sexual infantil, incluindo material em CGI (*computer-generated imagery*), técnica de produção de imagens, animações e vídeos tridimensionais por meio de software gráfico, amplamente utilizada no cinema e em videogames, mas que, nesse contexto, é empregada para criar representações fotorrealistas de cenas que não existem no mundo físico. O autor explicou: “Este vídeo não tem necessariamente a intenção de ser emocionante, na verdade, pode ter o efeito oposto. Por que fizemos isso? Para mostrar as capacidades da IA e o que ela pode trazer para o mundo da CP [sigla em inglês para material ou conteúdo de violência sexual contra crianças e adolescentes].

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA



FIGURAS 36, 37 E 38

Publicação no fórum sobre como aprender a nudificar imagens com uso de IA (segunda ocorrência), usuários conversam sobre práticas de exploração e/ou abuso sexual infantojuvenil e "idades preferidas" para vítimas.

O realismo, como ela pode transportá-lo para um mundo diferente, como você quase pode sentir e cheirar o ambiente." Nos comentários, outros usuários elogiaram o realismo dos vídeos. "Simplesmente UAU! Onde isso vai parar? Ou talvez não pare?", perguntou um deles. A resposta do criador foi inequívoca: "Fim? Oh... isso é apenas o começo."

O caso do fórum ilustra uma inflexão qualitativa no ecossistema de exploração sexual infantil on-line. Não se trata mais apenas de distribuição de material preexistente, mas da consolidação de comunidades organizadas em torno da produção industrial de conteúdo sintético, com divisão de funções, transferência de conhecimento entre membros e aprimoramento contínuo de técnicas.

A retórica de que imagens artificiais protegem jovens reais — refutada pela literatura especializada e pelo próprio comportamento observado nos fóruns — serve, na prática, como mecanismo de recrutamento e retenção. Na avaliação da SaferNet Brasil, essas comunidades também buscam normalizar conteúdos artificiais que contenham erotização, sexualização, abuso ou exploração sexual de crianças e adolescentes com a ambição de mudar o que é moralmente aceito em nossa cultura.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

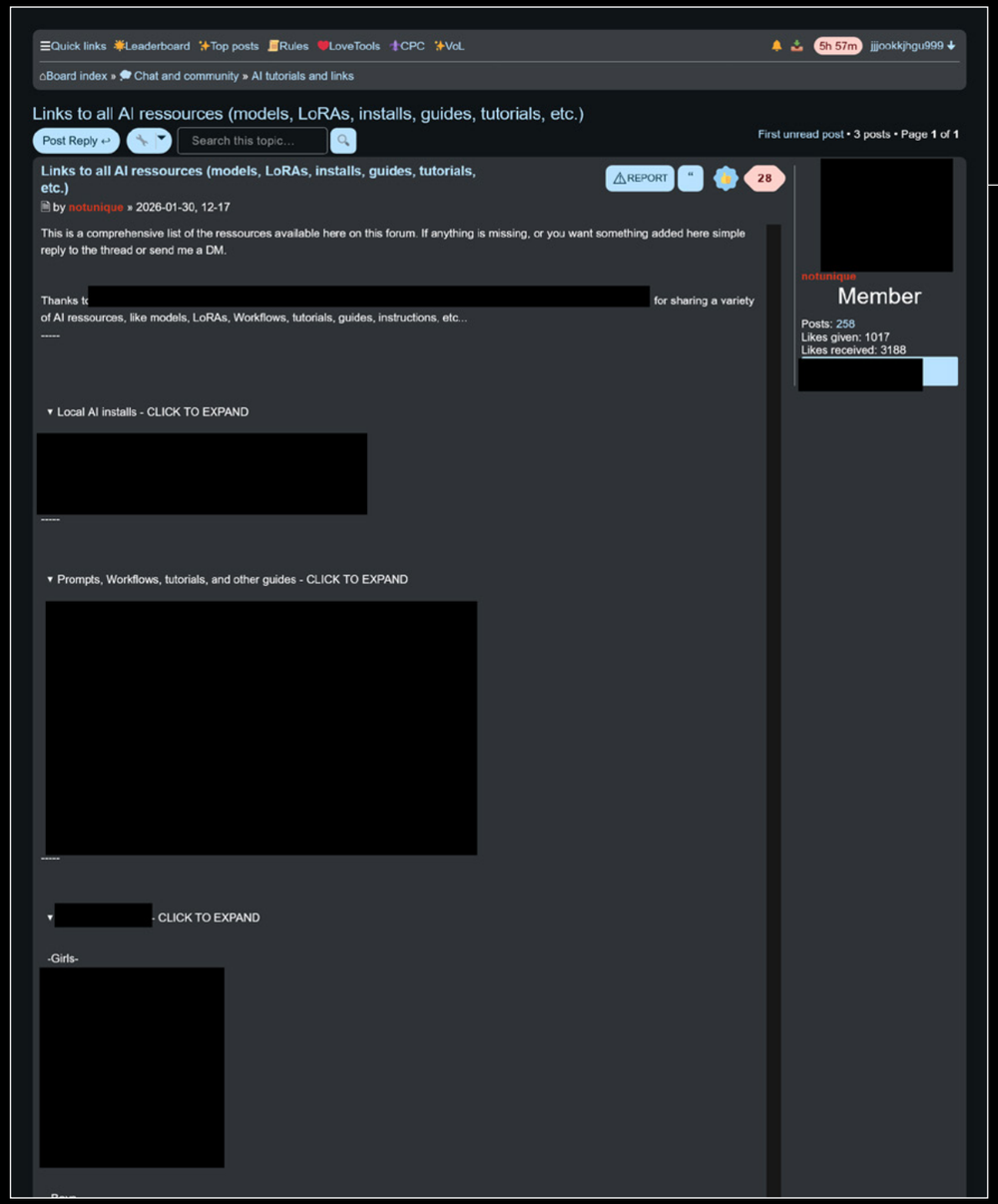
4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA





FIGURAS 39 E 40

Publicação no fórum com repositório de plataformas, links, prompts, tutoriais e guias para criação de imagens de abuso sexual de crianças e adolescentes.

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

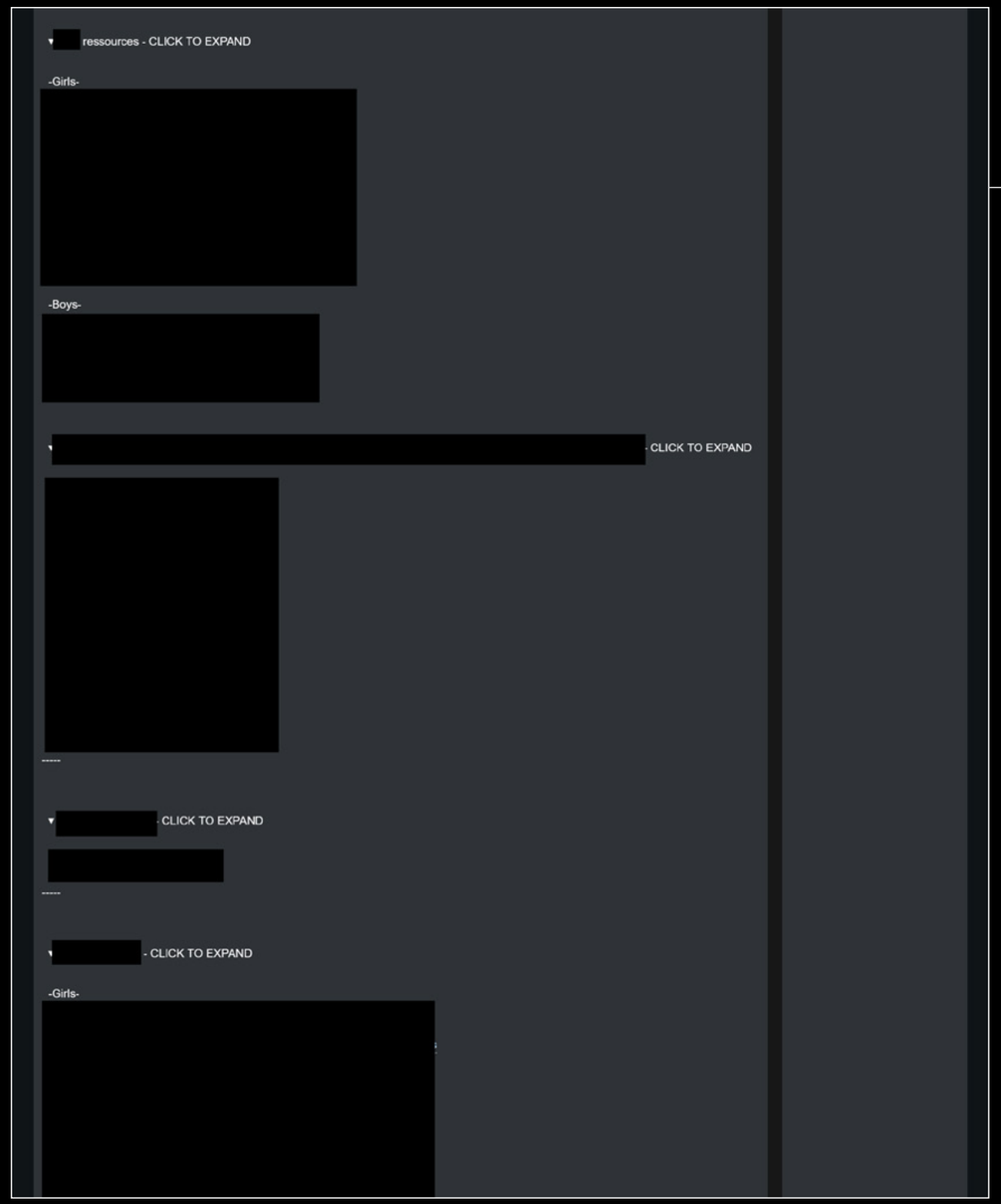
4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA



FIGURAS 41 E 42

Publicação no fórum com repositório de plataformas, links, prompts, tutoriais e guias para criação de imagens de abuso sexual de crianças e adolescentes. (continuação).



ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

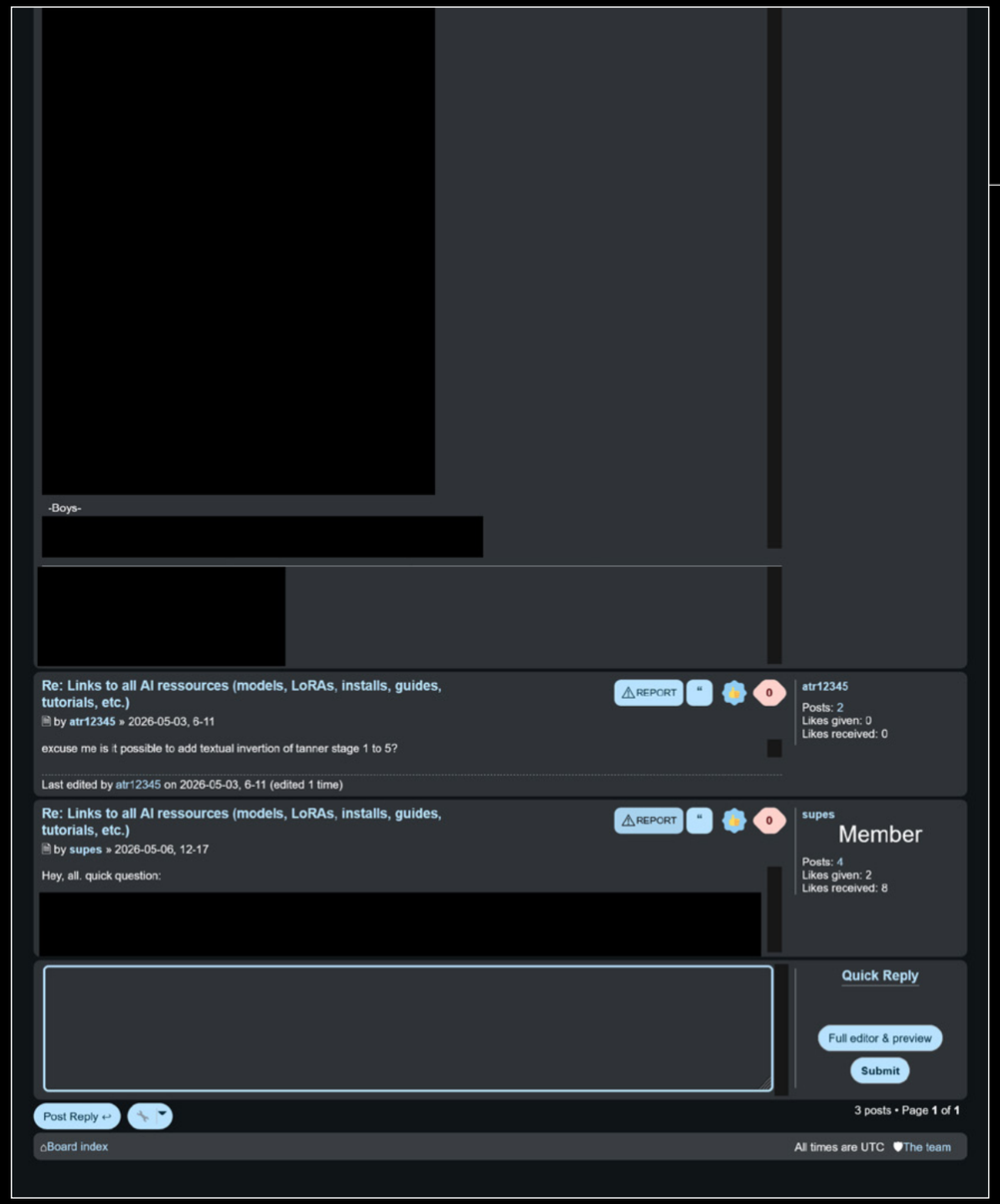


FIGURA 43

Leaderboard de principais prompts de criações contendo abuso ou exploração sexual infantojuvenil (continuação).

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

4.2.2. ESTUDO DE CASO

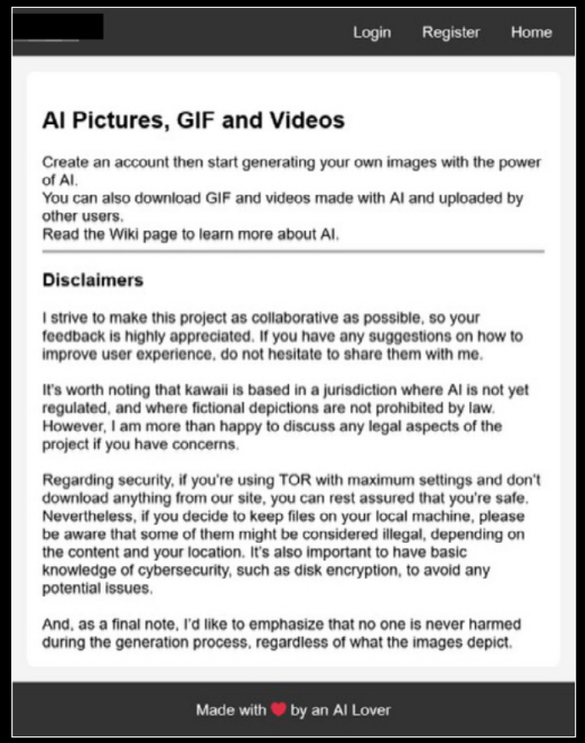
GERADOR DE IMAGEM POR IA PARA ABUSO SEXUAL INFANTIL

O segundo domínio selecionado para esta análise é um gerador de imagens baseado em IA generativa dedicado exclusivamente à produção de material ou conteúdo de violência sexual contra crianças e adolescentes. O site não possui nome comercial, razão pela qual será referido como "o gerador" nesta seção. *[detalhes disponíveis às autoridades — A-05]*

A documentação técnica do site detalha as ferramentas utilizadas, todas gratuitas e de código aberto, que qualquer pessoa pode baixar sem custo ou rastreamento. Todo o processo ocorre no computador do próprio usuário, sem servidor central que possa ser desligado. O site também ensina técnicas para acelerar a geração de vídeos e indica softwares profissionais gratuitos para edição e montagem. *[detalhes disponíveis às autoridades — A-06]*

FIGURA 44

Página principal do gerador de imagens.



ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

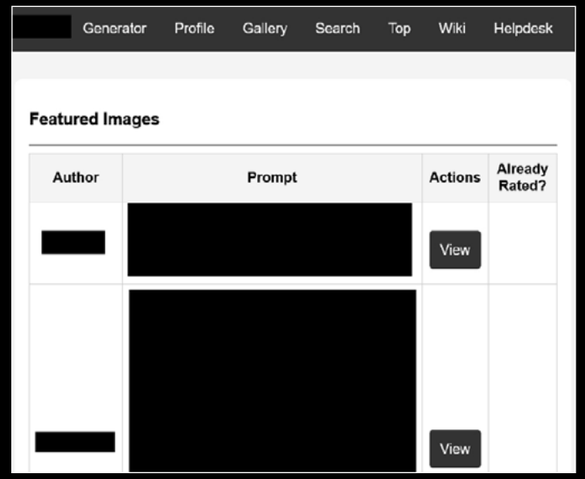
7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

O acesso à plataforma é deliberadamente anônimo: o cadastro exige apenas nome de usuário e senha, sem e-mail ou qualquer outro identificador. Após gerar uma imagem inicial, o usuário obtém acesso à galeria de conteúdos do site e passa a acumular créditos ao avaliar — positiva ou negativamente — as mídias criadas por outros usuários. Na avaliação da SaferNet Brasil, esse mecanismo configura uma estratégia de aprendizado por reforço com feedback humano (*Reinforcement Learning from Human Feedback* — RLHF), nos mesmos moldes do que ocorre no treinamento de *chatbots* de IA generativa comerciais. O dado é grave: indica que o site está ativamente utilizando o engajamento de sua comunidade de usuários para aprimorar o próprio modelo de geração de material ou conteúdo de violência sexual contra crianças e adolescentes.

FIGURA 45

Exemplos de *prompts* que geraram imagens de exploração sexual infantil com a palavra "*girl*" no estudo de caso.





ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

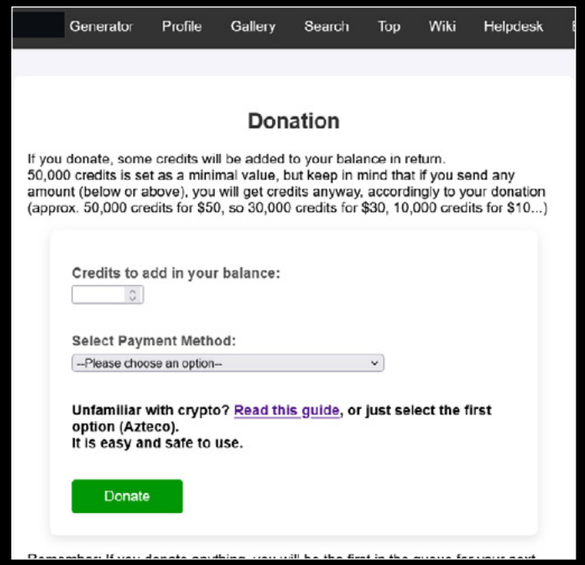
SOBRE ESTA PESQUISA

No que diz respeito ao financiamento, a criação de imagens não exige necessariamente contrapartida econômica — basta a avaliação de mídias já existentes —, embora o site ofereça a opção de doações para ampliar a quantidade de créditos disponíveis.

Os pagamentos são aceitos exclusivamente em criptomoedas, e a cada solicitação é informada uma carteira diferente. Uma das modalidades permite que o comprador use meios convencionais — como cartão de crédito ou serviços de pagamento digital — para adquirir créditos convertidos em criptomoeda antes de chegar aos operadores. *[detalhes disponíveis às autoridades — A-07]*

FIGURA 46

Página para doações e/ou compra de créditos para gerar imagens de abuso ou exploração sexual infantojuvenil com IA generativa.



O modelo de financiamento opera, portanto, em duas frentes com graus distintos de rastreabilidade. Nas transações diretas em criptomoeda, a rotação sistemática de carteiras reduz significativamente a capacidade de rastreamento, exigindo técnicas avançadas de análise de blockchain. Já na modalidade intermediada pela Azteco, a dependência de meios de pagamento convencionais na etapa de aquisição do cartão-presente constitui um ponto de vulnerabilidade explorável: as transações ficam registradas junto às operadoras de pagamento e podem ser vinculadas a dados cadastrais do comprador — o que representa, para as autoridades, uma das poucas vias rastreáveis em todo o ecossistema analisado.

Embora não seja possível estimar o volume total de transações ou o número de usuários recorrentes, a pesquisa buscou dimensionar o alcance da ferramenta por meio de funcionalidades públicas do site, como o número de resultados associados a *prompt*s específicos. Em março de 2026, o gerador já havia produzido mais de 510 mil materiais artificiais de abuso sexual infantil contendo a palavra “girl” (menina, em inglês).

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

FIGURA 47

Exemplos de prompts que geraram imagens de exploração sexual infantil com a palavra "*girl*" no estudo de caso.



Outra estratégia adotada pela SaferNet Brasil foi monitorar o crescimento do volume de mídias geradas com a palavra "Brazil". Em 20 de janeiro de 2026, a busca pelo termo retornou 4.401 registros; em 22 de fevereiro, 5.419; e em 20 de março, 6.056 — um crescimento de 37,6% em 60 dias. Isso equivale a aproximadamente 1.655 novas mídias retratando o abuso sexual de crianças e adolescentes brasileiras no período, ou cerca de 27,6 conteúdos por dia. A intensidade de uso é tal que, no intervalo de escrita deste parágrafo — cerca de um minuto —, outras duas novas imagens com o *prompt* "brazil" foram criadas.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

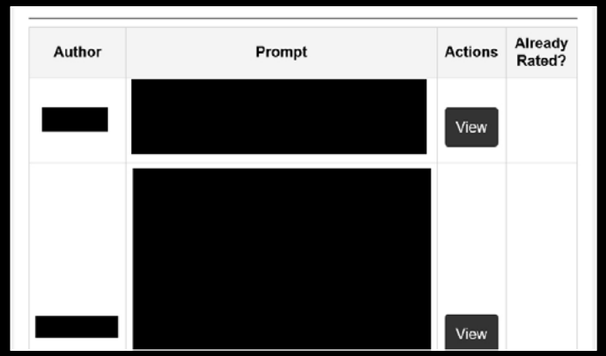
7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

FIGURAS 48

Capturas de tela realizadas em 20 de janeiro de 2026, 22 de fevereiro de 2026 e 20 de março de 2026 com resultados de pesquisa sobre o termo "*brazilian*".







992

PROMPTS COM PALAVRAS "BRAZIL" OU "BRASIL" ANALISADOS PELA SAFERNET.

Para compreender a diversificação de idiomas e as idades das crianças representadas, a SaferNet Brasil salvou manualmente 992 *prompt*s contendo as palavras "Brasil" ou "Brazil". A contagem por idioma identificou 937 registros em inglês, 29 em português e 26 em espanhol — predomínio que se explica, em parte, pelo fato de o gerador processar com maior precisão comandos formulados em inglês.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Em relação às idades, a SaferNet Brasil aplicou uma fórmula de extração numérica aos *prompt*s salvos, obtendo dados em 889 dos 992 comandos analisados. As idades mais frequentes foram 12 anos (111 menções), 10 anos (109), 7 anos (107), 11 anos (87) e 6 anos (84) — revelando que a maior parte dos conteúdos retrata crianças entre 6 e 12 anos. Chama atenção a presença de idades como 1, 2 e 3 anos nos *prompts*, com 22, 20 e 65 menções respectivamente, indicando que bebês e crianças muito pequenas também são alvos dessas produções criminosas. A concentração de menções na faixa de 6 a 12 anos, combinada com a presença não desprezível de idades inferiores a 3 anos, evidencia que o gerador atende a um espectro amplo de demandas — das mais "convencionais" dentro desse ecossistema às mais extremas.[38]

Algumas das imagens geradas mencionam pessoas reais. Em um dos casos, uma autoridade pública brasileira foi retratada em cena de abuso sexual junto a uma criança de sua família. *[detalhes disponíveis às autoridades — A-08]* O mesmo usuário responsável por essa imagem — com nome de perfil em português — havia gerado outras 130 mídias criminosas de abuso e exploração sexual infantil na plataforma.

FIGURA 49

Usuário do gerador cria imagens de abuso sexual infantil retratando pessoa pública brasileira.



ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

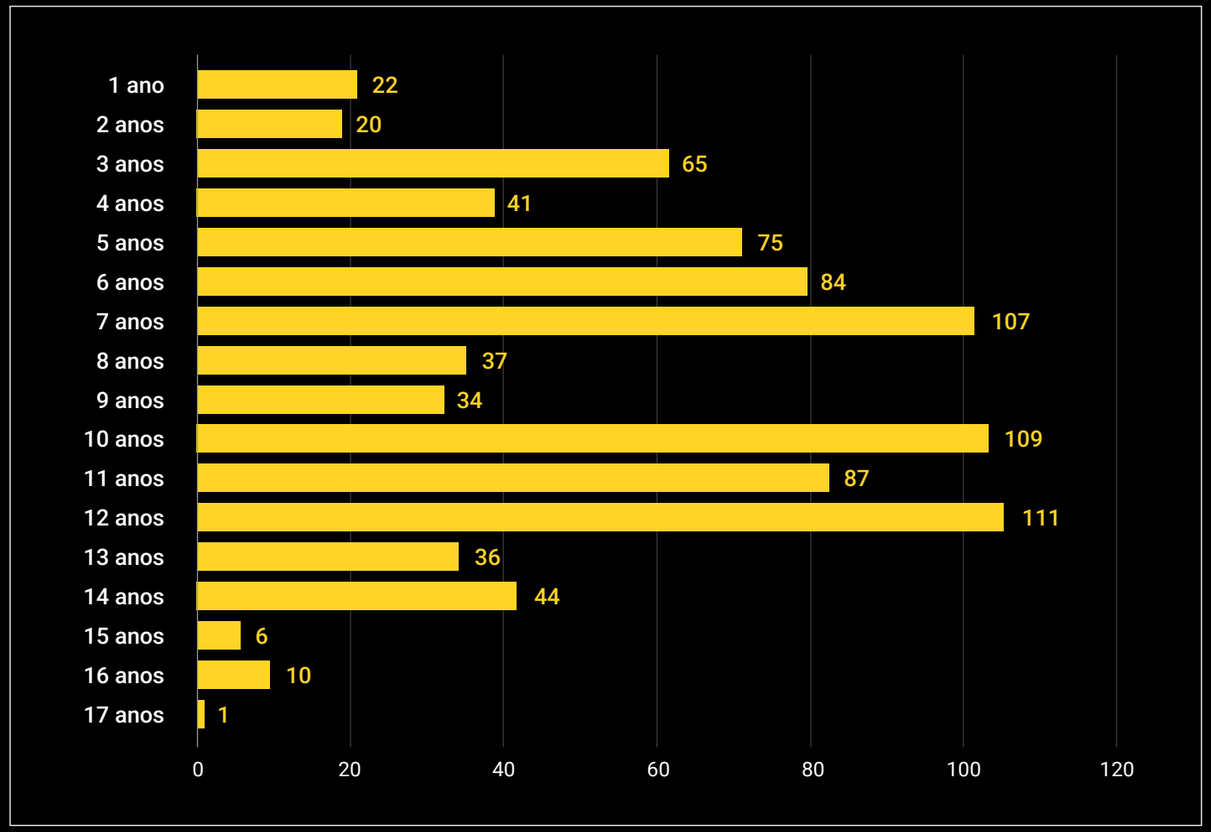
6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

FIGURA 50

Distribuição de idade mencionada nos *prompt*s coletados pela SaferNet Brasil.



O gerador analisado representa, na avaliação da SaferNet Brasil, um estágio de maturação técnica e operacional que distingue qualitativamente a ameaça atual daquelas observadas em anos anteriores. O volume de mais de 510 mil imagens com a palavra “girl” e o crescimento de 37,6% em 60 dias nos conteúdos referentes ao Brasil são indicadores de uma capacidade de produção industrial que já superou, em escala e velocidade, qualquer possibilidade de resposta manual por parte de autoridades ou organizações de proteção à infância. Se o fórum ilustra a consolidação de comunidades organizadas em torno do aprendizado criminoso, este gerador ilustra o próximo passo lógico dessa cadeia: a automação da produção em escala, operada por ferramentas que qualquer usuário com conhecimentos básicos pode utilizar — e que, a cada avaliação de imagem, se tornam mais precisas.

4.3. VETOR DE RISCO N.º3 → PLATAFORMAS DE PORNOGRAFIA

As ferramentas de IA desempenham um papel central na produção e proliferação de conteúdos de exploração sexual infantojuvenil; contudo, o problema transcende a criação. Em grande parte dos casos analisados e monitorados pela SaferNet Brasil, a infraestrutura de distribuição e monetização desses materiais recai sobre sites de pornografia adulta — plataformas que não apenas hospedam e exibem *deepfake*s sexuais e conteúdos artificiais de exploração sexual infantojuvenil, mas funcionam como amplificadores de alcance e visibilidade e, mais criticamente, obtêm lucro substancial por meio de sua exibição, da publicidade associada ou de modelos de assinatura. O fenômeno não se restringe a sites de nicho: plataformas como XVideos e Xnxx — dois dos três maiores sites de pornografia do mundo, com mais tráfego que Wikipedia, Amazon e Reddit — hospedam *deepfake*s sexuais e não demonstram tratar com seriedade a moderação de conteúdo não consensual.[39] Apesar de esforços pontuais de moderação de material gerado por IA, a aplicação das regras permanece inconsistente e conteúdos ilícitos frequentemente escapam ao escrutínio.

Esta seção se dedica à análise comparativa de duas plataformas de conteúdo adulto. O propósito não é isolar ou condenar sites específicos, mas demonstrar, de forma factual, que a presença e a exploração desses conteúdos constituem um problema sistêmico que afeta o setor como um todo — e que esses ambientes digitais contêm e, em muitos casos, dependem, direta ou indiretamente, da presença de *deepfake*s sexuais e de materiais de exploração sexual infantil gerados por IA.

Merece registro, ainda, uma dimensão que extrapola o escopo principal deste relatório, mas que os dois estudos de caso tornaram impossível ignorar: o uso de *deepfake*s sexuais como instrumento de violência contra mulheres adultas com vidas públicas — vereadoras, prefeitas, deputadas, senadoras e ativistas. Nos dois sites analisados, foram identificados *deepfake*s sexuais de deputadas e ex-deputadas federais e estaduais, publicados com nomes completos e facilmente localizáveis pelos mecanismos de busca internos das plataformas. Em nenhum dos casos o conteúdo havia sido moderado, removido ou sua indexação impedida. *Deepfake*s sexuais contra mulheres em posições de poder não são apenas uma violação à intimidade — são instrumentos de intimidação política que operam pela humilhação pública e pelo silenciamento, e cuja impunidade é viabilizada pela mesma infraestrutura que hospeda material de exploração sexual infantojuvenil.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

O QUE ENCONTRAMOS NESTE VETOR DE RISCO

O canal de denúncias da SaferNet Brasil acumula aproximadamente 8,6 mil registros de potencial conteúdo de exploração e abuso sexual infantojuvenil apenas na plataforma do primeiro estudo de caso (seção 4.3.1) desde 2018, com crescimento exponencial: de 3 denúncias em 2018 para 2.108 em 2025.

Na amostra qualitativa de 112 URLs da primeira plataforma, 27% dos links ainda disponíveis continham material ou conteúdo de violência sexual contra crianças e adolescentes, e 5,4% continham materiais artificiais de exploração sexual de adolescentes — demonstrando um padrão contínuo ao longo de sete anos. Pesquisas internas nesta plataforma por termos como "*AI teen*" retornaram mais de 5 mil resultados, com vídeos acumulando mais de 80 mil visualizações.

Na plataforma do, segundo estudo de caso (4.3.2), foram identificados 263 conteúdos gerados por IA sem consentimento envolvendo 22 vítimas; em 19 dos 22 casos, as vítimas tinham menos de 18 anos no momento da captura das imagens originais ou ainda são adolescentes.

Ambas as plataformas operam sem verificação efetiva de idade — em violação ao ECA Digital — e promovem serviços de transmissão ao vivo com categorias que exploram a aparência jovem das produtoras de conteúdo.

Táticas de ofuscação de nomes de vítimas adolescentes (uso de números para substituir letras e evitar moderação) foram documentadas na segunda plataforma, indicando esforço deliberado para evadir tanto a moderação automatizada quanto a detecção pelas próprias vítimas.

Em ambas as plataformas, foram identificados *deepfakes* sexuais de autoridades públicas brasileiras, publicados com nomes completos e localizáveis pelos mecanismos de busca internos — sem que qualquer moderação ou remoção tivesse sido realizada, o que evidencia que a mesma infraestrutura que hospeda material de exploração infantojuvenil viabiliza também a intimidação política de mulheres com vidas públicas.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

4.3.1. ESTUDO DE CASO

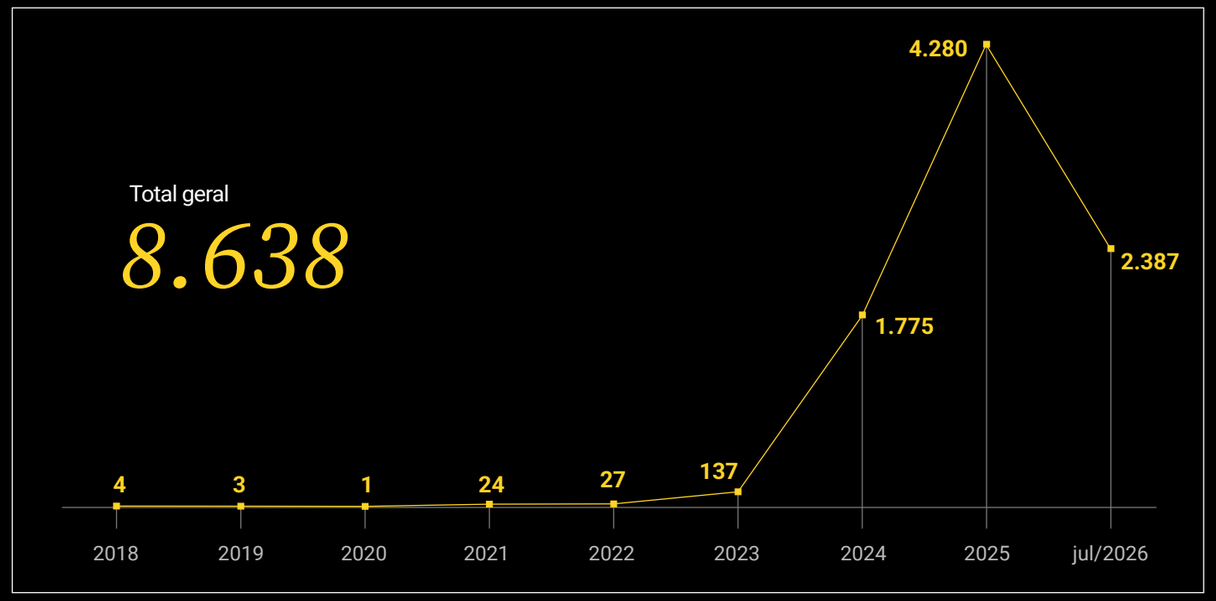
PRIMEIRA PLATAFORMA [DETALHES DISPONÍVEIS ÀS AUTORIDADES — A–13]

A primeira plataforma é voltada ao upload, organização e compartilhamento de fotos e vídeos adultos. O conteúdo pode ser exibido pública ou privadamente, e os criadores frequentemente utilizam links compartilháveis para distribuir álbuns diretamente ao público, sem que os visitantes precisem criar contas. Diferentemente de plataformas como Pornhub e XVideos, que monetizam por meio de anúncios e assinaturas premium, esta plataforma se concentra em permitir o compartilhamento direto. O site aceita uploads anônimos, sem verificação robusta de idade ou consentimento formal. Além do serviço principal de hospedagem, a plataforma promove serviços complementares voltados ao público adulto, como um *chatbot* de conteúdo sexual com função de roleplay (ourdream.ai) e uma plataforma de transmissões ao vivo.

Desde 2018, o canal de denúncias da SaferNet Brasil recebeu 8,6 mil registros associados a potencial conteúdo de exploração e abuso sexual infantojuvenil nesta plataforma — volume que evidencia a persistência e a escala do problema. Entre as denúncias, há relatos recorrentes de conteúdos íntimos de adultos publicados sem consentimento que, mesmo após denunciados, não foram removidos — o que aponta para falhas sistemáticas nos mecanismos de moderação da plataforma. A gravidade do cenário é reforçada por um dado mais amplo: a plataforma figura entre os 20 domínios mais denunciados ao canal da SaferNet Brasil desde 2007, posição que reflete não apenas a frequência das violações, mas a dificuldade estrutural de responsabilizar a empresa por conteúdos ilegais hospedados em seu ambiente.

FIGURA 51

Denúncias sobre a primeira plataforma para a SaferNet Brasil de 2018 (ano da primeira denúncia) a jul/2026.



ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

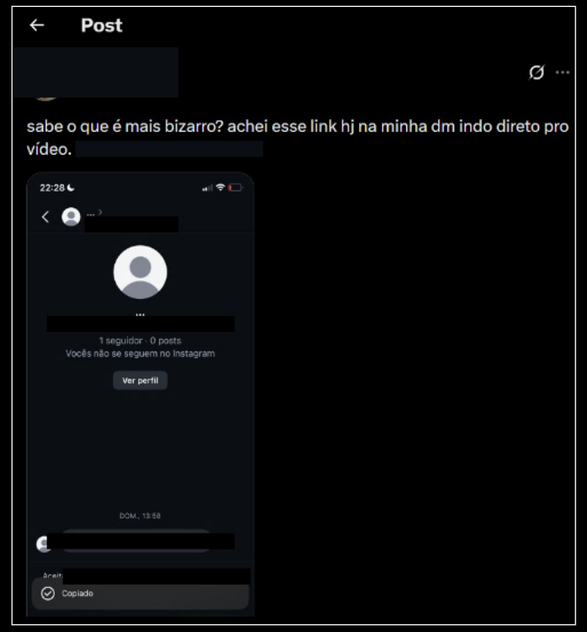
SOBRE ESTA PESQUISA

Desde meados de 2023, a primeira plataforma vem sendo citada em denúncias públicas como vetor de disseminação de conteúdo ilegal em redes sociais. *[detalhes disponíveis às autoridades — A-09]*[40] Em novembro de 2025, um caso envolveu uma usuária brasileira que reportou tentativa de extorsão sexual após *deepfake*s com sua imagem serem publicados na plataforma. A usuária afirmou ter recebido ameaças de chantagem mesmo após realizar a denúncia.[41]

Na avaliação da SaferNet Brasil, a popularidade da plataforma decorre do crescimento da pornografia amadora e da produção descentralizada de conteúdo, impulsionada por smartphones e redes sociais. Plataformas com cadastro facilitado e pouca verificação atraem usuários em busca de conteúdo fora dos grandes sites comerciais. A ausência de verificação efetiva de idade e de consentimento, no entanto, eleva significativamente os riscos, facilitando a circulação de material ilegal — desde imagens íntimas não autorizadas até material ou conteúdo de violência sexual contra crianças e adolescentes. A título de exemplo, mesmo após a implementação do ECA Digital em março de 2026, a plataforma não exige verificação de idade além da autoidentificação.

FIGURA 52

Publicação no X (antigo Twitter) sobre caso de sextorsão com *deepfakes* sexuais criadas sem consentimento (novembro de 2025).



Devido ao alto volume de URLs denunciadas desde 2018, a SaferNet Brasil selecionou aleatoriamente 112 denúncias para análise qualitativa. Cada publicação foi examinada individualmente e classificada conforme a natureza do conteúdo identificado. Por tratar-se de seleção aleatória, a amostra abrange links registrados entre 2018 e 2025.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

FIGURA 53

Distribuição da análise da SaferNet Brasil sobre os 112 links de denúncias.

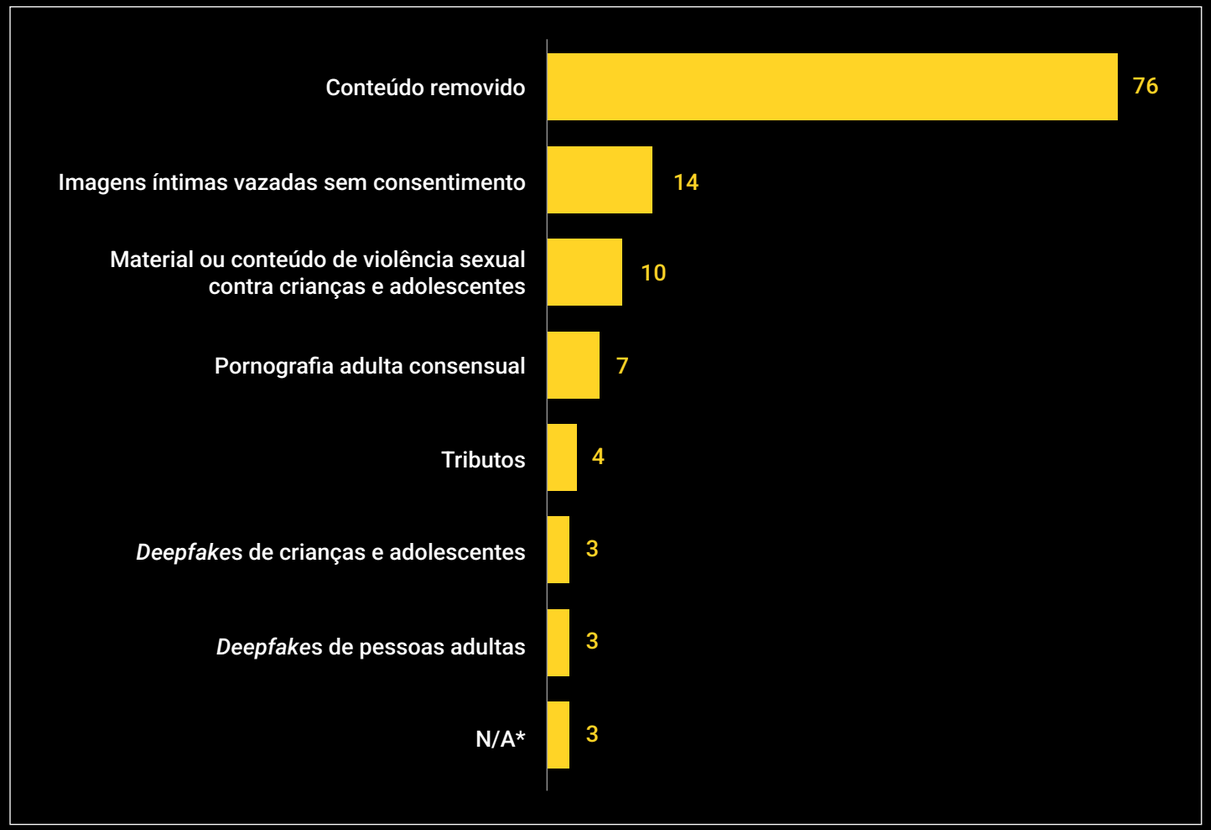
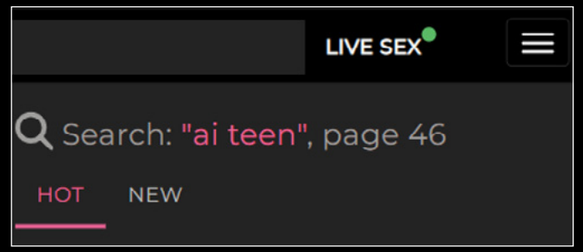


FIGURA 54

Exemplos de resultados de pesquisa.



ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

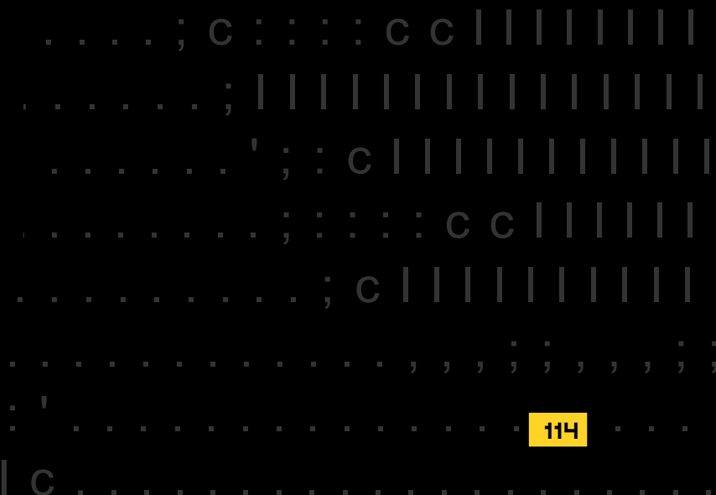
SOBRE ESTA PESQUISA

Das 112 publicações analisadas, 76 (67,9%) já haviam sido removidas — a plataforma permite remoção por direitos autorais a pedido da vítima. Material ou conteúdo de violência sexual contra crianças e adolescentes foi identificado em 10 URLs (27,8% dos disponíveis); imagens íntimas publicadas sem consentimento, em 14 (38,9%); *deepfake*s sexuais envolvendo crianças ou adolescentes, em 3 (8,3%); *deepfake*s de mulheres, em 4 (11,1%); e conteúdos classificados como "tributo" — prática na qual um indivíduo se masturba diante da imagem de uma mulher, com ou sem consentimento —, em 4 (11,1%). Ao todo, 6 URLs (16,7%) continham algum tipo de *deepfake* sexual. Três links foram classificados como N/A por não se enquadrarem nas categorias analisadas.

A sofisticação técnica dos conteúdos varia consideravelmente. Parte consiste em *deepfake*s rudimentares; outros, porém, são inteiramente gerados por IA a partir de múltiplas imagens reais da vítima, frequentemente extraídas de selfies ou fotos casuais publicadas em redes sociais. Entre os conteúdos denunciados, a SaferNet Brasil documentou *deepfake*s sexuais de crianças e adolescentes criados a partir de fotografias do Instagram. Em um dos exemplos, imagens de uma jovem foram extraídas de seu perfil e processadas por um aplicativo de nudificação — ferramenta especificamente desenvolvida para criar imagens sexuais não consensuais de indivíduos reais, incluindo crianças e adolescentes. *[detalhes disponíveis às autoridades — A-10]* [42]

Em buscas realizadas no mecanismo interno da plataforma, a SaferNet Brasil identificou mais de 5 mil publicações com descrições como "*deepfake*s", "adolescente de IA" e "AI teen" nos títulos. Consultas específicas por "AI teen" retornaram volume semelhante, com vídeos nessa categoria alcançando mais de 80 mil visualizações. Além disso, foram localizados conteúdos envolvendo uma dezena de *deepfake*s sexuais de autoridades brasileiras femininas.

Esse conjunto de evidências indica não apenas alta disponibilidade, mas também significativa circulação e consumo desse tipo de conteúdo. A padronização dos termos utilizados sugere a consolidação de categorias próprias dentro da plataforma, facilitando a busca e organização do material. Ao mesmo tempo, o volume de visualizações aponta para uma demanda consistente, enquanto a presença de figuras públicas entre as vítimas revela a ampliação do fenômeno para além de alvos anônimos, atingindo também mulheres em posições de visibilidade e poder.



ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

4.3.2. ESTUDO DE CASO

SEGUNDA PLATAFORMA [DETALHES DISPONÍVEIS ÀS AUTORIDADES — A–14]

A segunda plataforma analisada se descreve como "uma plataforma líder para descobrir conteúdo exclusivo de seus modelos favoritos do OnlyFans, influenciadores do Instagram e criadores de conteúdo adulto." Em julho de 2026, a empresa responsável foi multada por um órgão regulador estrangeiro por não ter adotado verificação de idade que impedisse o acesso de crianças a conteúdo pornográfico.[43]

A plataforma se organiza sob uma lógica de rede social baseada na segmentação de perfis e na oferta de um *feed* de conteúdos recomendados, cuja barreira de acesso se limita a um sistema de autoidentificação de idade — sem autenticação ou login, desde que não haja interação direta com as publicações.

Ao todo, a análise da SaferNet Brasil sobre a segunda plataforma identificou 263 *deepfake*s sexuais de mulheres jovens (faixa de 18 a 20 anos) e materiais artificiais de exploração sexual de adolescentes, envolvendo 22 vítimas, todas mulheres com milhões de seguidores nas redes sociais, cuja exposição pública desde a adolescência resultou na ampla disponibilidade de imagens suas ainda com menos de 18 anos.

Na maioria dos casos (234), os *deepfake*s utilizavam fotografias das vítimas quando adolescentes, independentemente de terem sido publicados antes ou depois de atingirem a maioridade, o que também configura crime pela lei brasileira. Em 29 casos, as vítimas ainda são adolescentes. Mais especificamente, observou-se predominância de materiais retratando vítimas com fotografias suas entre 15 e 17 anos.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

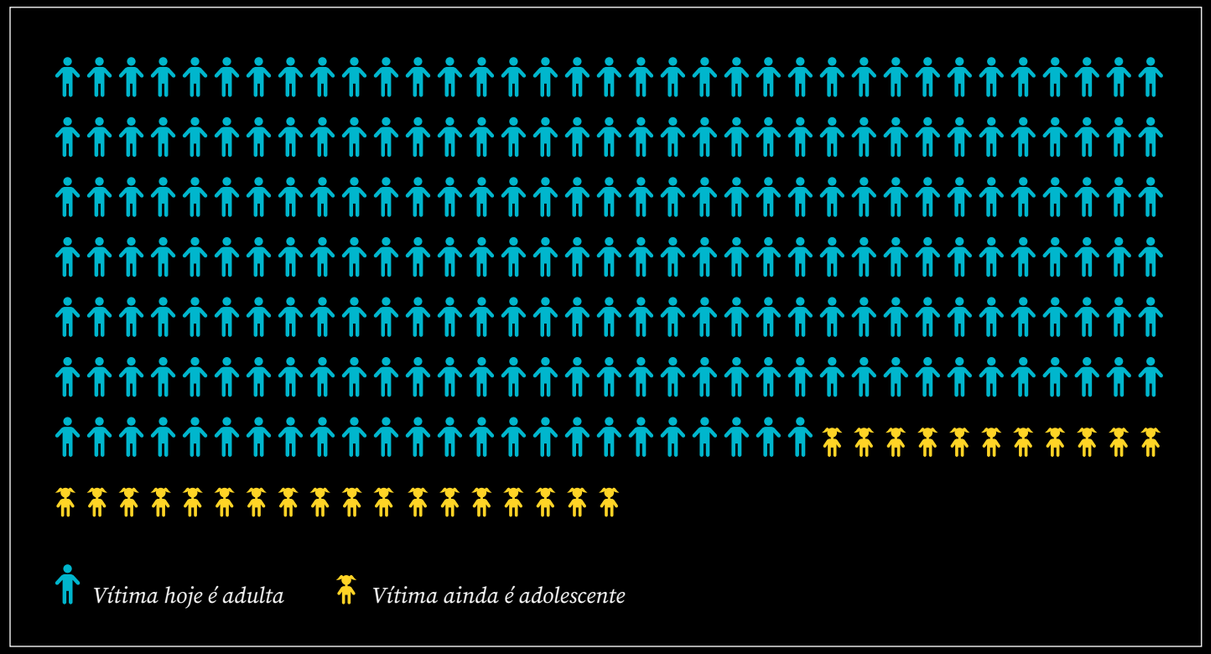
6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

FIGURA 55

Distribuição das idades atuais das vítimas de *deepfakes* sexuais na segunda plataforma.



A análise identificou que os conteúdos utilizados como base para a criação dos *deepfakes* sexuais originam-se, frequentemente, de fragmentos de vídeos curtos extraídos dos perfis oficiais das vítimas. A seleção dos materiais-fonte (mídias originais que foram adulteradas com nudificação) concentra-se em registros do cotidiano — rotinas de atividades físicas, viagens ou exibição de bens de consumo —, o que indica que conteúdos de exposição rotineira são frequentemente extraídos para a produção de materiais sintéticos.

Em vítimas que ganharam visibilidade ainda quando crianças ou adolescentes, como atrizes de televisão infantil ou cantoras mirins, observou-se um volume significativamente maior de interação por parte dos usuários, com dezenas de comentários sexualmente explícitos e agressivos, incluindo ameaças de violência física e de invasão de domicílio dirigidas a uma adolescente. *[detalhes disponíveis às autoridades — A-11]*[44]

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

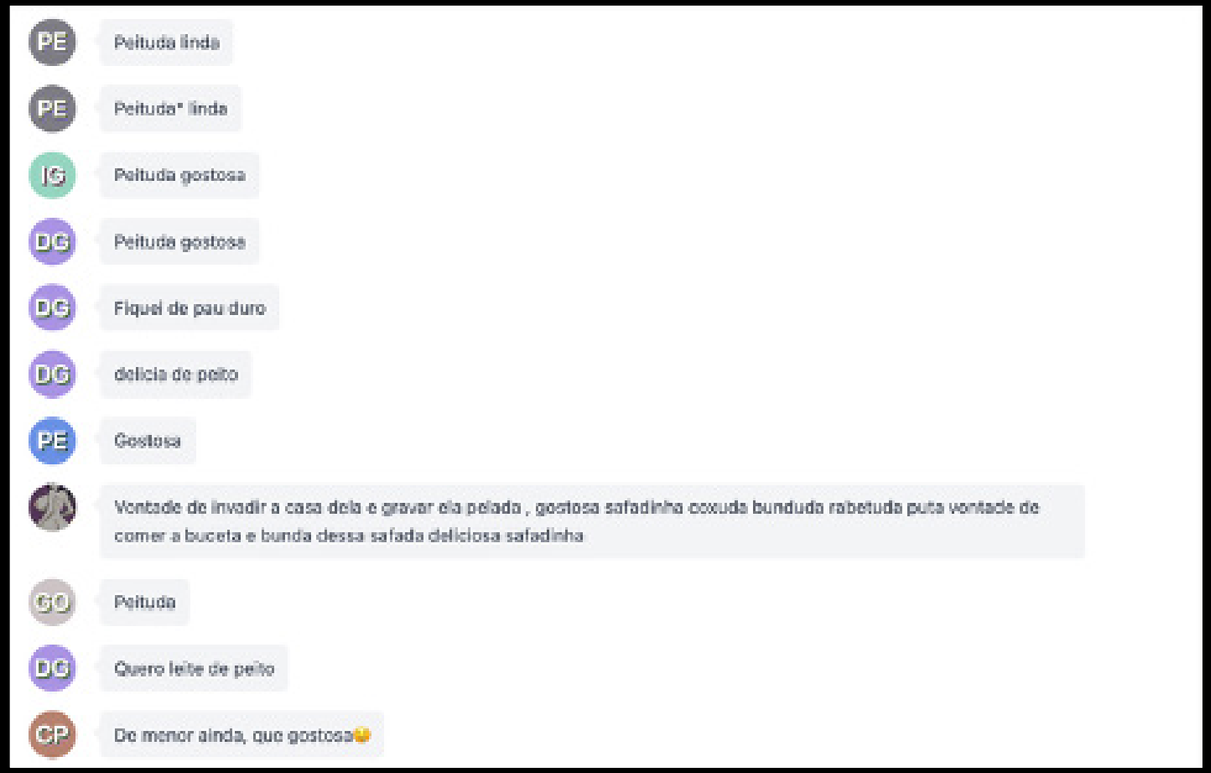
6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

FIGURA 56

Conversa entre usuários em publicação com *deepfake* sexual não consentido.



Esse fenômeno se insere em uma prática histórica e documentada de sexualização e exploração sexual de meninas e jovens mulheres com vidas públicas. Figuras como Brooke Shields, sexualizada em campanhas publicitárias e produções cinematográficas ainda na infância, e Natalie Portman, que aos 12 anos — após estrear em *O Profissional* (1994) — recebeu a primeira de diversas cartas descrevendo fantasias de estupro, ilustram como a exposição midiática precoce historicamente converteu meninas em alvos de objetificação sexual.

A pesquisa também identificou que, sobretudo nos casos envolvendo perfis de adolescentes, a plataforma e seus usuários empregam táticas de ofuscação para dificultar a localização dos conteúdos — como uso de números ao invés de letras (4 para A, 1 para I, 0 para O, assim por diante), espaçamento excessivo (mais de vinte toques) ou a exclusão das vogais dos nomes das vítimas. Por isso, a SaferNet Brasil conclui que a prática cumpre uma dupla função: burlar mecanismos de moderação automatizada e algoritmos de detecção de palavras-chave e criar

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

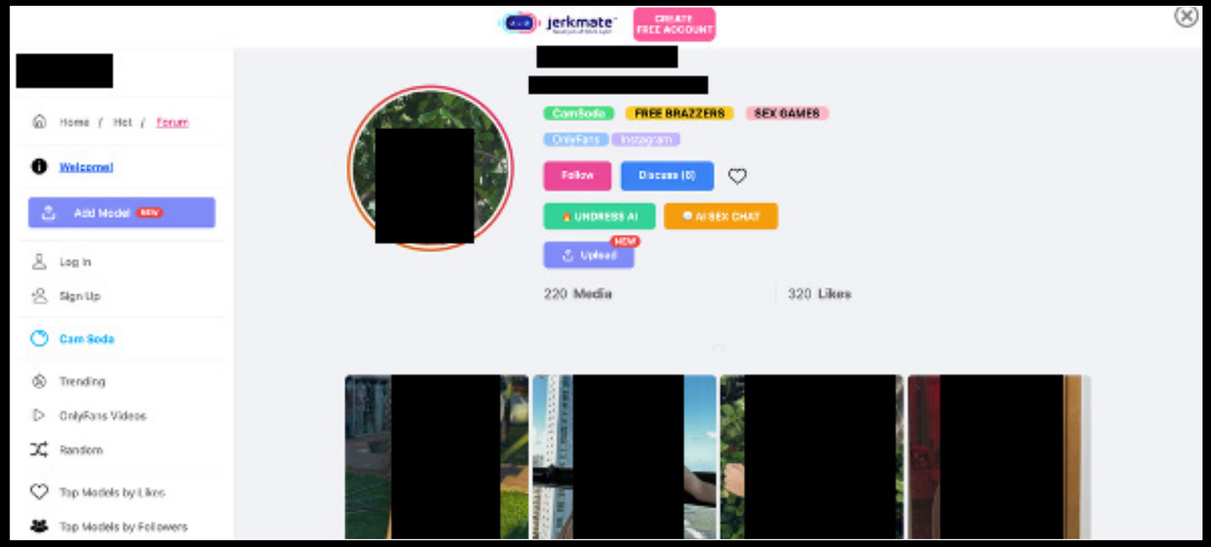
SOBRE ESTA PESQUISA

uma barreira de busca orgânica que dificulta às próprias vítimas localizarem os materiais, retardando a identificação do conteúdo ilícito e, consequentemente, prolongando o tempo de exposição ao impedir denúncias imediatas.

A segunda plataforma também possui uma seção de fórum de discussão interno — aberta para leitura também para usuários não registrados —, onde usuários podem pedir melhorias na plataforma ou conversar entre si. Muitos usam o espaço para compartilhar imagens íntimas sem o consentimento das vítimas, muitas vezes colocando elementos de identificação das mesmas nas postagens, como nome completo ou perfil de rede social. Há também discussões sobre uso de IA para criação de *deepfake*s não consensuais. Até março de 2026, cerca de 86 publicações no fórum mencionaram IA nos títulos.

FIGURA 57

Exemplo de página com *deepfakes* sexuais de influenciadora brasileira na segunda plataforma.



A arquitetura da plataforma apresenta desafios significativos. A estrutura de publicação omite metadados elementares, como registro de data e hora das postagens, o que inviabiliza o estabelecimento de uma cronologia da disseminação de materiais ilícitos. O sistema de busca interno é igualmente limitado e omite parte dos conteúdos hospedados. Durante a análise, a SaferNet Brasil observou que muitas publicações contendo *deepfake*s sexuais e materiais de exploração de adolescentes não aparecem nos resultados de busca direta da plataforma, sendo necessário rastrear links indexados externamente que levam a páginas ocultas dentro do próprio site.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

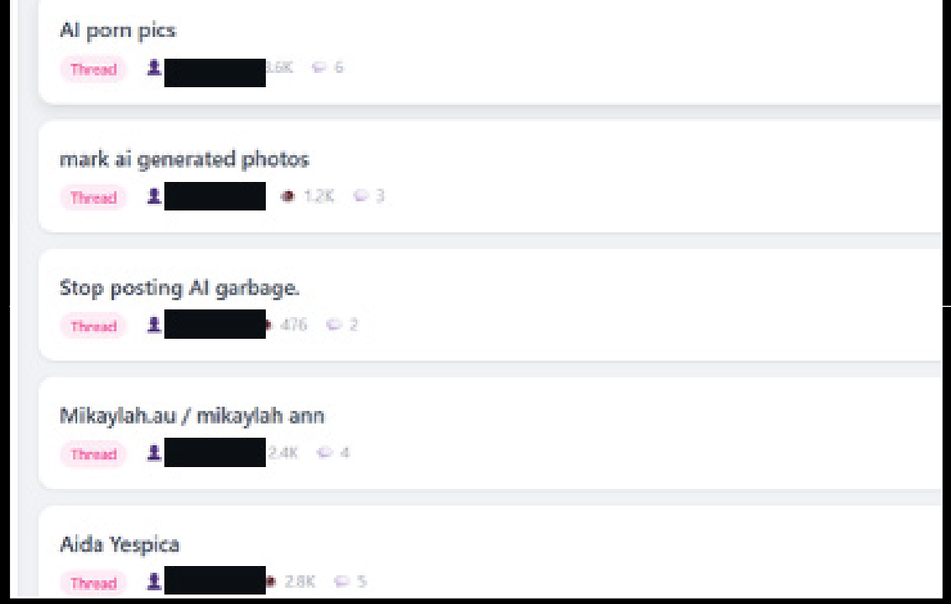
4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

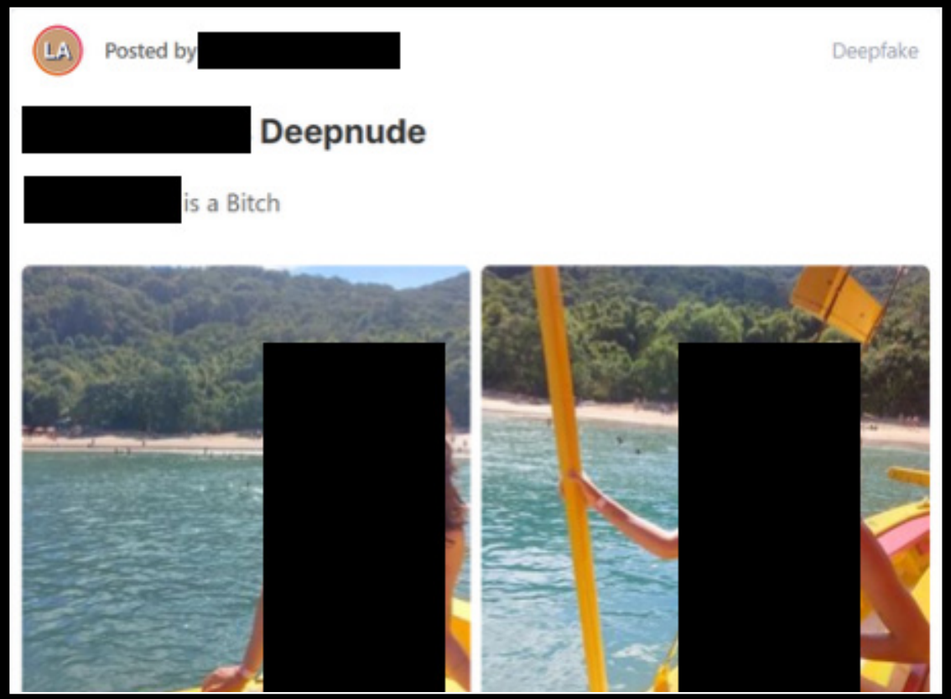
5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA





FIGURAS 58 E 59
Discussões no fórum de usuários da plataforma sobre deepfakes sexuais não consentidos e imagens nudificadas. Exemplo de imagem nudificada sem o consentimento da vítima.

Quanto à identificação dos usuários, a análise revelou uma arquitetura deliberadamente voltada ao anonimato: os perfis não exibem dados biográficos, nomes personalizados ou fotos, sendo identificados apenas por nomes de usuário. A ausência de links ou históricos de navegação impossibilita qualquer forma de rastreabilidade das atividades realizadas na plataforma — o que constitui uma barreira adicional à identificação de agentes e à análise de padrões comportamentais.

Assim como a primeira plataforma, esta segunda também promove em seu site serviços de transmissão ao vivo de conteúdo sexual (livecams). Nesta, o serviço vinculado é o Stripchat, que exibe categorias como “Adolescentes” e “Adolescentes +18 se masturbando”, com perfis que informam nacionalidade, traços físicos, atividades realizadas em transmissão, sexualidade e idade das produtoras de conteúdo. Observou-se ainda que a busca pelo termo “*child*” na plataforma não retorna resultados, enquanto a busca por “*teen*” (adolescente) localiza dezenas de conteúdos e perfis — em alguns dos quais as envolvidas utilizam máscaras ou filmam abaixo do rosto para não serem reconhecidas. A plataforma também disponibiliza uma funcionalidade de geração de imagens artificiais personalizáveis, na qual o usuário pode selecionar características para criar uma personagem que simule maior proximidade com a realidade, incluindo perfis que aparentam ter acabado de atingir a maioridade. Para essa funcionalidade específica, é exigida a criação de cadastro.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

CAPÍTULO CINCO

Respostas de diferentes atores ao problema da violência sexual mediada por IA

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

A emergência de *deepfake*s sexuais e de materiais artificiais de abuso sexual infantil gerados por IA constitui uma crise global de proteção à infância que exige respostas coordenadas, multissetoriais e juridicamente robustas. Este capítulo examina como diferentes atores — empresas de IA generativa, legisladores, autoridades públicas, organismos internacionais e a sociedade civil brasileira — estão respondendo ao problema, identificando tanto avanços significativos quanto lacunas persistentes.

As respostas institucionais até aqui revelam tensões fundamentais entre inovação tecnológica e proteção de direitos, entre autorregulação corporativa e *enforcement* estatal, entre tipificação penal e capacidade investigativa real. Enquanto algumas empresas implementam salvaguardas técnicas sofisticadas — como filtragem de *prompts*, curadoria de dados de treinamento e moderação específica para conteúdo sintético—, outras liberam modelos de código aberto sem controles posteriores, criando vetores permanentes de risco. Enquanto outros países avançam em legislações específicas que criminalizam não apenas a produção de conteúdo sintético ilícito, mas também as ferramentas e guias que ensinam sua criação, o Brasil ainda depende de tipificações do ECA — aprovado originalmente em 1990 e atualizado para o contexto digital apenas em 2008 — e assiste a uma disputa fragmentada no Congresso Nacional, com dezenas de projetos de lei sobre *deepfake*s apresentados por parlamentares de diferentes partidos e orientações, sem que se tenha alcançado, até o momento, uma resposta legislativa unificada à altura da urgência do problema.

A premissa central que orienta esta análise é que respostas isoladas — sejam elas exclusivamente tecnológicas, jurídicas, legislativas ou educacionais — são estruturalmente insuficientes para enfrentar um problema de natureza sistêmica. O UNICEF alerta que o cenário permanece desigual e que modelos de IA em número excessivo estão sendo desenvolvidos sem salvaguardas adequadas, o que representa um risco agravado quando ferramentas generativas são integradas diretamente a plataformas, redes sociais e outros produtos onde imagens manipuladas se disseminam com velocidade incompatível com qualquer resposta de moderação.[45]

A proteção efetiva de crianças e adolescentes nesse contexto exige um ecossistema integrado de responsabilidades compartilhadas, no qual empresas implementam salvaguardas robustas desde o design do produto, legisladores criam tipificações claras e instrumentos processuais ágeis, autoridades desenvolvem capacidade investigativa especializada, educadores formam cidadãos digitalmente conscientes e a sociedade civil atua como fiscalizadora crítica e provedora de suporte às vítimas. Essa articulação não é aspiracional — é a condição mínima para romper a lógica de impunidade que sustenta a proliferação de conteúdos criminosos documentada nos capítulos anteriores deste relatório.

PARA COMPREENDER O ALCANCE DO PROBLEMA, É NECESSÁRIO EXAMINAR CADA CATEGORIA DE ATOR DO ECOSSISTEMA DIGITAL SEPARADAMENTE: AÇÕES INDIVIDUAIS, PRESSÃO REGULATÓRIA E SALVAGUARDAS TÉCNICAS.

Em abril de 2024, Amazon, Anthropic, Civitai, Google, Meta, Metaphysic, Microsoft, Mistral AI, OpenAI e Stability AI aderiram aos princípios voluntários de *Safety by Design* (seguro por design) propostos pela Thorn e a All Tech Is Human, comprometendo-se a implementar nos seus produtos medidas para prevenir a geração e disseminação de material ou conteúdo de violência sexual contra crianças e adolescentes por IA ao longo de todo o ciclo de vida dos modelos, do desenvolvimento à manutenção posterior ao lançamento.[46] Essas medidas incluem, entre outras, a remoção de material ou conteúdo de violência sexual contra crianças e adolescentes de bases de treinamento, mecanismos de reporte e filtragem e a hospedagem responsável de modelos para minimizar abusos conhecidos.

Os princípios abrangem desde a curadoria responsável de bases de treinamento e a remoção de material abusivo dos conjuntos de dados até a realização de *red-teaming* contínuo, a moderação de conteúdo gerado e a hospedagem responsável de modelos. Em abril de 2025, a Thorn publicou relatórios de progresso das empresas signatárias, nos quais algumas reportaram avanços como auditorias completas de bases de dados de treinamento, adoção do padrão C2PA[47] para rotulagem de conteúdo gerado por IA e implementação de monitoramento de *prompt*s por meio de ferramentas de terceiros.[48]

Do ponto de vista técnico, as estratégias de mitigação implementadas pelas empresas operam em diferentes camadas da cadeia de geração. No nível de *input* — isto é, antes que o comando do usuário chegue ao modelo generativo —, diversas soluções atuam como *firewalls* para sistemas de IA. O *PromptGuard*, framework publicado em janeiro de 2025, introduziu o conceito de *soft prompt*s de segurança para modelos texto-para-imagem, inspirando-se no mecanismo de *system prompt*s dos grandes modelos de linguagem: em vez de depender de modelos auxiliares para verificar ou desintoxicar conteúdo, o *PromptGuard* neutraliza diretamente elementos maliciosos no *prompt* de entrada sem afetar a qualidade da geração de imagens legítimas. Nos testes, a técnica superou oito defesas anteriores, reduzindo a taxa de conteúdo inseguro gerado para 5,84% e operando 3,8 vezes mais rápido que métodos de moderação precedentes.[49]

No plano comercial, as grandes provedoras de infraestrutura também oferecem soluções de filtragem integradas a seus serviços.

- A Amazon disponibiliza o *Bedrock Guardrails*, que filtra conteúdo nocivo, bloqueia tópicos proibidos e oculta informações sensíveis tanto na entrada quanto na saída dos modelos.
- O Google desenvolveu classificadores de conteúdo dentro de seu *Responsible GenAI Toolkit* (Kit de ferramentas para IA generativa responsável), projetados para verificar a conformidade de entradas e saídas com políticas de uso seguro, e detalhou publicamente como seus sistemas de IA responsável atuam na detecção e no bloqueio de material ou conteúdo de violência sexual contra crianças e adolescentes.
- A OpenAI mantém uma *Moderation API* que avalia em tempo real a probabilidade de que o conteúdo gerado contenha violência, discurso de ódio ou material adulto, atribuindo pontuações de risco que permitem o bloqueio automático de respostas; em abril de 2026, a empresa publicou um *Child Sa-*

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

fety Blueprint elaborado em conjunto com o NCMEC, a Attorney General Alliance e a Thorn, abordando a modernização legislativa, o aprimoramento de relatórios de denúncia e a incorporação de princípios de segurança desde o design dos sistemas.

- A Microsoft, por sua vez, detalhou compromissos que incluem a revisão rigorosa de dados de treinamento, a aplicação de camadas de proteção durante o *deployment* e o desenvolvimento de tecnologia para detectar uso adversarial de seus sistemas com foco em segurança infantil. Essas camadas de filtragem são complementadas por mecanismos de monitoramento de saída, que analisam o conteúdo gerado antes de entregá-lo ao usuário — um processo que, nas implementações mais avançadas, inclui análise semântica, detecção de dados sensíveis e verificação de conformidade regulatória.

Não obstante esses compromissos, há também evidências consistentes de que políticas restritivas não necessariamente se traduzem em resultados práticos robustos quando os modelos são disponibilizados como de código aberto. O Wilson Center observa que modelos de código aberto apresentam risco mais elevado de abuso porque as restrições sobre o uso "são difíceis de aplicar e fáceis de ignorar por agentes maliciosos", ao contrário de modelos fechados, cujos desenvolvedores podem, em tese, restringir ou bloquear o acesso.[50]

O problema não reside em um único modelo, mas em um ecossistema em expansão. Dito isso, é importante registrar que parte das empresas e plataformas do ecossistema de código aberto tem adotado medidas de mitigação, algumas delas com resultados mensuráveis. A Black Forest Labs, desenvolvedora do Flux, implementou um protocolo de segurança em múltiplas camadas antes de cada lançamento: filtragem de material ou conteúdo de violência sexual contra crianças e adolescentes dos dados de pré-treinamento em parceria com a Internet Watch Foundation, múltiplas rodadas de *fine-tuning* de segurança pós-treinamento para inibir a geração de material abusivo, e avaliações externas por terceiros independentes — incluindo testes adversariais com mais de 2.800 *prompt*s focados na geração artificial de material ou conteúdo de violência sexual contra crianças e adolescentes e de imagens íntimas não consensuais. No caso do modelo Flux.1 Kontext, foram avaliados 21 *checkpoints* distintos antes da liberação. A empresa também incorporou marcas d'água digitais e implementação do padrão C2PA para rastreabilidade de conteúdo gerado.[51]

O CivitAI, principal plataforma de hospedagem de modelos de terceiros, reportou ter detectado mais de 252 mil *prompt*s violadores em seus filtros de entrada, removido retroativamente 183 modelos otimizados para geração de *deepfake*s sexuais ou materiais artificiais de abuso sexual infantojuvenil, banido mais de 17 mil contas de usuários e encaminhado 178 relatórios ao NCMEC referentes a casos confirmados de material ou conteúdo de violência sexual contra crianças e adolescentes gerado por IA.[52] A plataforma também desenvolveu filtros conceituais chamados *Semi-Permeable Membranes* (SPMs), que operam no nível semântico — bloqueando conceitos e não apenas termos — e removeu proativamente mais de 5.400 modelos que violavam suas políticas de conteúdo. A Stability AI, signatária dos princípios de *Safety by Design*, também implementou filtros em versões mais recentes do Stable Diffusion e conduziu duas rodadas de *red-teaming* em preparação para implementações em 2025.[53]

No entanto, pesquisas recentes indicam que mesmo as melhores práticas de mitigação ainda enfrentam limites estruturais. Pesquisadores da Universidade de Georgetown demonstraram, em resultados preliminares, que mesmo após a remoção automática de cerca de 94% dos

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

dados de treinamento referentes a um conceito específico — como imagens de crianças —, o modelo ainda era capaz de produzir conteúdo abusivo por meio da combinação de conceitos aprendidos separadamente. Esse fenômeno, conhecido como generalização composicional, ocorre porque os modelos de IA não armazenam imagens como unidades indivisíveis: eles aprendem atributos visuais de forma modular — rostos, corpos, cenários, idades, contextos — e podem recombiná-los de maneiras inéditas. Na prática, isso significa que um modelo pode nunca ter sido exposto a uma imagem de abuso sexual infantil e, ainda assim, ser capaz de gerar uma, combinando o conceito de "criança" com o conceito de "nudez" ou "ato sexual", cada um aprendido em contextos distintos e legítimos. A filtragem de dados de treinamento, embora necessária, não elimina essa capacidade emergente. Em artigo do *Tech Policy Press*, a pesquisadora Elissa Redmiles observou que "as pessoas encarregadas de testar esses modelos trabalham com uma mão amarrada nas costas, sabendo que agentes maliciosos não se impõem a mesma limitação" — referindo-se ao fato de que empresas que conduzem *red-teaming* para conteúdo artificial de abuso e exploração sexual infantojuvenil enfrentam riscos legais, enquanto ofensores operam sem restrições.[54]

Dessa forma, a disponibilização dos pesos dos modelos torna, em última instância, todas essas proteções estruturalmente reversíveis: uma vez publicados, filtros e *fine-tuning* de segurança podem ser removidos por qualquer usuário com conhecimento técnico básico. O padrão identificado pela literatura é recorrente — um pequeno número de modelos abertos, treinados em dados da web com conteúdo pornográfico e liberados sem salvaguardas suficientes, é responsável pela maioria esmagadora do conteúdo abusivo gerado por IA. O resultado é um ecossistema no qual a inovação em capacidade generativa avança em velocidade incomparavelmente superior à inovação em salvaguardas — e no qual cada novo modelo lançado representa, simultaneamente, um avanço técnico e um novo vetor de risco.

Crimes desta natureza costumam ser praticados entre diferentes plataformas, portanto é essencial haver um compartilhamento de informações e cooperação entre diferentes empresas do setor tecnológico. Um exemplo dessa estratégia é o **Lantern**, uma infraestrutura de inteligência cibernética lançada em 2023 para unificar o combate à exploração sexual infantil on-line. A iniciativa funciona por meio do compartilhamento seguro e em tempo real de 'sinais' — que abrangem desde dados técnicos, como *hashes* de conteúdos e URLs nocivas, até padrões comportamentais de aliciamento e extorsão financeira. Contando também com a participação consultiva do setor financeiro para rastrear fluxos de pagamento ilícitos, o Lantern viabilizou, entre 2023 e 2025, o compartilhamento de mais de 2 milhões de sinais, resultando diretamente em 353 mil ações de moderação e derrubada de contas. Essa articulação demonstra que, diante de ameaças complexas potencializadas por IA, a eficácia das salvaguardas técnicas depende também da quebra de sigilos corporativos.

A crescente pressão regulatória global também tem forçado respostas mais concretas das empresas. Em janeiro de 2026, o Ofcom, regulador de comunicações do Reino Unido, abriu investigação formal contra o X (antigo Twitter) sob o *Online Safety Act*, após o caso Grok. O governo britânico declarou que, caso o X não adotasse medidas adequadas, o Ofcom teria respaldo para aplicar multas de até 10% da receita global da empresa ou, nos casos mais graves, solicitar ordem judicial para bloquear o acesso de usuários britânicos à plataforma. Na mesma esteira, o *Information Commissioner's Office* (ICO) abriu investigação sobre o proces-

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

samento de dados pessoais pelo Grok e sua capacidade de produzir imagens sexualizadas de indivíduos, incluindo crianças.

Apesar desses avanços técnicos e colaborativos descritos, as salvaguardas enfrentam limitações importantes de evasão e eficácia que são, em parte, inerentes à própria arquitetura de certos sistemas de IA generativa. Após o lançamento de um modelo de IA, as estratégias de mitigação geralmente concentram-se na filtragem de *prompts* ou na avaliação em tempo real das saídas — ou seja, as imagens geradas —, mecanismos que, embora necessários, operam em uma desvantagem estrutural diante de redes e indivíduos criminosos de alto risco, com poder financeiro o suficiente para contornar essas barreiras.[55]

Um desafio adicional reside nos chamados *prompts* composicionais — comandos que, individualmente, não violam nenhum filtro de segurança, mas que, ao serem combinados, produzem conteúdo ilícito. Um *prompt* que solicite a criação de uma imagem de uma "menina de 12 anos" será, na maioria das plataformas e dos modelos de IA generativa, bloqueado se associado a termos sexuais explícitos; no entanto, usuários com intenção criminosa fragmentam a solicitação em elementos inofensivos — como descrições de idade, cenário, vestuário ou posição corporal. Esses elementos, em conjunto, podem resultar em material artificial de exploração sexual infantil sem que nenhum termo isolado acione os filtros. Essa técnica explora a mesma propriedade de generalização composicional dos modelos descrita na seção anterior: assim como o modelo pode combinar conceitos aprendidos separadamente para gerar imagens inéditas, um usuário também pode compor comandos fragmentados para contornar salvaguardas baseadas em palavras-chave. Nesse sentido, a defesa contra essa estratégia exige sistemas de moderação capazes de interpretar o significado semântico do conjunto do *prompt* — e não apenas a presença ou ausência de termos proibidos —, o que permanece um problema em aberto na pesquisa de segurança de IA.

O cenário resultante é de uma assimetria fundamental. De um lado, empresas investem em compromissos voluntários, relatórios de progresso, ferramentas de filtragem e parcerias com organizações de proteção à infância — esforços reais, mas cuja implementação permanece aquém dos esforços do setor para eliminar o problema e desafia a capacidade dos países para regular e fiscalizar. De outro, ofensores operam sem restrições legais, éticas ou técnicas, explorando cada novo modelo lançado como um vetor adicional de risco.

5.2. RESPOSTAS DE AUTORIDADES PÚBLICAS BRASILEIRAS E DO CONGRESSO

Nos últimos quatro anos, dezenas de projetos de lei (na sigla PL) buscando regulamentar, impedir ou coibir o uso de *deepfake*s — não necessariamente apenas os de natureza sexual — e de materiais artificiais de abuso e exploração sexual infantil foram apresentados no Congresso Nacional.

Até maio de 2026, a SaferNet Brasil identificou 70 projetos de lei na Câmara dos Deputados e no Senado relacionados ao uso de *deepfake*s e IA, apresentados por parlamentares de ao menos 15 partidos. O ritmo de apresentação vem se acelerando: dos 70 PLs mapeados, 20 foram apresentados somente entre fevereiro e maio de 2026. Os projetos se distribuem por ao menos doze eixos temáticos distintos, com sobreposições frequentes e baixa coordenação entre as propostas.

A análise temática dos PLs com base em suas ementas evidencia que a maior parte não aborda diretamente a exploração sexual: apenas 21,4% (15 de 70) tratam do tema. Metade dos projetos (35 de 70) se concentra em transparência e rotulagem de conteúdo sintético (18,6%), legislação eleitoral (14,3%), proteção de direito de imagem (11,4%) e tipificação de fraude (5,7%) — temas relevantes, mas que não enfrentam o núcleo do problema documentado neste relatório.

FIGURA 60

Distribuição temática dos projetos de lei identificados pela SaferNet Brasil.

Tema	PLs	% do total
Direito de imagem (inclusive pós-morte)	16	22,9%
Transparência e rotulagem	14	20,0%
Exploração sexual infantojuvenil	12	17,1%
Legislação eleitoral	11	15,7%
Violência de gênero / misoginia digital	10	14,3%
Fraude, estelionato ou violência psicológica	7	10,0%
Uso de dados pessoais	6	8,6%
Educação digital e literacia midiática	6	8,6%
Regulação geral de *deepfakes*	4	5,7%
Exploração sexual (todas as faixas etárias)	3	4,3%
Governança e regulação geral de IA	2	2,9%
Outros	1	1,4%
Total de menções	92	—

A tabela contabiliza todas as menções temáticas — um mesmo projeto pode ser classificado em até três categorias, somando 92 menções. Já os percentuais citados no texto consideram apenas o tema principal de cada projeto (total de 70)

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Em maio de 2026, o governo federal publicou um decreto presidencial que veda aos provedores de aplicação de internet a prática de nudificação.[56] Trata-se do Decreto n.º 12.976, que estabelece diretrizes para a proteção de mulheres na internet e para o enfrentamento da violência contra mulheres no ambiente digital. Conforme o art. 9º do decreto, fica vedada aos provedores de aplicações de internet a geração e a modificação de conteúdo íntimo de terceiro mediante uso de inteligência artificial ou de qualquer outro recurso tecnológico que altere imagem ou som da vítima.

Complementarmente, o art. 10 determina que os provedores baseados em funcionalidades de IA implementem salvaguardas técnicas e procedimentais para identificar e bloquear solicitações de geração desses conteúdos, de forma escalonada e proporcional ao volume de acessos e ao nível de risco da aplicação. O decreto também prevê a remoção célere de conteúdo íntimo divulgado sem autorização, incluindo aqueles gerados por IA. De acordo com o decreto, após notificação da vítima ou de seu representante, o provedor deve indisponibilizar o material no prazo de até duas horas.

Uma das respostas institucionais mais significativas até o momento foi a atuação coordenada da Autoridade Nacional de Proteção de Dados (ANPD), do Ministério Público Federal (MPF) e da Secretaria Nacional do Consumidor (Senacon) no caso Grok. Em janeiro de 2026, após denúncias de que o sistema de IA generativa integrado à plataforma X estava sendo utilizado para produzir *deepfake*s sexualizados de mulheres, crianças e adolescentes a partir de fotografias reais, as três instituições expediram recomendação conjunta exigindo que a empresa implementasse medidas imediatas para impedir a geração desses conteúdos em todas as versões do Grok.

A Nota Técnica n.º 1/2026 da ANPD consolidou um entendimento juridicamente relevante: conteúdos sintéticos gerados por IA que reproduzam ou simulem pessoas identificáveis constituem dados pessoais para fins da LGPD e, quando envolvem elementos biométricos, configuram dados pessoais sensíveis.[57] Na prática, esse enquadramento submete a geração de *deepfake*s ao regime da LGPD, exigindo base legal, finalidade compatível e deveres de segurança e governança — independentemente de legislação específica sobre IA.

Em fevereiro de 2026, após avaliar a resposta da plataforma X, as três instituições consideraram as providências informadas insuficientes: a empresa alegou ter removido milhares de publicações e suspendido centenas de contas, mas não apresentou evidências concretas, relatórios técnicos ou mecanismos de monitoramento que permitissem aferir a efetividade das medidas.[58] Testes realizados pelas equipes técnicas das instituições indicaram a persistência das falhas, com a continuidade da geração e circulação de conteúdos incompatíveis com as recomendações. A Senacon expediu medida cautelar administrativa determinando implementação imediata de soluções técnicas, com prazo de cinco dias úteis para comprovação; o MPF determinou a apresentação de relatórios mensais sobre remoções e bloqueios; e a ANPD sinalizou a possibilidade de multa diária e instauração de processo administrativo sancionador.

Embora o caso Grok represente o episódio de maior repercussão, a atuação do governo federal não se restringe a ele. A ANPD incluiu a IA e a proteção de crianças e adolescentes entre os temas prioritários de seu mapa de fiscalização para o biênio 2026–2027. O conjunto dessas ações indica que o aparato institucional brasi-

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

leiro dispõe de instrumentos legais e capacidade de articulação para responder ao problema — como demonstrado pela coordenação entre ANPD, MPF e Senacon no caso Grok. No entanto, a efetividade dessas respostas permanece condicionada à capacidade de fiscalização continuada, à cooperação das plataformas e à existência de tipificações penais específicas que acompanhem a evolução técnica das ferramentas de IA generativa.

No plano investigativo, tanto a Polícia Federal (PF) quanto as polícias civis estaduais vêm deflagrando operações relacionadas à produção e distribuição de material ou conteúdo de violência sexual contra crianças e adolescentes envolvendo IA generativa, embora a escala dessas ações permaneça limitada diante do volume documentado neste relatório. Em outubro de 2025, a PF prendeu, na Operação Miopia, um homem investigado por participar ativamente de fóruns da *dark web* que conectavam abusadores sexuais em várias partes do mundo, com apreensão de oito computadores, quatro celulares e dezenas de mídias de armazenamento.[59] Em abril de 2026, a polícia prendeu em flagrante um homem na região metropolitana de São Paulo suspeito de utilizar IA para produzir material ou conteúdo de violência sexual contra crianças e adolescentes a partir de fotos de crianças obtidas em redes sociais, com apreensão de 18 dispositivos eletrônicos contendo centenas de milhares de imagens. [60]

No plano legislativo estadual, Mato Grosso tornou-se, em março de 2026, o primeiro estado brasileiro a aprovar uma legislação específica sobre o tema: a Lei n.º 13.257/2026 proíbe o uso de IA para criar, distribuir ou divulgar *deepfake*s sexuais no estado, obrigando plataformas a remover conteúdos e a oferecer canais de denúncia.[61] A aprovação foi motivada diretamente pelos dados da SaferNet Brasil, que identificaram 30 vítimas de *deepfake*s em escolas mato-grossenses — o segundo maior número entre os estados analisados —, um impacto significativo desta pesquisa no mundo real.

A atuação dos estados, contudo, permanece heterogênea e predominantemente reativa. Não há, até o momento, um protocolo nacional padronizado de respostas aos incidentes envolvendo as escolas, com diretrizes do MEC, no âmbito do Escola que Protege, voltadas especificamente ao enfrentamento do fenômeno no ambiente escolar. A capacitação de delegacias especializadas em crimes cibernéticos para lidar com material gerado por IA — que exige competências técnicas distintas das investigações tradicionais de material ou conteúdo de violência sexual contra crianças e adolescentes — varia significativamente entre as unidades da federação. A lei de Mato Grosso, embora pioneira, evidencia justamente a lacuna federal: uma tentativa de legislar no âmbito estadual de forma fragmentada, com alcance territorial limitado e sem a necessária harmonização com a legislação nacional vigente.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

5.3. RESPOSTAS DE AUTORIDADES INTERNACIONAIS

O conjunto de respostas legislativas internacionais ao problema dos *deepfake*s sexuais e do material ou conteúdo de violência sexual contra crianças e adolescentes gerado por IA revela um campo normativo em rápida transformação — mas ainda marcado por profundas assimetrias. Enquanto alguns países avançam com marcos regulatórios abrangentes e penas específicas, outros recorrem à analogia penal ou permanecem sem qualquer tipificação. Compreender esse panorama é fundamental para situar o estágio atual da proteção jurídica disponível e identificar as lacunas que ainda precisam ser enfrentadas. Nas páginas a seguir, a SaferNet apresenta um mapeamento de legislações internacionais feito entre outubro de 2025 e abril de 2026.

Legenda do "status de legislação"

Tipificação específica	*País tem legislação que criminaliza expressamente material artificial de abuso sexual infantil e/ou deepfakes gerados por IA*
Aplicável por analogia / parcial	*Leis existentes podem ser aplicadas por analogia, mas sem tipificação autônoma para conteúdo gerado por IA*
Em tramitação / debate	*Projetos de lei em discussão ou legislação parcial aprovada*
Sem legislação / lacuna crítica	*Ausência de marcos normativos ou lacuna regulatória identificada*

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

5.3.1.
AMÉRICA LATINA (SEM O BRASIL)



ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA



A Colômbia avançou com a Lei 2502/2025, que introduz agravante específico para fraude de identidade por IA, incluindo *deepfakes*,[62] e tem em tramitação o Projeto de Lei 042/2025, voltado à proteção de crianças e adolescentes em ambientes digitais e à imposição de obrigações às plataformas. [63]

Chile e México encontram-se em estágios iniciais: o Chile está debatendo um projeto de lei para regulamentação da IA considerado por alguns como um dos mais avançados da região, com foco em direitos humanos e proteção de grupos vulneráveis;[64] o México discute projeto federal no Senado e, sem lei específica vigente, recorre à interpretação analógica da legislação penal existente.[65]

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

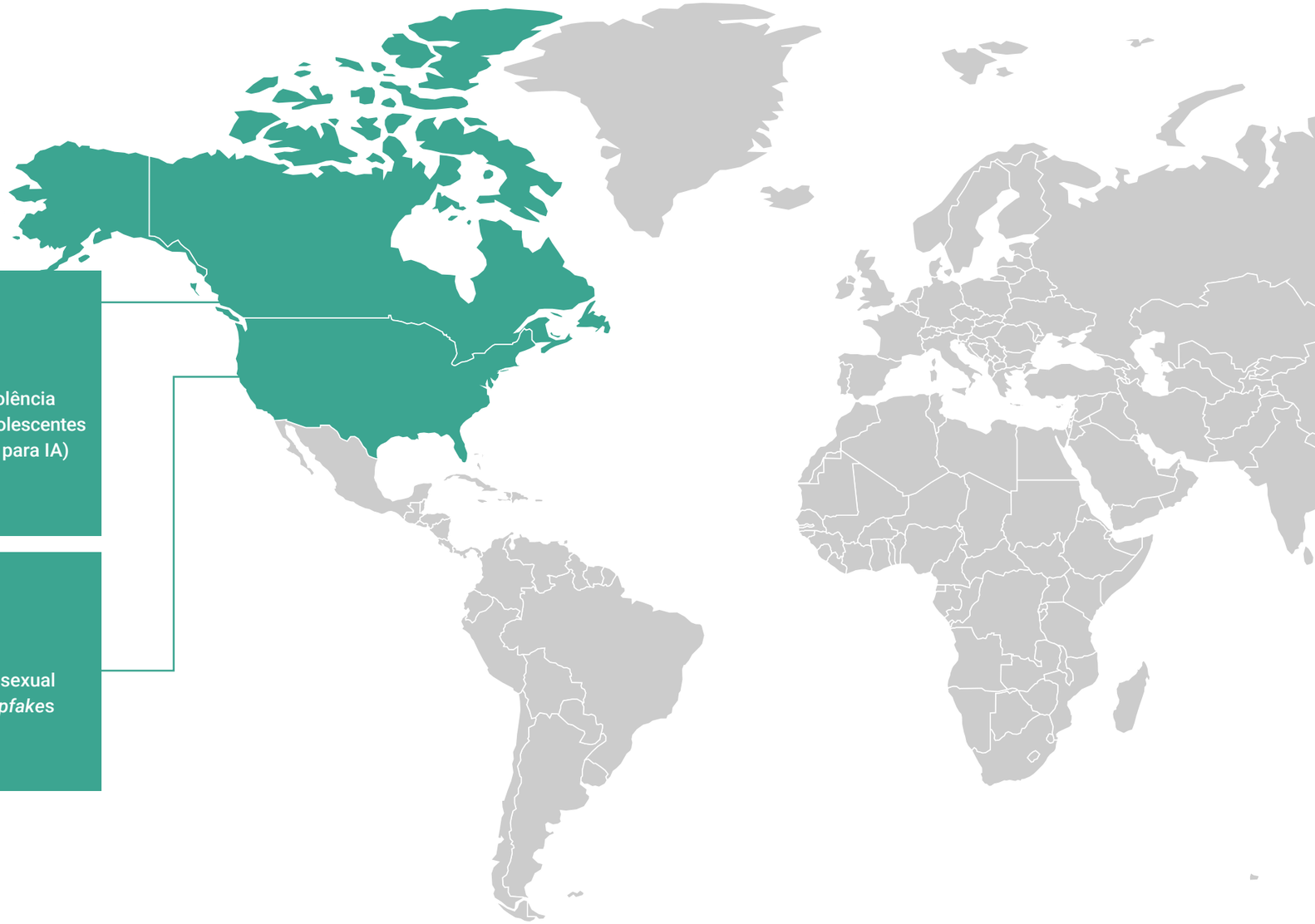
SOBRE ESTA PESQUISA

5.3.2. AMÉRICA DO NORTE

Status da legislação

Foco principal

Pena máxima



CANADÁ

Aplicável por analogia

Material ou conteúdo de violência sexual contra crianças e adolescentes (sem tipificação específica para IA)

Variável

ESTADOS UNIDOS

Tipificação específica

Material artificial de abuso sexual infantil gerado por IA + *deepfakes*

Variável por estado

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

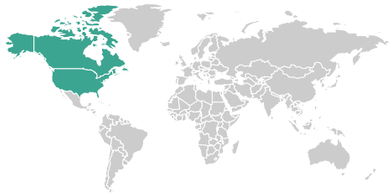
4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA



Os Estados Unidos lideram em abrangência normativa no continente: 45 estados já criminalizam o material ou conteúdo de violência sexual contra crianças e adolescentes gerado por IA.[66] No plano federal, em 2025 foi aprovada a Lei *Take It Down*, que criminaliza a criação e a distribuição de conteúdo sexual profundo gerado por IA sem consentimento, incluindo casos envolvendo crianças e adolescentes, e exige que plataformas retirem rapidamente o material, com penalidades semelhantes às aplicadas ao material convencional.[67]

O Canadá ainda apresenta lacuna crítica em relação à tipificação expressa de material ou conteúdo de violência sexual contra crianças e adolescentes gerado exclusivamente por IA. Embora o Código Criminal já considere crime a criação, distribuição e posse de imagens sexuais de crianças, mesmo quando sinteticamente manipuladas, a legislação não incorpora redação específica para conteúdo produzido por IA, dependendo de enquadramento por analogia.[68] Em 2025, o governo canadense aprovou uma reforma importante do Código Penal, com a Lei de Proteção de Vítimas (*Protecting Victims Act / Bill C-16*), que endurece penas para crimes sexuais contra crianças e adolescentes e proíbe a distribuição de *deepfake*s sexuais não consensuais, mas mesmo esse marco não cria um delito autônomo para material de exploração infantil gerado por IA, mantendo a responsabilização em quadros normativos gerais.[69]

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

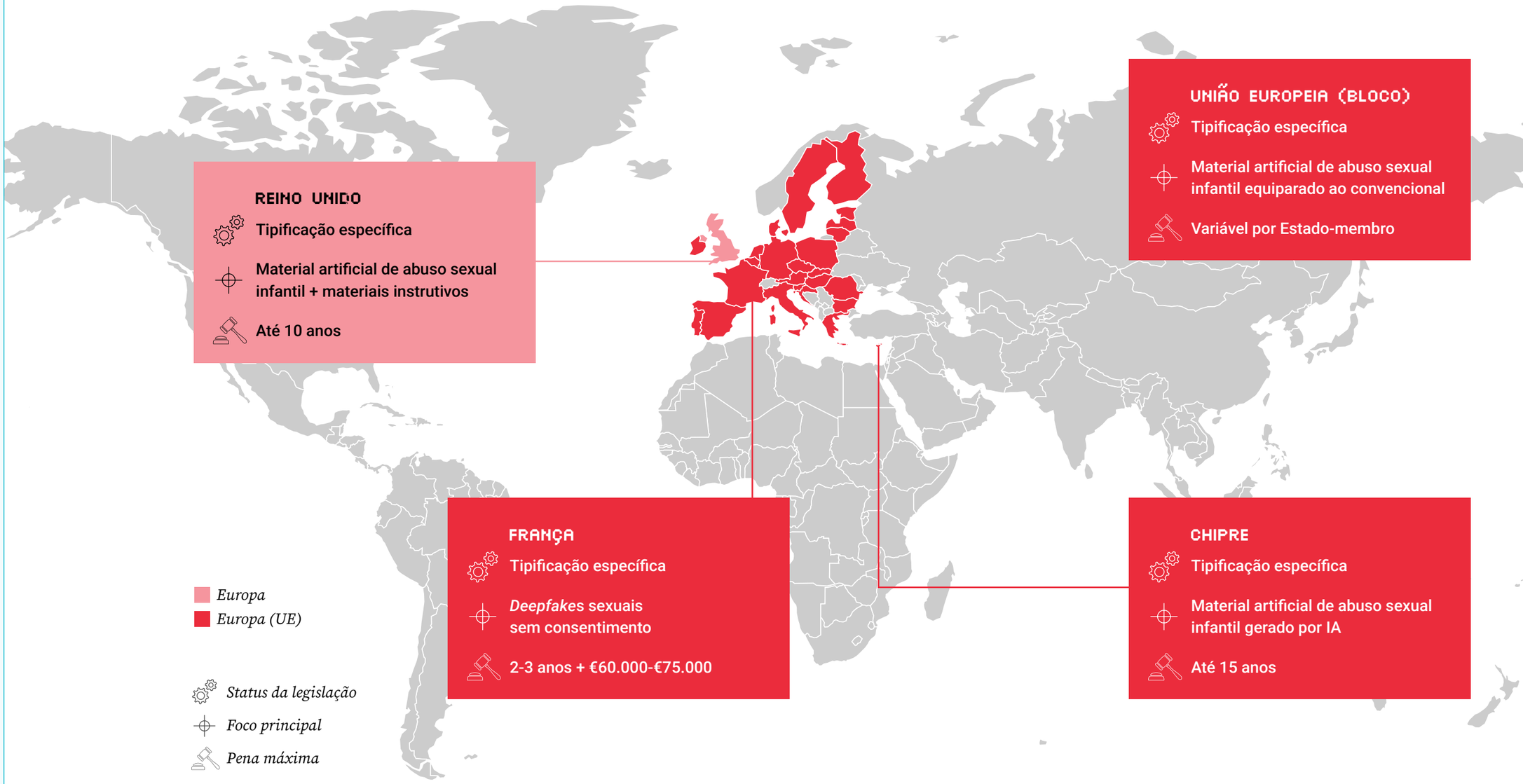
5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

5.3.3. EUROPA



ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA



O panorama europeu é marcado por dois níveis de regulação sobrepostos: o marco comunitário, aplicável a todos os Estados-membros da União Europeia, e as legislações nacionais, em diferentes estágios de desenvolvimento.

No âmbito da União Europeia, o Parlamento Europeu aprovou propostas que buscam, em conjunto, equiparar conteúdos de abuso sexual infantil gerados por IA ao material convencional, incluindo *grooming*, *livestreaming* de exploração sexual e sextorsão, com forte ênfase em responsabilização de provedores e plataformas. O *AI Act* proíbe explicitamente sistemas de IA criados especificamente para gerar ou disseminar material ou conteúdo de violência sexual contra crianças e adolescentes, operando, de forma indireta, como uma barreira legal à disponibilização desse tipo de conteúdo no mercado interno. [70]

DENTRO DO BLOCO, ALGUNS PAÍSES DA UE TAMBÉM SE DESTACAM POR MEDIDAS ADICIONAIS JÁ VIGENTES.

O Chipre criminaliza explicitamente a produção, posse e distribuição de material ou conteúdo de violência sexual contra crianças e adolescentes gerado por IA, com penas de até 15 anos de prisão, regulando inclusive imagens sintéticas baseadas em fotos reais de crianças e adolescentes; a primeira condenação com base nesta lei aconteceu em março de 2026. A França criminaliza a divulgação de conteúdo sexual gerado por algoritmos sem consentimento; a pena base é 2 anos e €60.000. Quando a publicação é feita por serviço de comunicação on-line, sobe para 3 anos e €75.000.

O Reino Unido, fora da União Europeia, avança em direção ao marco mais abrangente já proposto por qualquer país. Em fevereiro de 2025, tornou-se o primeiro país do mundo a criminalizar explicitamente a criação, posse e disseminação de materiais instrutivos — guias e tutoriais — que ensinem como gerar imagens sintéticas de abuso sexual infantil, reconhecendo que a infraestrutura de conhecimento técnico que viabiliza esses conteúdos constitui, em si mesma, um vetor autônomo de risco. A legislação introduz dois tipos penais complementares: o primeiro prevê sanções para a posse, criação ou distribuição de modelos de IA especificamente projetados para gerar material ou conteúdo de violência sexual contra crianças e adolescentes; o segundo criminaliza a posse de guias ou instruções que ensinem como utilizar ferramentas legítimas de IA para produzir imagens ilegais — abordando diretamente o fenômeno da conversão de tecnologias de uso dual para fins criminosos.[71] O *Crime and Policing Bill*, combinado ao *Online Safety Act*, prevê penas de até cinco anos para posse de geradores e até 10 anos para administração de plataformas.[72]

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

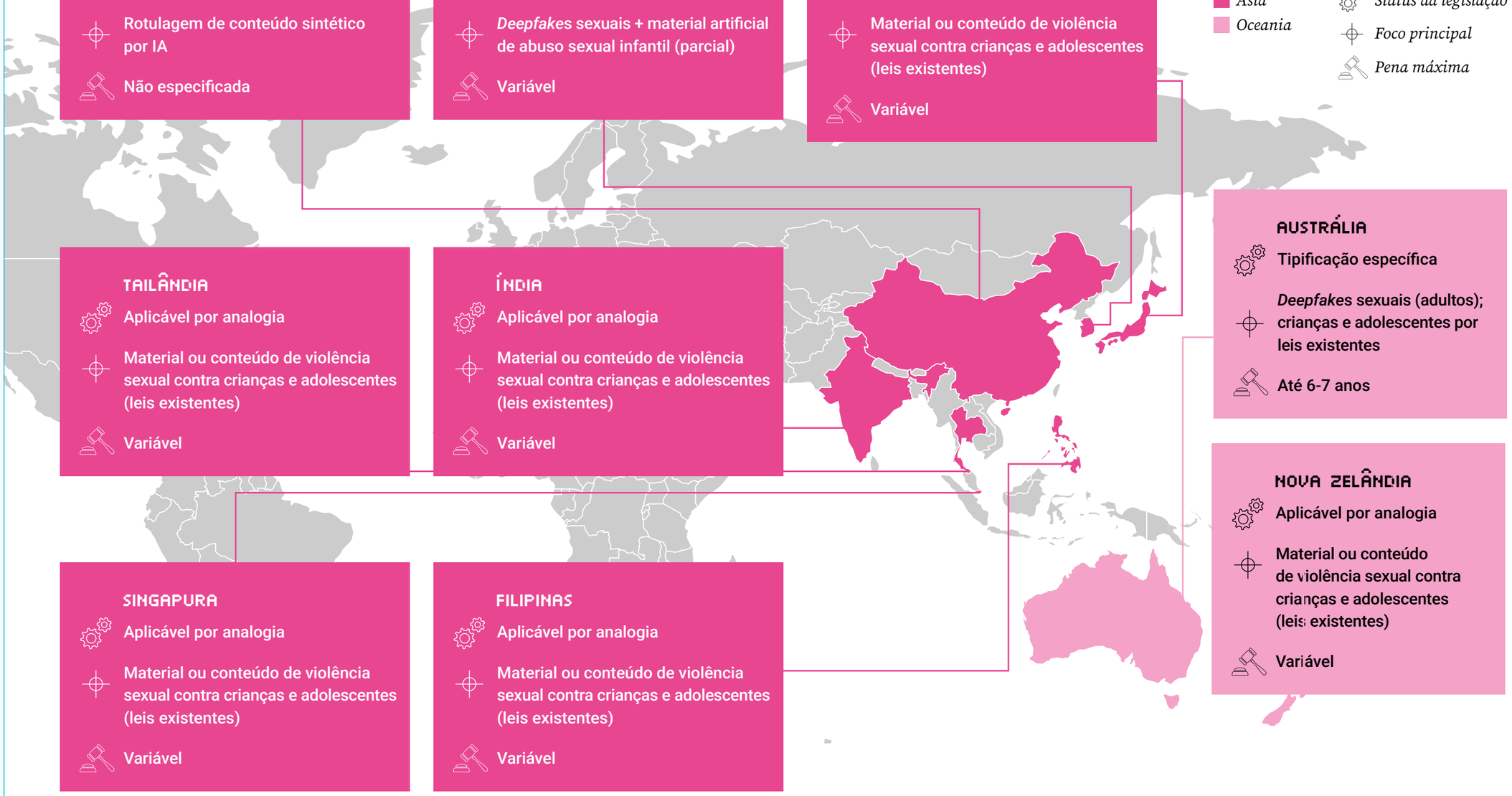
7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

5.3.4. ÁSIA E OCEANIA

Ásia
Oceania

Status da legislação
Foco principal
Pena máxima



CHINA
- Legislação parcial (rotulagem)
- Rotulagem de conteúdo sintético por IA
- Não especificada

COREIA DO SUL
- Tipificação específica
- *Deepfakes* sexuais + material artificial de abuso sexual infantil (parcial)
- Variável

JAPÃO
- Aplicável por analogia
- Material ou conteúdo de violência sexual contra crianças e adolescentes (leis existentes)
- Variável

AUSTRÁLIA
- Tipificação específica
- *Deepfakes* sexuais (adultos); crianças e adolescentes por leis existentes
- Até 6-7 anos

NOVA ZELÂNDIA
- Aplicável por analogia
- Material ou conteúdo de violência sexual contra crianças e adolescentes (leis existentes)
- Variável

TAILÂNDIA
- Aplicável por analogia
- Material ou conteúdo de violência sexual contra crianças e adolescentes (leis existentes)
- Variável

ÍNDIA
- Aplicável por analogia
- Material ou conteúdo de violência sexual contra crianças e adolescentes (leis existentes)
- Variável

SINGAPURA
- Aplicável por analogia
- Material ou conteúdo de violência sexual contra crianças e adolescentes (leis existentes)
- Variável

FILIPINAS
- Aplicável por analogia
- Material ou conteúdo de violência sexual contra crianças e adolescentes (leis existentes)
- Variável

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA



A Coreia do Sul aprovou emenda à Lei de Punição de Crimes Sexuais em resposta à crise nacional de *deepfake*s em 2024, criminalizando conteúdos gerados por IA, inclusive quando adultos são representados como crianças e adolescentes.[73] A nova redação criminaliza a produção, distribuição, posse e até mesmo a visualização de vídeos e imagens contendo *deepfake*s sexualmente explícitos, com penas de até sete anos de prisão para produção e distribuição de conteúdos criminosos gerados por IA, e de até três anos para posse ou visualização. A reforma também aumenta as penas para quem usa tecnologia de IA para gerar conteúdo sexual de pessoas reais, inclusive adultos representados como crianças e adolescentes, embora a tipificação específica para exploração sexual infantil gerada por IA permaneça em debate.

Em 2025, a China aprovou uma diretriz para rotular conteúdos sintéticos gerados por IA.[74] A norma exige que todo material gerado ou alterado por IA seja claramente identificado, com rótulos visíveis e metadados incorporados, além de exigir consentimento, verificação de identidade e registro ou aprovação governamental dos provedores de serviços que utilizam essas tecnologias. Essas medidas são vinculadas a critérios de autorização para lançamento ou listagem de aplicativos, o que impõe obrigações diretas às plataformas de distribuição e pode dificultar a produção e disseminação de materiais artificiais de abuso ou exploração sexual infantil e *deepfake*s sexuais.

Na Austrália, uma nova norma recém-aprovada criou um novo delito específico para o envio não consensual de material sexual *deepfake*, com penas de até seis anos de prisão, e até sete anos em hipóteses agravadas.[75] A norma abrange tanto imagens quanto vídeos criados ou alterados por tecnologia, e se aplica quando o agente sabe ou é conscientemente indiferente quanto à ausência de consentimento da pessoa representada, reforçando a proteção contra abuso de imagens íntimas geradas por IA. Já existem relatos de condenações iniciais e de medidas de reforço da segurança on-line, incluindo obrigações às plataformas para remoção rápida de conteúdo ilícito, embora o foco principal da lei seja material sexual de adultos, já que condutas relacionadas a crianças são reguladas por diplomas específicos.[76]

Vários outros países da Ásia e da Oceania — Singapura, Índia, Tailândia, Filipinas, Japão e Nova Zelândia — enfrentam esse tipo de conteúdo com leis que já existem contra o abuso e a exploração sexual de crianças, sem terem criado uma lei nova e dedicada só ao material gerado por IA, como fizeram o Reino Unido e o Chipre. O alcance dessas leis, porém, varia bastante. Na Índia, nas Filipinas e na Nova Zelândia, a própria legislação já abrange de forma expressa imagens criadas por computador ou por IA — e a Nova Zelândia, inclusive, já tem condenações nesse sentido. Em Singapura e na Tailândia, as definições legais são amplas o bastante para, segundo especialistas, alcançar também o conteúdo artificial. Já o Japão é o caso mais frágil: sua lei protege apenas crianças reais e não cobre desenhos ou imagens totalmente sintéticas. De modo geral, relatórios da ONU, da UNICEF e de organizações não governamentais apontam que a maioria desses países já consegue aplicar suas leis a esse tipo de conteúdo, mas nenhum aprovou até agora uma lei específica e separada voltada apenas ao material ou conteúdo de violência sexual contra crianças e adolescentes gerado por IA.

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

5.3.5.
ÁFRICA



ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA



A África do Sul dispõe do enquadramento mais estruturado do continente, combinando o *Films and Publications Act* e a *Protection of Personal Information Act* para coibir a distribuição de imagens sexuais privadas e o uso indevido de dados pessoais, com aplicação extensiva a *deepfake*s — embora sem tipificação específica para material ou conteúdo de violência sexual contra crianças e adolescentes gerado por IA. [77]

A ausência de marcos normativos específicos no restante do continente africano configura uma lacuna estrutural de especial gravidade, agravada pelo rápido crescimento do acesso digital entre a juventude — uma situação evidente em países como Nigéria, Quênia e Gana, onde políticas de proteção infantil on-line ainda carecem de dispositivos próprios para *deepfake*s e IA.[78] A carência de regulamentação, particularmente no que diz respeito à IA, é recorrentemente apontada em relatórios regionais sobre a proteção de crianças e adolescentes na internet.

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

CAPÍTULO SEIS

Recomendações detalhadas para múltiplos atores

ÍNDICE

1.
DEFINIÇÕES & TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3.
PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4.
ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5.
RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6.
RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

As recomendações contidas nesta seção resultam de um processo aprofundado de análise e diagnóstico realizado pela equipe multidisciplinar da SaferNet Brasil ao longo de dez meses de pesquisa. Elas estão ancoradas nos achados empíricos documentados nos capítulos anteriores — o mapeamento de *deepfakes* sexuais em escolas, a análise de 264 URLs denunciadas, os cinco estudos de caso sobre vetores de risco, a pesquisa de opinião com 679 jovens, a escuta de educadores e o exame das respostas institucionais e legislativas — e dialogam com experiências internacionais que oferecem modelos replicáveis ou adaptáveis ao contexto brasileiro.

O objetivo central é ir além da descrição do problema, propondo caminhos concretos e estratégicos para aprimorar políticas públicas, fortalecer protocolos institucionais e ampliar a capacidade do país em três frentes: prevenção, mitigação de danos e resposta rápida a incidentes. A premissa que orienta todas as recomendações é a de que nenhuma resposta isolada — exclusivamente tecnológica, jurídica, legislativa ou educacional — será suficiente para enfrentar um problema de natureza sistêmica. As recomendações estão organizadas por ator, mas foram concebidas para funcionar de forma integrada.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Após dez meses de pesquisa, a principal recomendação da SaferNet Brasil é que autoridades e indústria tecnológica atuem de forma coordenada na asfixia financeira dos sistemas que produzem e distribuem conteúdo sintético de abuso e exploração sexual de crianças e adolescentes — ou seja, na interrupção do fluxo de recursos que sustenta essas atividades.

RECOMENDA-SE QUE O BRASIL CRIE MECANISMO EQUIVALENTE NO ÂMBITO NACIONAL, COM PARTICIPAÇÃO DO BANCO CENTRAL, DO COAF E DAS PRINCIPAIS INSTITUIÇÕES FINANCEIRAS DO PAÍS. A ASFIXIA FINANCEIRA DEVE OPERAR EM TRÊS FRENTES COMPLEMENTARES.

A primeira frente busca cortar o fluxo financeiro direto. Recomenda-se que o poder público desenvolva protocolos de monitoramento de transações suspeitas vinculadas a sites, *bots* e serviços de *deepfakes* sexuais envolvendo crianças e adolescentes, fundados em indícios objetivos e com preservação dos dados de terceiros não envolvidos. Referência relevante é a orientação emitida em 2020 pela *Financial Crimes Enforcement Network* (FinCEN), dos EUA, com indicadores específicos para identificar transações financeiras associadas à exploração. Recomenda-se ainda integrar plataformas de criptoativos a sistemas de alerta e bloqueio de carteiras, em conformidade com os padrões internacionais de prevenção à lavagem de dinheiro (AML ou *Anti Money Laundering* no inglês) e de identificação de clientes (KYC).

No caso específico da TON Network — blockchain vinculada ao Telegram e utilizada para a conversão de Telegram Stars em criptomoeda, conforme documentado na seção 4.1.1 —, por exemplo, recomenda-se que as autoridades brasileiras solicitem informações à Fragment sobre o volume e a natureza das transações originadas no Brasil, como também incluam no template de declaração de dados dos ESPs, para fins de cumprimento da obrigação prevista no Art. 39 do Decreto n.º 12.880/2026, campos para o fornecimento de informações relevantes para inteligência financeira, a exemplo — mas não limitando-se — a wallets de origem e destino, histórico e rastreabilidade de transações (*hashes* e metadados), presença e vinculação de contas a redes de publicidade digital (*ads networks*), e outros indicadores de atividade econômica ou de perfil de risco associados aos usuários suspeitos.

A segunda frente busca limitar a monetização indireta, atuando sobre lojas de aplicativos, plataformas de anúncios e provedores de hospedagem. Este relatório documentou que Apple e Google retêm um percentual de 30% das transações de compra de Telegram Stars, meio de pagamento utilizado para a compra e venda de material ou conteúdo de violência sexual contra crianças e adolescentes e geração de *deepfake* nudes no Telegram. Igualmente, evidenciou-se a existência de plataformas de conteúdo adulto que funcionam como verdadeiros repositórios de conteúdo ilegal, sem moderação. No tocante às plataformas de anúncios, é necessário aprofundar o estudo sobre o funcionamento das redes **TrafficJunky** e **AdultForce**.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Recomenda-se ao Poder Executivo a criação de Grupo de Trabalho para aperfeiçoar os mecanismos de fiscalização com vistas à desmonetização, rastreamento das transações ilícitas e auditoria dos sistemas de pagamento utilizados pelos aplicativos de nudificação, websites e *bots* do Telegram. Modelos internacionais relevantes incluem a iniciativa **Child Protection Toolkit**, desenvolvida pelo ICMEC.

A terceira frente se concentra na identificação e no bloqueio dos beneficiários econômicos. Recomenda-se que o Estado mapeie a cadeia que envolve desenvolvedores, distribuidores e intermediários, com observância do devido processo legal. Recomenda-se também acionar mecanismos de cooperação internacional para o bloqueio e o confisco de ativos, com base em tratados já ratificados pelo Brasil. Para os casos de *deepfakes* sexuais nas escolas brasileiras, recomenda-se que o MEC, em articulação com o Ministério da Justiça e Segurança Pública e o Ministério dos Direitos Humanos, desenvolva e publique um protocolo nacional de resposta, abrangendo: identificação e registro do caso pela escola e acolhimento da vítima sem revitimização; capacitação de conselheiros tutelares alinhada a um fluxo envolvendo a rede de proteção. O protocolo deve incluir diretrizes específicas para situações em que o infrator é menor de idade e para casos envolvendo professoras como vítimas.

Nesse sentido, a SaferNet Brasil também considera de extrema relevância a implementação de um programa nacional para capacitar delegacias especializadas na investigação de crimes que utilizam IA generativa. O foco dessa capacitação deve incluir a análise forense de imagens sintéticas, o rastreamento de transações com criptomoedas, técnicas de infiltração em fóruns da *dark web* e a colaboração com *hotlines* internacionais. É igualmente recomendada a expansão do uso de ferramentas de IA investigativas — como o Nêmesis e os classificadores da Thorn — para todas as unidades da federação.

METADADOS GERADOS DURANTE AS TRANSAÇÕES

CAMADA DA APLICAÇÃO (WEB2 – TELEGRAM / BOT API)
Quando um usuário compra ou gasta Telegram Stars (XTR) dentro de um *bot*, o servidor do Telegram e o servidor do desenvolvedor do *bot* registram metadados específicos através de objetos da API (como Invoice, PreCheckoutQuery e SuccessfulPayment):

USER_ID
O identificador numérico único, permanente e universal do usuário no Telegram (mesmo que ele mude de @username ou nome de exibição).

CHAT_ID
O ID da conversa onde a transação foi iniciada.

INVOICE_PAYLOAD
Uma *string* customizada definida pelo desenvolvedor do *bot* (ex: premium_credits_user_9482). Esse metadado é crucial porque vincula a transação financeira ao produto/comando específico gerado (por exemplo, o comando que acionou a IA generativa).

TELEGRAM_PAYMENT_CHARGE_ID E PROVIDER_PAYMENT_CHARGE_ID
Identificadores únicos da transação gerados pelo Telegram para controle financeiro e conciliação.

TIMESTAMP
Registro de data e milissegundo exatos da criação da fatura, da autorização e da conclusão do pagamento.

ENDEREÇO IP E LOGS DE DISPOSITIVO
Coletados nativamente pelos servidores do Telegram no momento em que o usuário abre a interface de checkout.

Fonte: https://core.telegram.org/bots/payments

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

6.2. RECOMENDAÇÕES PARA O PODER LEGISLATIVO E O PODER JUDICIÁRIO

Esta seção se fundamenta no consenso internacional estabelecido pelo manifesto global *No to Nudify*, lançado em fevereiro de 2026 e endossado por mais de uma centena de organizações de proteção à infância e direitos digitais em todo o mundo, incluindo a SaferNet Brasil. Partindo da premissa de que as tecnologias dedicadas à "nudificação" por IA não têm qualquer finalidade social legítima e atuam estritamente como vetores de exploração sexual e violência de gênero, recomendam-se as seguintes diretrizes para o poder legislativo:

Sobre a criminalização e proibição de ferramentas de nudificação, recomenda-se que a legislação brasileira incorpore, a exemplo do *Crime and Policing Bill* do Reino Unido, a criminalização das ferramentas de nudificação e dos materiais instrutivos — guias e tutoriais — que ensinam a produzir imagens de abuso sexual infantil com IA. O relatório documentou, nos estudos de caso da *dark web* (seções 4.2.1 e 4.2.2), a existência de centenas de guias e manuais disponíveis para esse fim, e a porosidade entre tutoriais acessíveis na internet aberta e fóruns criminosos.

Além disso, é necessário ressaltar a necessidade de uma responsabilização civil e criminal de toda a cadeia de fornecimento. O arcabouço legal deve estabelecer a responsabilização civil e criminal de proprietários, desenvolvedores e plataformas que hospedem ou que viabilizem, facilitem ou aufiram lucro com a criação ou circulação de imagens nudificadas não consensuais de pessoas reais. Nesse sentido, é fundamental asfixiar economicamente o ecossistema cibercriminoso, estendendo a responsabilização a quem fornece a infraestrutura técnica que sustenta e monetiza a violência.

No escopo do diagnóstico desta pesquisa, cumpre contextualizar as contribuições de estudo anterior desenvolvido pela SaferNet Brasil no âmbito do Grupo de Trabalho instituído pela Portaria CGJ 274/2023-GSEC do Tribunal de Justiça da Bahia (TJBA) voltado a normativos e outras medidas relacionadas à qualificação da atuação do Poder Judiciário e de sua relação com demais instituições para a proteção de direitos no ambiente digital. Os achados da referida articulação institucional evidenciaram a necessidade premente de aprimorar o gerenciamento de dados por meio de ferramentas de *Business Intelligence* (BI) e de promover o aperfeiçoamento das *Tabelas Processuais Unificadas* (TPU). Essa reestruturação projeta o desmembramento da classificação por tipos penais previstos na Lei 11.829/2008, a inclusão de novas condutas previstas na legislação extravagante e o agrupamento dos assuntos existentes em uma árvore temática especificamente dedicada aos direitos digitais.

Como desdobramento prático desse diagnóstico: recomenda-se ao Conselho Nacional de Justiça (CNJ) e ao Conselho Nacional do Ministério Público (CNMP) a criação e a implementação de mecanismos específicos de indexação, triagem e monitoramento — tais como marcadores, *tags* ou subclasses específicas nas Tabelas Processuais Unificadas (TPU) — voltados à identificação de processos judiciais e procedimentos de investigação criminal que envolvam indícios de utilização de ferramentas de Inteligên-

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

cia Artificial Generativa (notadamente para a criação de *deepfakes*, conteúdos sintéticos ou manipulação digital de imagens e vozes) no cometimento de crimes tipificados na Lei n.º 11.829/2008.

Esta medida mostra-se imprescindível não apenas para mapear a evolução dessa nova vertente tecnológica da criminalidade digital, mas também para garantir a celeridade prioritária na tramitação, subsidiar a atuação estratégica dos órgãos de persecução penal e do Poder Judiciário, e viabilizar a produção de dados estatísticos confiáveis que orientem futuras políticas públicas de proteção integral à criança e ao adolescente.

Complementarmente, conforme destacado pela representação da SaferNet Brasil na oportunidade, essa diretriz deve contemplar a criação de um fluxo de validação dos assuntos processuais por servidores do próprio Poder Judiciário, providência indispensável para sanar as recorrentes inconsistências geradas pelo modelo atual, no qual a classificação inicial dos temas correlatos compete exclusivamente à advocacia ou à Defensoria Pública no ato do peticionamento.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

As empresas que desenvolvem, hospedam, financiam ou monetizam tecnologias digitais ocupam posição central na arquitetura de proteção contra a exploração sexual sintética de crianças e adolescentes. Recomenda-se que adotem o princípio da segurança desde a concepção (*Safety by Design*), incorporando salvaguardas técnicas, jurídicas e éticas em todas as etapas do ciclo de vida de seus produtos — e não apenas como resposta reativa a incidentes já consumados.

Este relatório documentou que o ecossistema de código aberto constitui o principal vetor técnico de produção de material ou conteúdo de violência sexual contra crianças e adolescentes por IA, com modelos como Stable Diffusion, Flux, Wan 2.x e Qwen sendo explicitamente utilizados e recomendados em fóruns criminosos. Recomenda-se que as empresas desenvolvedoras de modelos abertos implementem avaliações obrigatórias de segurança infantil antes da publicação dos parâmetros dos modelos de geração de imagem, incluindo testes adversariais focados em material ou conteúdo de violência sexual contra crianças e adolescentes criado por IA e imagens íntimas não consensuais — como o protocolo adotado pela Black Forest Labs antes do lançamento do Flux.

Recomenda-se ainda que as licenças de uso dos modelos proíbam explicitamente a geração de material ou conteúdo de violência sexual contra crianças e adolescentes e de imagens íntimas não consensuais, com cláusulas que responsabilizem distribuidores que redistribuam versões sem salvaguardas. Reconhece-se, contudo, que a disponibilização dos parâmetros dos modelos torna essas proteções estruturalmente reversíveis — razão pela qual as medidas de mitigação na origem devem ser complementadas por ações de fiscalização, detecção e responsabilização ao longo de toda a cadeia.

Recomenda-se que plataformas de hospedagem de modelos de terceiros implementem avaliações de segurança infantil como requisito prévio à publicação de novos modelos de geração de imagem, e que realizem auditorias retroativas dos modelos já hospedados. Recomenda-se a expansão dos filtros conceituais (como os *Semi-Permeable Membranes* do CivitAI) e a integração com bases internacionais de *hashes* de conteúdos de abuso ou exploração sexual infantil conhecidos por analistas e autoridades públicas.

Às empresas do setor financeiro, dos meios de pagamento e das plataformas de criptoativos, reforça-se a necessidade de fortalecer os controles internos de prevenção à lavagem de dinheiro e de conhecimento do cliente (AML/KYC), com adaptação dos modelos de risco para identificar fluxos típicos da monetização de conteúdo sintético de abuso. Recomenda-se a participação em iniciativas setoriais de troca segura de indicadores de fraude e crime, sempre com observância dos marcos legais de proteção de dados pessoais e bancários e com salvaguardas que impeçam a estigmatização ou o bloqueio indevido de usuários legítimos.

Aos provedores de hospedagem, infraestrutura em nuvem, registradores de domínio, redes de distribuição de conteúdo (*CDN*) e

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

plataformas de anúncios, recomenda-se a aplicação consistente e tempestiva dos termos de uso já vigentes, com suspensão imediata de contas, domínios e campanhas vinculados à comercialização ou à divulgação de conteúdo de abuso. Recomenda-se ainda o estabelecimento de canais especializados para o recebimento de notificações qualificadas oriundas de autoridades, *hotlines* e organizações da sociedade civil, com fluxos internos que assegurem análise rápida e fundamentada das demandas. Os mecanismos de moderação devem operar de forma a não comprometer a privacidade ou a integridade de dados de terceiros não envolvidos.

De forma transversal a todos os setores, também se recomenda que as empresas instituam estruturas internas de governança dedicadas à proteção de crianças e adolescentes, incluindo a designação de pontos de contato responsáveis pela interlocução com autoridades públicas, conselhos tutelares e organizações de defesa de direitos. A capacitação contínua de equipes técnicas, jurídicas e de moderação sobre as especificidades dos crimes digitais contra a infância é condição indispensável para que as medidas técnicas se traduzam em proteção efetiva. Por fim, recomenda-se que as empresas se comprometam publicamente com princípios de prestação de contas (*accountability*), submetendo-se a auditorias independentes e divulgando, de forma clara, as políticas adotadas, os resultados alcançados e as limitações persistentes, em diálogo permanente com a sociedade civil e os órgãos reguladores.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

6.4. RECOMENDAÇÕES PARA O SETOR EDUCACIONAL

A escola ocupa posição estratégica e insubstituível no sistema de garantia de direitos de crianças e adolescentes. No entanto, o diagnóstico deste relatório revela que escolas não dispõem de protocolo padronizado de resposta às violências on-line. O cenário atual é marcado pela desorientação institucional de educadores e, em alguns casos, pela minimização dos incidentes, o que contribui para a percepção de desresponsabilização dos infratores e isolamento das vítimas. Importante lembrar que o Brasil já tem um amplo conjunto de leis e diretrizes educacionais que estabelecem o dever das escolas de incluir nos currículos de Educação Básica (educação infantil, ensino fundamental e ensino médio) e na formação continuada de professores temas relacionados ao uso seguro e consciente das tecnologias e os impactos desses usos na saúde física e mental dos estudantes.

Essa base legal ganhou capilaridade pedagógica com a aprovação da Base Nacional Comum Curricular (BNCC) em 2017, que consagrou como uma de suas competências gerais a “Cultura Digital”, compreendida como o direito dos estudantes de desenvolver habilidades para compreender, utilizar e criar tecnologias digitais de informação e comunicação de forma ética, crítica e reflexiva. Esse compromisso foi aprofundado com o complemento da BNCC para o Ensino de Computação em 2022, com a instituição da Política Nacional de Educação Digital em 2023, e com a Resolução do Conselho Nacional de Educação n.º 2/2025, que trouxe orientações para a integração curricular obrigatória, a partir de 2026, da educação digital e midiática nas escolas.

No Ensino Médio, as habilidades estabelecidas pela BNCC Computação determinam que os estudantes compreendam o impacto das transformações tecnológicas na segurança e privacidade, capacitando-os a reconhecer os limites da IA. As diretrizes preveem ainda a análise crítica sobre a confiabilidade das informações e as interações em comunidades on-line, avaliando seus reflexos na saúde física e mental, além de conscientizar sobre direitos e deveres na Internet para promover a convivência e a denúncia de condutas abusivas on-line.

Para materializar essas obrigações legais, a escola precisa deixar de ser o local onde os incidentes de violência digital repercutem para ser o espaço onde eles são prevenidos. Por isso, recomenda-se a integração curricular de conteúdos como manipulação digital de imagens, consentimento, empatia e os impactos psicossociais da circulação não consensual de materiais sexualizados, de forma transdisciplinar e não episódica, alinhando-se ao já previsto pela BNCC. No campo da privacidade e segurança digital, é crucial que os currículos de educação digital e midiática orientem os alunos sobre proteção de dados pessoais, gestão da própria imagem e mitigação dos riscos em plataformas digitais. O objetivo é consolidar uma cidadania digital crítica, onde jovens compreendam o ecossistema digital e reconheçam seus direitos e responsabilidades para tornar os ambientes digitais mais seguros.

É indispensável a formação continuada de professores e gestores sobre o funcionamento de ferramentas de IA generativa e de apli-

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

cativos de nudificação — o desconhecimento técnico por parte dos adultos expõe a falha de resposta institucional e impede a identificação precoce de riscos.

AS ESCOLAS DEVEM INSTITUIR E DAR AMPLA PUBLICIDADE A PROTOCOLOS FORMAIS PARA O RECEBIMENTO DE DENÚNCIAS, ACOLHIMENTO DAS VÍTIMAS E ENCAMINHAMENTO DOS CASOS. ESSES PROCEDIMENTOS DEVEM SER INSTITUCIONAIS E DEVEM GARANTIR QUE NÃO GEREM REVITIMIZAÇÃO, EVITANDO JUÍZOS MORAIS.

A escola não deve agir de forma isolada. A gestão de incidentes e as ações preventivas devem ocorrer em articulação permanente com as famílias, Secretarias de Educação, Conselhos Tutelares, MEC, e organizações da sociedade civil. Essa governança em rede é fundamental para consolidação de fluxos interinstitucionais de proteção e para compartilhamento de boas práticas de educação digital.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

O RELATÓRIO EVIDENCIOU LACUNAS SIGNIFICATIVAS DE DADOS SOBRE O FENÔMENO NO BRASIL: A DEPENDÊNCIA DE CASOS NOTICIADOS PELA MÍDIA COMO FONTE PRIMÁRIA, A INEXISTÊNCIA DE PROTOCOLO UNIFICADO DE REGISTRO E A SUBNOTIFICAÇÃO ESTRUTURAL DOS CASOS COMPROMETEM QUALQUER TENTATIVA DE MENSURAÇÃO PRECISA.

Por esse motivo, recomenda-se que organizações de pesquisa e instituições acadêmicas desenvolvam estudos sobre a prevalência de *deepfakes* sexuais em escolas brasileiras em diferentes regiões considerando recortes de classe, gênero e etnia, com amostras representativas e metodologias padronizadas — superando a dependência atual de mapeamentos baseados em cobertura jornalística.

Recomenda-se que organizações da sociedade civil atuem como articuladoras entre os diferentes atores do ecossistema de proteção — escolas, famílias, autoridades e empresas —, produzindo materiais de referência com respaldo técnico e promovendo a capacitação de educadores, gestores escolares e profissionais de segurança pública. A SaferNet Brasil, por sua atuação como *Safer Internet Centre* reconhecido internacionalmente e responsável pelo canal de denúncias de crimes cibernéticos e violações de direitos humanos on-line, continuará desempenhando esse papel.

Recomenda-se ainda que organizações de pesquisa monitorem sistematicamente a evolução das ferramentas de IA generativa utilizadas para a produção de material ou conteúdo de violência sexual contra crianças e adolescentes, incluindo o surgimento de novos modelos de código aberto, novas técnicas de evasão de filtros e novas formas de monetização. O ritmo de inovação técnica documentado neste relatório exige monitoramento contínuo e atualização permanente das estratégias de enfrentamento.

Por fim, recomenda-se que organizações da sociedade civil mantenham redes de apoio acessíveis para vítimas de *deepfakes* sexuais e outras formas de violência sexual mediada por IA, incluindo orientação jurídica, acolhimento psicossocial e suporte para remoção de conteúdos — serviços que a SaferNet Brasil já oferece por meio de seus canais de atendimento e que precisam ser ampliados diante da escalada documentada neste relatório.

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

CAPÍTULO SETE

Considerações finais

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Esta pesquisa de quase um ano sintetiza o cenário complexo e sem precedentes enfrentado com o impacto do desenvolvimento da IA generativa em todos os setores da sociedade brasileira. O avanço acelerado e a popularização dos sistemas de inteligência artificial generativa colocaram os direitos fundamentais de crianças e adolescentes em uma encruzilhada histórica.

A urgência do tema se impõe pela radical mudança na velocidade e na escala de produção desse conteúdo. Pela primeira vez, a literatura e o monitoramento de campo documentam uma violência capaz de provocar processos severos de dissociação e estigma nas vítimas. O sofrimento psíquico, o medo e o trauma experimentados por uma adolescente ao ter sua imagem despida ou manipulada são reais, concretos e devastadores, ainda que a imagem em si tenha sido inteiramente fabricada por um algoritmo.

A violência sexual mediada por IA no Brasil tem um recorte estrutural que não pode ser tratado como coincidência: ela é cirurgicamente direcionada contra meninas e mulheres. Os 17 casos mapeados em instituições de ensino brasileiras demonstraram que a totalidade das centenas de vítimas eram meninas, adolescentes e professoras, enquanto a totalidade dos infratores eram meninos da mesma faixa etária. As consequências para as jovens sobreviventes são avassaladoras. O estigma e a velocidade de disseminação nas redes impõem uma espécie de "morte social" no ambiente escolar.

Como já endossado pelo manifesto global *No to Nudify*, lançado em fevereiro de 2026 por mais de uma centena de organizações internacionais e mencionado em capítulos anteriores, as ferramentas de nudificação não têm qualquer utilidade social ou finalidade legítima e, na realidade, atuam estritamente como vetores de exploração sexual de crianças e adolescentes e violência de gênero. Por esse motivo, a SaferNet Brasil reforça a necessidade de atuação do governo brasileiro e do legislativo nacional para reforçar uma proibição ao desenvolvimento de tais sistemas, ferramentas ou produtos em solo nacional.

Os achados empíricos deste relatório revelam que, em 2026, produzir material ou conteúdo de violência sexual contra crianças e adolescentes exige menos esforço do que criar uma conta de e-mail. Como destrinchado ao longo do capítulo de evidências deste documento, as comunidades de cibercriminosos dedicados à exploração sexual infantojuvenil digital têm diariamente aperfeiçoado ferramentas de IA generativa capazes de produzir e reproduzir múltiplas formas de violência sexual contra crianças e adolescentes de todo o mundo.

O exemplo mais emblemático é o gerador de imagens apresentado entre os estudos de caso deste relatório. Em março de 2026, os membros dessa comunidade já tinham gerado mais de 510 mil imagens de abuso sexual de meninas. Na prática, isso significa mais de 1.300 novas imagens geradas por dia, ou cerca de uma imagem nova a cada minuto. Já as buscas e criações com descrições para crianças e adolescentes brasileiros em cenas de abuso sexual cresceram quase 38% em apenas dois meses nesse mesmo site.

O aprimoramento tecnológico desses sistemas é descentralizado e impulsionado ativamente pela própria comunidade de usuários por meio de mecanismos de avaliação e aprendizado por reforço com feedback humano (*Reinforcement Learning from Human Feedback* – RLHF).

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

MAIS DE **1.300** IMAGENS POR DIA

RETRATANDO MENINAS

OU APROXIMADAMENTE UMA NOVA IMAGEM A CADA MINUTO

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

Cibercriminosos também estão desenvolvendo novas aplicações para revitimizar crianças e adolescentes que já sofreram alguma forma de abuso sexual e tiveram registros desses crimes publicados na internet, submetendo-as a um novo ciclo de violência por meio da manipulação e redistribuição de suas imagens. E, ao mesmo tempo, com essa nova crise, crianças e adolescentes que nunca passaram por uma violência do tipo no mundo real se tornam, pela primeira vez, vítimas de uma violência sexual inteiramente digital.

O conjunto de evidências reunidas nesta pesquisa permite afirmar que a violência sexual contra crianças e adolescentes mediada por inteligência artificial não é um fenômeno emergente: é uma crise instalada. A eliminação quase total da barreira técnica representa uma ruptura sem precedentes. Em 2026, qualquer pessoa com um smartphone e uma fotografia pode gerar conteúdo sexualmente explícito envolvendo crianças em segundos e com pouco dinheiro. Essa facilidade não apenas amplia o volume de material em circulação a um ritmo que já ultrapassa a capacidade de resposta das autoridades, como altera fundamentalmente o perfil dos agressores: adolescentes produzem conteúdos de colegas de escola, usuários casuais se tornam consumidores recorrentes e comunidades inteiras se organizam em torno do aperfeiçoamento coletivo dessas técnicas.

O aperfeiçoamento técnico desses produtos, sistemas e ferramentas de IA generativa nessas comunidades cibercriminosas tem um único propósito: a reprodução em escala industrial da violência sexual contra crianças e adolescentes. A retórica interna dessas comunidades busca neutralizar o peso moral da conduta sob a premissa de que as representações são artificiais e funcionariam como uma alternativa para proteger crianças e adolescentes no mundo real. Trata-se de uma racionalização contraditória e sem amparo científico, estruturada para tentar se evadir da responsabilidade legal, normalizar o consumo de abusos e violar as diretrizes constitucionais brasileiras e os tratados internacionais de proteção integral à infância.

Diante deste diagnóstico, o ecossistema de proteção à infância se encontra em uma corrida contra o tempo e em direção a um cenário devastador. A velocidade da inovação tecnológica na capacidade generativa avança em um ritmo incomparavelmente superior ao desenvolvimento de salvaguardas técnicas, aos marcos regulatórios e às capacidades investigativas do Estado. A disponibilização sem travas dos parâmetros de modelos de código aberto e a proliferação de aplicações descentralizadas indicam que as respostas das empresas de forma voluntária e isolada são, até o momento, insuficientes para conter uma infraestrutura criminosa que opera em rede.

Contudo, este cenário não deve ser lido sob a ótica do fatalismo ou da paralisia. O Brasil e a comunidade internacional ainda dispõem de uma estreita janela de oportunidade para moldar o futuro da integridade digital. Estamos diante da necessidade urgente de uma coordenação intersetorial robusta. Para evitar que a sociedade atinja um ponto de não retorno nos próximos anos — em que a proliferação exponencial de materiais sintéticos de abuso sobrecarregue completamente as autoridades e impossibilite o resgate de vítimas reais —, a resposta reativa e fragmentada precisa ser substituída por uma estratégia de rede.

Por fim, é por este motivo que a SaferNet Brasil continuará desenvolvendo seu trabalho de pesquisa, a fim de contribuir com as autoridades, comunidade técnica e especialistas em proteção à infância para que o enfrentamento da

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

violência sexual mediada por IA deixe de ser uma reação fragmentada e passe a se constituir como uma política de Estado permanente, baseada em evidências, dotada de capacidade técnica e orientada pela centralidade dos direitos de crianças e adolescentes. Os dados, os relatos e os estudos de caso reunidos neste relatório não são um ponto de chegada, mas um ponto de partida: um chamado urgente para que governos, legisladores, plataformas, instituições de ensino, organismos internacionais e a sociedade civil atuem de forma coordenada, antes que a escala da violência torne irreversíveis os danos a uma geração inteira de jovens brasileiros.

ÍNDICE

1.
DEFINIÇÕES
& TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL
MEDIADA POR
IA EM ESCOLAS
BRASILEIRAS

3.
PONTOS DE VISTA
DE DIFERENTES
ATORES
ENVOLVIDOS
NO FENÔMENO

4.
ECOSSISTEMA
DIGITAL QUE
FACILITA A
VIOLÊNCIA SEXUAL
MEDIADA POR IA

5.
RESPOSTAS
DE DIFERENTES
ATORES AO
PROBLEMA DA
VIOLÊNCIA SEXUAL
MEDIADA POR IA

6.
RECOMENDAÇÕES
DETALHADAS PARA
MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES
FINAIS

SOBRE ESTA
PESQUISA

Sobre esta pesquisa

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

DADOS BÁSICOS E RESUMO

Esta pesquisa investiga o uso de inteligência artificial generativa por crianças, adolescentes e jovens adultos brasileiros para a criação de material ou conteúdo de violência sexual contra crianças e adolescentes ou de nudez não consentida. A análise se estrutura em torno de três eixos centrais: (i) o fenômeno dos *deepfake*s sexuais no ambiente escolar; (ii) o mapeamento das ferramentas tecnológicas empregadas na produção e na disseminação desses conteúdos; e (iii) o exame dos debates institucionais (autoridades, terceiro setor, sociedade civil e empresas) e legislativos em curso no país sobre o tema.

COMO CITAR ESTE RELATÓRIO?

Este relatório é disponibilizado sob a licença Creative Commons Atribuição–Não Comercial–Sem Derivações 4.0 Internacional (CC BY-NC-ND 4.0). Permite-se copiar e redistribuir o material em qualquer meio ou formato, desde que seja fornecida atribuição adequada, sem uso comercial e sem qualquer modificação do conteúdo original.

Exemplo de citação:
SAFERNET BRASIL, 2026. Rostos fabricados, danos reais: um panorama dos *deepfakes* sexuais entre jovens brasileiros. Disponível em bit.ly/rostosfabricados.

A SAFERNET BRASIL UTILIZOU FERRAMENTAS DE INTELIGÊNCIA ARTIFICIAL GENERATIVA NA REVISÃO DESTE DOCUMENTO, SEGUINDO AS DIRETRIZES DE USO RESPONSÁVEL DA ORGANIZAÇÃO.

EXPEDIENTE

PUBLICAÇÃO
SaferNet Brasil

APOIO
SafeOnline e Tech Coalition

COORDENAÇÃO DE PESQUISA
Juliana Cunha

PESQUISA, ANÁLISE DE DADOS E GRÁFICOS
Sofia Schurig

REDAÇÃO
Thiago Tavares, Juliana Cunha, Sofia Schurig

REVISÃO
Renata Leal (Ibirá Comunicação)

COLABORAÇÃO
Bianca Orrico, Isabella Ferro, Júlia Fernandes, Guilherme Alves

DIREÇÃO DE ARTE E DIAGRAMAÇÃO
Natália Durães Design (Ibirá Comunicação)

DATA DE PUBLICAÇÃO
Novembro/2026

ISBN
978-85-65151-12-2

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

METODOLOGIA DO MAPEAMENTO

O mapeamento de episódios de *deepfake*s sexuais em instituições de ensino brasileiras foi conduzido entre setembro de 2025 e fevereiro de 2026, combinando três fontes de dados complementares. A primeira consistiu em um formulário on-line permanente, mantido pela SaferNet Brasil, para o recebimento de relatos diretos de vítimas, familiares, educadores e profissionais de redes de proteção. A segunda consistiu no monitoramento ativo e sistemático de matérias jornalísticas, por meio de alertas configurados no Google Alerts e de buscas periódicas com palavras-chave relacionadas ao fenômeno "nudificação de alunos", "*deepfake* em escola", etc). A terceira fonte foi um questionário específico, lançado pela SaferNet Brasil em julho de 2025, voltado ao recebimento de relatos de casos de *deepfake*s sexuais em escolas brasileiras ou envolvendo mulheres e adolescentes como vítimas. O questionário recebeu 105 respostas, das quais duas resultaram na incorporação de novos casos ao mapeamento.

Os casos identificados foram catalogados e classificados segundo as seguintes variáveis: unidade da federação, tipo de instituição (pública ou privada), número estimado de vítimas e de infratores, gênero e faixa etária dos envolvidos, natureza do conteúdo produzido e evidência de comercialização. A estimativa do número de vítimas e infratores se baseou nas informações disponíveis nas fontes consultadas, o que constitui uma limitação metodológica — os dados refletem o que foi reportado à SaferNet Brasil ou noticiado pela imprensa, não a totalidade dos casos ocorridos. O baixo índice de conversão do questionário (1,9% das respostas geraram casos novos) reforça a hipótese de subnotificação discutida na seção 2.4 e sugere que a maioria das vítimas que responderam ao formulário relatou casos já identificados por outras vias.

Registra-se que, em pré-divulgações do mapeamento antes do lançamento oficial desta pesquisa, a SaferNet Brasil informou a ocorrência de casos em 10 unidades da federação. A informação foi corrigida para 11 estados: AL, BA, CE, MG, MT, PA, PB, PE, RJ, RS e SP. A discrepância decorreu de erro na identificação das unidades da federação, e não na contagem total de casos mapeados.

METODOLOGIA DA SURVEY

A pesquisa de opinião foi aplicada por meio de formulário on-line com linguagem adaptada ao público de 16 a 21 anos, divulgado por anúncios patrocinados no Instagram entre janeiro e março de 2026. Do total de 784 respondentes, 90 optaram por não participar da pesquisa e 13 respostas foram consideradas inválidas por apresentarem mais de um gênero informado. Assim, a amostra final analisada foi composta por 679 jovens.

A composição da amostra apresentou predominância de meninos cisgênero (501 respondentes, 73,8%), seguidos por meninas cisgênero (91, 13,4%), respondentes que preferiram não declarar gênero (68, 10,0%), pessoas não-binárias (13, 1,9%), meninos transgênero (5, 0,7%) e meninas transgênero (1, 0,1%). A sobrerrepresentação masculina pode estar associada ao mecanismo de distribuição dos anúncios no Instagram, cujo algoritmo prioriza a entrega com base em perfis de engajamento — o que pode ter resultado em maior exposição ao público masculino. Essa limitação amostral deve ser considerada na interpretação dos resultados desagregados por gênero.

ÍNDICE

1.
DEFINIÇÕES & TIPOLOGIAS

2.
VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3.
PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4.
ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5.
RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6.
RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7.
CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

METODOLOGIA DA ANÁLISE DE FERRAMENTAS E PLATAFORMAS

A análise adotou delineamento misto em duas etapas complementares. A primeira consistiu na triagem e classificação qualitativa das denúncias recebidas pela central da SaferNet Brasil, com identificação de padrões recorrentes relativos a perfis de autores e vítimas, plataformas envolvidas e tipologias de conteúdo. A segunda compreendeu o mapeamento das rotas de produção e disseminação — modelos de IA, tutoriais, fluxos financeiros e dinâmicas comunitárias —, conduzido por meio de navegação controlada em ambientes da *surface web* e da *dark web*.

A amostra inicial foi constituída por 264 URLs extraídas da base de denúncias da SaferNet Brasil, selecionadas a partir de palavras-chave presentes nos relatos dos denunciantes (como *deepfake*, *deepnude*, IA, etc.) e de domínios previamente identificados como recorrentes na veiculação de material ou conteúdo de violência sexual contra crianças e adolescentes. Ao longo dos estudos de caso, URLs adicionais foram incorporadas à análise. Cada URL foi classificada manualmente, resultando em 308 registros — já que uma mesma URL podia conter mais de um tipo de conteúdo.

A classificação dos conteúdos obedeceu a três eixos não excludentes: (i) faixa etária das figuras representadas (crianças, adolescentes ou personagens fictícios inspirados nesses grupos); (ii) natureza do conteúdo (nudez parcial ou total, simulação de atos sexuais, cenas de exploração ou coerção); e (iii) origem do material (produção real, conteúdo integralmente gerado por IA ou material manipulado digitalmente). A não exclusividade entre eixos implica que um mesmo registro pode ser classificado simultaneamente em múltiplas categorias.

O acesso às comunidades e ferramentas analisadas foi realizado em equipamentos dedicados, em conformidade com os protocolos de segurança institucional da SaferNet Brasil. Conteúdos que indicassem risco iminente a crianças ou adolescentes foram encaminhados às equipes competentes para providências legais.

No estudo de caso sobre a segunda plataforma (seção 4.3.2), a estimativa de idade das vítimas no momento da captura das imagens originais se baseou em análise cronológica comparativa: as imagens manipuladas foram cotejadas com os registros originais disponíveis nos perfis das vítimas, cruzando-se a data de publicação com a data de nascimento. Nos casos em que os registros originais haviam sido removidos ou publicados em formatos efêmeros, a idade foi estimada a partir de marcadores contextuais — características físicas datáveis (comprimento e cor do cabelo, tatuagens, piercings), elementos tecnológicos (modelo de dispositivo visível) e metadados disponíveis.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

1. CNN Brasil. **Brasil registrou um estupro de crianças e adolescentes a cada 8 minutos, aponta Unicef.** Disponível em: <https://www.cnnbrasil.com.br/nacional/brasil-registrou-um-estupro-de-criancas-e-adolescentes-a-cada-8-minutos-aponta-unicef/>.

2. BRACKET FOUNDATION. A new threat for online child sexual exploitation and abuse. Disponível em: <https://unicri.org/sites/default/files/2024-09/Generative-AI-New-Threat-Online-Child-Abuse.pdf>.

3. STANFORD UNIVERSITY. **Identifying and Eliminating CSAM in Generative ML Training Data and Models.** Disponível em: <https://stacks.stanford.edu/file/druid:kh752sm9123/ml_training_data_csam_report-2023-12-23.pdf>.

4. HU, Edward J. et al. **LoRA: Low-Rank Adaptation of large language models.** 2021. Disponível em: <http://arxiv.org/abs/2106.09685>.

5. WILSON CENTER. **Combatting AI-Generated CSAM.** Disponível em: <https://www.wilsoncenter.org/article/combatting-ai-generated-csam>.

6. 404 Media. (23 dez. 2023). **Largest Dataset Powering AI Images Removed After Discovery of Child Sexual Abuse Material.** Disponível em: https://www.404media.co/laion-datasets-removed-stanford-csam-child-abuse/

7. SEKHARI, A. et al. **Remember What You Want to Forget: Algorithms for Machine Unlearning.** Disponível em: <https://papers.nips.cc/paper/2021/file/9627c45df543c816a3ddf2d8ea686a99-Paper.pdf>.

8. Yuan, H. et al. (2024). **Towards Robust Knowledge Unlearning: An Adversarial Framework for Assessing and Improving Unlearning Robustness in Large Language Models.** arXiv:2408.10682.

9. **TIC Kids Online Brasil: 65% das crianças e dos adolescentes usam IA generativa para estudar, criar conteúdo e lidar com emoções.** Disponível em: <https://cetic.br/pt/noticia/tic-kids-online-brasil-65-das-criancas-e-dos-adolescentes-usam-ia-generativa-para-estudar-criar-conteudo-e-lidar-com-emocoes/>.

10. KOPECKÝ, Kamil; VORÁČ, Dominik. **The phenomenon of deep nudes—a new threat to children and adults.** AI & Society, v. 41, n. 1, p. 545–556, 2026.

11. FLYNN, Asher et al. **Sexualized deepfake abuse: Perpetrator and victim perspectives on the motivations and forms of non-consensually created and shared sexualized deepfake imagery.** Journal of Interpersonal Violence, n. 08862605251368834, p. 8862605251368834, 2025.

12. SZPACENKOPF, Marta. **Desertos de notícias no Brasil reduzem 7,7% em dois anos, impulsionados pelo crescimento digital.** Disponível em: <https://latamjournalismreview.org/pt-br/articles/desertos-de-noticias-no-brasil-reduzem-77-em-dois-anos-impulsionados-pelo-crescimento-digital/>.

13. **Nota metodológica:** A maioria dos respondentes era composta por meninos cisgênero, totalizando 513 pessoas, ou 73,9% da amostra. A SaferNet Brasil avalia que esse índice pode estar ligado à forma de divulgação da pesquisa, que utilizou anúncios no Instagram. O algoritmo da plataforma tende a priorizar a entrega de conteúdo com base no perfil de engajamento, o que pode ter gerado uma maior exposição do anúncio ao público masculino.

14. NÚCLEO JORNALISMO. **Telegram diz priorizar segurança, mas *bots* contam outra história.** Disponível em: <https://nucleo.jor.br/especiais/2025-02-07-telegram-diz-priorizar-seguranca-de-criancas-mas-seus-robos-contam-outra-historia/>.

15. DE LIRA, Camila. **Ilegal, pornografia por IA é negócio dentro do Telegram.** Disponível em: <https://fastcompanybrasil.com/tech/ilegal-pornografia-por-ia-e-negocio-dentro-do-telegram/>.

16. UNICEF; ECPAT INTERNATIONAL; INTERPOL. **Deepfake abuse is abuse.** Nova York: UNICEF, 4 fev. 2026. Disponível em: https://www.unicef.org/press-releases/deepfake-abuse-is-abuse.

17. SAFERNET BRASIL, Pulitzer Center. **Guia Literacia em IA.** Disponível em: <https://arquivos.safernet.org.br/pdfs/guialetramentoia.pdf>

18. FLYNN, Asher; POWELL, Anastasia; EATON, Asia; SCOTT, Adrian J. **Sexualized deepfake abuse: perpetrator and victim perspectives on the motivations and forms of non-consensually created and shared sexualized deepfake imagery.** Journal of Interpersonal Violence, v. 40, 9 set. 2025. DOI: 10.1177/08862605251368834. Disponível em: https://journals.sagepub.com/doi/10.1177/08862605251368834.

19. UMBACH, Rebecca; HENRY, Nicola. **Perpetration of image-based sexual abuse in the digital age: prevalence, motivations, and community attitudes in 10 countries.** Journal of Cybersecurity, v. 11, n. 1, p. tyaf033, jan. 2025. DOI: 10.1093/cybsec/tyaf033. Disponível em: https://academic.oup.com/cybersecurity/article/11/1/tyaf033/8322282.

20. HAN, C. et al. **Characterizing the MrDeepFakes Sexual Deepfake Marketplace.** Disponível em: <https://arxiv.org/abs/2410.11100>.

21. THORN. **Sexual extortion & young people: navigating threats in digital environments.** Thorn, nov. 2025. Disponível em: https://www.thorn.org/blog/the-state-of-sextortion-in-2025/.

22. UPI. **Some 60% of deepfake sex crime victims in South Korea are minors, police say.** Seoul: UPI, 30 ago. 2024. Disponível em: https://www.upi.com/Top_News/World-News/2024/08/30/South-Korea-deepfake-sex-crime-digital-pornography-scandal/4571725004448/.

23. FLYNN, Asher et al. **Sexualized deepfake abuse: Perpetrator and victim perspectives on the motivations and forms of non-consensually created and shared sexualized deepfake imagery.** Journal of Interpersonal Violence, n. 08862605251368834, p. 8862605251368834, 2025.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

24. CIARDHA, Caoilte Ó.; BUCKLEY, John; PORTNOFF, Rebecca S. **AI generated child sexual abuse material — what's the harm?** 2025. Disponível em: <http://arxiv.org/abs/2510.02978>.

25. GROSSMAN, Shelby et al. **AI-generated child sexual abuse material: insights from educators, platforms, law enforcement, legislators, and victims.** Stanford: Stanford Internet Observatory, 2025. Disponível em: <https://stacks.stanford.edu/file/mn692xc5736/AI-CSAM-paper-2025-05-29.pdf>.

26. INHOPE. **No To Nudify.** https://inhope.org/articles/unifying-voices-worldwide-no-to-nudify. Disponível em: <https://inhope.org/articles/unifying-voices-worldwide-no-to-nudify>.

27. G1. **Adolescentes processam empresa de Musk nos EUA por imagens sexualizadas geradas por *chatbot*.** Disponível em: <https://g1.globo.com/tecnologia/noticia/2026/03/17/adolescentes-processam-empresa-de-musk-nos-eua-por-imagens-sexualizadas-geradas-por-chatbot.ghtml>.

28. NEW YORK TIMES. **Musk's Chatbot Flooded X With Millions of Sexualized Images in Days, New Estimates Show.** Disponível em: <https://www.nytimes.com/2026/01/22/technology/grok-x-ai-elon-musk-deepfakes.html>.

29. MAIA, Mariana Rosetti. **"Tire a roupa dela": o ecossistema de violência dos deepfakes.** Disponível em: <https://nucleo.jor.br/reportagem/2026-01-22-ecossistema-violencia-deepfakes/>.

30. THOMPSON, Polly. **One year after his arrest, Pavel Durov says he'd "rather die" than give third parties access to messages on Telegram.** Business Insider. Disponível em: <https://www.businessinsider.com/pavel-durov-telegram-ceo-user-data-privacy-arrest-2025-8>.

31. SCHWIETER, Christian et al. **Online Safety Regulation & Private Online Communications.** Disponível em: <https://www.isdglobal.org/wp-content/uploads/2025/02/Online-Safety-Regulation-and-Private-Online-Communications-State-of-Play.pdf>.

32. *Exchanges* são plataformas digitais criadas para intermediar a negociação de ativos digitais, principalmente os criptoativos, como o bitcoin. Elas funcionam de forma semelhante às corretoras de valores. No caso da Fragment, seus termos de uso (https://fragment.com/about, versão de 6 de maio de 2026) dão a entender que a plataforma somente permite converter Stars em TON, criptomoeda oficial da Ton Network.

33. SHIEBER, J. **Telegram to pay SEC fine of $18.5 million and return $1.2 billion to investors as it dissolves TON | TechCrunch.** Disponível em: <https://techcrunch.com/2020/06/26/telegram-to-pay-sec-fine-of-18-5-million-and-return-1-2-billion-to-investors-as-it-dissolves-ton/>.

34. https://x.com/durov/status/2051315383466897663

35. Telegram Stars in TON Wallet. Disponível em: <https://help.wallet.tg/article/632-telegram-stars-in-ton-space>.

36. DESMARAIS, Anna. **Telegram hosts vast organised abuse networks in Spain and Italy, report finds.** Disponível em: <https://www.euronews.com/next/2026/04/08/telegram-hosts-vast-organised-abuse-networks-in-spain-and-italy-report-finds>.

37. INTERNET WATCH FOUNDATION. **AI CSAM Report 2026: Harm without limits.** Disponível em: <https://www.iwf.org.uk/about-us/why-we-exist/our-research/how-ai-is-being-abused-to-create-child-sexual-abuse-imagery/>.

38. **Nota metodológica:** Para isto, utilizou-se a fórmula `=SEERRO(VALOR(REGEXEXTRACT(coluna_com_prompt; "\d+")); "")`, que extrai apenas o primeiro valor numérico presente no campo analisado e ignora eventuais números subsequentes. Dessa forma, nem todas as informações quantitativas eventualmente contidas nos prompts são integralmente capturadas (como mais de um número ou números em extenso), o que pode gerar subestimações pontuais nos resultados obtidos.

39. FIGHT THE NEW DRUG. **AI accelerates rise of deepfakes on mainstream porn sites.** Disponível em: <https://fightthenewdrug.org/ai-accelerates-rise-of-deepfakes/>.

40. *[detalhes disponíveis às autoridades — A-09]*

41. Publicação em rede social, nov. 2025. *[detalhes disponíveis às autoridades — A-09]*

42. YALE LAW SCHOOL. **Clinics file suit against website that generates nonconsensual nude images.** Disponível em: <https://law.yale.edu/yls-today/news/clinics-file-suit-against-website-generates-nonconsensual-nude-images>.

43. *[detalhes disponíveis às autoridades — A-14]*

44. *[detalhes disponíveis às autoridades — A-12]*

45. ONU NEWS. Unicef alerta para aumento de imagens sexualizadas de crianças geradas por IA. Disponível em: <https://news.un.org/pt/story/2026/02/1852288>.

46. THORN. **Thorn and All Tech Is Human Forge Generative AI Principles with AI Leaders to Enact Strong Child Safety Commitments.** Disponível em: <https://www.thorn.org/press-releases/thorn-atih-generative-ai-principles>.

47. A Coalition for Content Provenance and Authenticity (C2PA) é um padrão técnico que permite registrar e verificar a origem e o histórico de alterações de um conteúdo digital, garantindo sua autenticidade e rastreabilidade.
PORTNOFF, Rebecca et al. **Safety by design: Annual progress report.** Disponível em: <https://info.thorn.org/hubfs/Thorn_SafetyByDesign_AnnualProgressReport_April2024-April2025.pdf>.

48. ALZAHRANI, Ahmed. **PromptGuard a structured framework for injection resilient language models.** Scientific Reports, v. 16, n. 1, p. 1277, 2026. Disponível em <https://www.nature.com/articles/s41598-025-31086-y/>.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

49. WILSON CENTER. **Combatting AI-generated CSAM.** Disponível em: <https://www.wilsoncenter.org/article/combatting-ai-generated-csam>.

50. AI-powered anti-CSAM filter for stable diffusion. Disponível em: <https://dbzer0.com/blog/ai-powered-anti-csam-filter-for-stable-diffusion/>.

51. THORN. **Thorn's Safety by Design for Generative AI: Progress Reports.** Disponível em: <https://www.thorn.org/blog/thorns-safety-by-design-for-generative-ai-progress-reports/>.

52. CAMERA FORENSICS. **A guide to AI-generated CSAM for investigators of online exploitation.** Disponível em: <https://www.cameraforensics.com/blog/2024/11/18/a-guide-to-ai-generated-csam-for-investigators-of-online-exploitation/>.

53. MITHANI, Jasmine. **How congress could stifle the onslaught of AI-generated child sexual abuse material.** Disponível em: <https://techpolicy.press/how-congress-could-stifle-the-onslaught-of-aigenerated-child-sexual-abuse-material>.

54. AMAZON WEB SERVICES. **Safe image generation and diffusion models with Amazon AI content moderation services.** Disponível em: <https://aws.amazon.com/de/blogs/machine-learning/safe-image-generation-and-diffusion-models-with-amazon-ai-content-moderation-services/>.

55. GOVERNO FEDERAL. DECRETO No 12.976, DE 20 DE MAIO DE 2026. Disponível em: <https://www.in.gov.br/web/dou/-/decreto-n-12.976-de-20-de-maio-de-2026-706921216>.

56. AGÊNCIA NACIONAL DE PROTEÇÃO DE DADOS. Nota Técnica n.º 1/2026/FIS/CGF/ANPD. Disponível em: <https://www.gov.br/anpd/pt-br/centrais-de-conteudo/documentos-tecnicos-orientativos/nota-tecnica-no-1-2026-fis-cgf-anpd.pdf/@@display-file/file>.

57. AGÊNCIA BRASIL. **Violações no Grok continuam, dizem MPF, ANPD e Senacon.** Disponível em: <https://agenciabrasil.ebc.com.br/geral/noticia/2026-02/violacoes-no-grok-continuam-dizem-mpf-anpd-e-senacon>.

58. PF deflagra operação contra abuso sexual infantojuvenil no Rio de Janeiro. Disponível em: <https://www.gov.br/pf/pt-br/assuntos/noticias/2025/10/pf-deflagra-a-operacao-miopia-em-botafogo>.

59. BRASIL, SBT. **Homem é preso em SP suspeito de usar IA para criar material ou conteúdo de violência sexual contra crianças e adolescentes.** Disponível em: <https://sbtnews.sbt.com.br/noticia/policia/homem-e-preso-em-sp-suspeito-de-usar-ia-para-criar-pornografia-infantil>.

60. **Mato Grosso sanciona lei que proíbe uso de inteligência artificial para criação de "deep nudes".** Disponível em: <https://www.al.mt.gov.br/midia/texto/mato-grosso-sanciona-lei-que-proibe-uso-de-inteligencia-artificial-para-criacao-de-deep-nudes/visualizar>.

61. **Suplantar identidad con IA ya es delito en Colombia: así se sancionará.** Disponível em: <https://www.ucentral.edu.co/noticentral/suplantar-identidad-ia-ya-es-delito-colombia-asi-se-sancionara>.

62. PROTECCIÓN A MENORES DE EDAD EN REDES SOCIALES - Cámara de representantes. Cámara de representantes, 11 fev. 2025. Disponível em: <https://www.camara.gov.co/proteccion-a-menores-de-edad-en-redes-sociales-529/>.

63. **Chile aposta em projeto de IA local para romper com preconceito.** Disponível em: <https://veja.abril.com.br/tecnologia/chile-aposta-em-projeto-de-inteligencia-artificial-local-para-romper-com-preconceito/>.

64. **Senado do México publica iniciativa de lei para regular a IA.** Disponível em: <https://www.mobiletime.com.br/noticias/06/10/2025/senado-mexico-regular-ia/>.

65. **IA: Escolas americanas enfrentam epidemia de deepfakes sexuais entre alunos.** Disponível em: <https://tmc.com.br/mundo/ia-escolas-americanas-enfrentam-epidemia-de-deepfakes-sexuais-entre-alunos/>.

66. TEDESCHI, Vitória. **EUA aprovam lei dura contra deepfakes e pornografia de vingança; veja detalhes.** Disponível em: <https://fastcompanybrasil.com/news/eua-aprovam-lei-dura-contra-deepfakes-e-pornografia-de-vinganca-veja-detalhes/>.

67. We are not ready for deepfakes: How Canadian privacy torts do not adequately address the unique harm posed by targeted deepfakes. Disponível em: <https://cba.org/fr-ca/sections/privacy-and-access/resources/we-are-not-ready-for-deepfakes-how-canadian-privacy-torts-do-not-adequately-address-the-unique-harm/>.

68. MARÍN, José Augusto. **Canadá endurece su Código Penal para proteger a víctimas y niños de depredadores.** Disponível em: <https://cbnnoticias.com/canada-endurece-su-codigo-penal-para-proteger-a-victimas-y-ninos-de-depredadores/>.

69. SPAGNUOLO, Sérgio. **O campo de batalha da IA.** Disponível em: <https://nucleo.jor.br/interativos/2025-03-25-campo-batalha-inteligencia-artificial/>.

70. DEPARTMENT FOR SCIENCE; TECHNOLOGY. **New law to tackle AI child abuse images at source as reports more than double.** Disponível em: <https://www.gov.uk/government/news/new-law-to-tackle-ai-child-abuse-images-at-source-as-reports-more-than-double>.

71. **Crime and Policing Bill: child sexual abuse material factsheet.** Disponível em: <https://www.gov.uk/government/publications/crime-and-policing-bill-2025-factsheets/crime-and-policing-bill-child-sexual-abuse-material-factsheet>.

72. CNN BRASIL. **Coreia do Sul criminaliza assistir a deepfakes sexualmente explícitas.** Disponível em: <https://www.cnnbrasil.com.br/internacional/coreia-do-sul-criminaliza-assistir-a-deepfakes-sexualmente-explicitas/>.

73. O CAFEZINHO. **China estabelece regras obrigatórias para identificação de conteúdo gerado por IA.** Disponível em: <https://www.ocafezinho.com/2025/03/15/china-estabelece-regras-obrigatorias-para-identificacao-de-conteudo-gerado-por-ia/>.

ÍNDICE

1. DEFINIÇÕES & TIPOLOGIAS

2. VIOLÊNCIA SEXUAL MEDIADA POR IA EM ESCOLAS BRASILEIRAS

3. PONTOS DE VISTA DE DIFERENTES ATORES ENVOLVIDOS NO FENÔMENO

4. ECOSSISTEMA DIGITAL QUE FACILITA A VIOLÊNCIA SEXUAL MEDIADA POR IA

5. RESPOSTAS DE DIFERENTES ATORES AO PROBLEMA DA VIOLÊNCIA SEXUAL MEDIADA POR IA

6. RECOMENDAÇÕES DETALHADAS PARA MÚLTIPLOS ATORES

7. CONSIDERAÇÕES FINAIS

SOBRE ESTA PESQUISA

74. PARLIAMENT HOUSE. **Criminal code amendment (Deepfake Sexual Material) Bill 2024.** Disponível em: <https://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/bd/bd2324a/24bd081>.

75. REPLICATION-RECEIVER. **NSW Government strengthens protections against deepfakes and image-based abuse.** Disponível em: <https://dcj.nsw.gov.au/news-and-media/media-releases/2025/nsw-government-strengthens-protections-against-deepfakes-and-ima.html>.

76. THE UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG. **Deepfakes and south African law.** Disponível em: <https://www.wits.ac.za/news/latest-news/opinion/2025/2025-09/deepfakes-and-south-african-law.html>.

77. CONCEI, Nota. **DCA 2023: OS DIREITOS DA CRIANÇA NO AMBIENTE DIGITAL.** Disponível em: <https://www.acerwc.africa/sites/default/files/2023-02/DAC%20CONCEPT%20NOTE%202023_PT.pdf>.


Safer net
Safe Online
Tech Coalition