



Nota Técnica 02/2025

Em 2025, cerca de 64% das denúncias recebidas pela SaferNet Brasil — quase sete em cada dez — envolveram conteúdos digitais de abuso e exploração sexual de crianças e adolescentes. Neste documento, além de registrar o aumento nas denúncias, a organização destaca pesquisa em andamento que evidencia como novas tecnologias, incluindo inteligência artificial, têm sido utilizadas para criar e disseminar esse tipo de conteúdo criminoso.

19 de agosto de 2025
Salvador, Bahia

Sumário

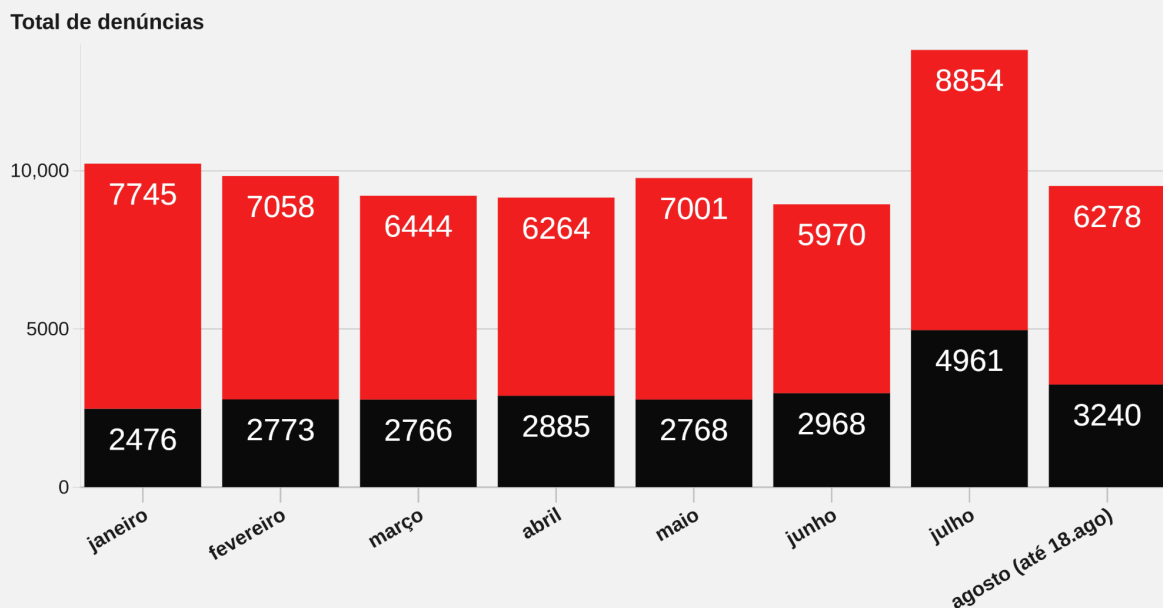
Introdução.....	2
Aumento de denúncias após vídeo-viral sobre exploração sexual-comercial de crianças e adolescentes.....	3
A reação no Congresso Nacional.....	4
Como funcionaria a moderação de conteúdos com o PL 2.628/2022?.....	7
Por que a UE adota um modelo diferente do proposto no PL 2.628/2022?.....	9
Uso da inteligência artificial generativa no abuso infantil digital....	11
O que são esses conteúdos?.....	11
As dimensões dos danos.....	12
Impactos no Brasil.....	13
Exemplos analisados pela SaferNet Brasil.....	14
Pesquisa inédita em andamento.....	16
Chamada pública para a pesquisa.....	17
Sobre a SaferNet Brasil.....	18
Imprensa.....	18
Expediente.....	19

Introdução

Entre 1º de janeiro e 31 de julho de 2025, o Canal Nacional de Denúncias de Crimes Cibernéticos registrou 49.336 denúncias anônimas de abuso e exploração sexual infantil, representando 64% do total de 76.997 notificações no período. Esse número equivale a quase sete em cada dez denúncias recebidas, considerando que o canal também recebe relatos de outros crimes digitais, como racismo e violência contra a mulher.

Quase 7 em cada 10 denúncias à SaferNet Brasil em 2025 são de abuso e exploração sexual infantil

■ Outras denúncias ■ Material de abuso sexual infantil



Dados: Central de Denúncias de Crimes Cibernéticos da SaferNet Brasil

Nota: Outras denúncias incluem crimes cibernéticos como racismo, intolerância religiosa, maus tratos contra animais ou xenofobia.

Em agosto de 2025, o canal de denúncias registrou um pico de denúncias coincidente com a viralização do vídeo “Adultização”, no qual o influenciador digital e humorista Felca expôs como alguns criadores de conteúdo exploram vídeos com teor sexual envolvendo crianças e adolescentes para obter lucro nas redes sociais.

Para Thiago Tavares, presidente da SaferNet, o aumento de denúncias em agosto reflete diretamente o impacto da repercussão do vídeo, que gerou amplo debate na sociedade e estimulou mais pessoas a reportarem conteúdos de abuso.

Além do crescimento de denúncias envolvendo vítimas reais, a SaferNet observa também o aumento de conteúdos sintéticos de abuso sexual infantil. Esses materiais incluem representações de crianças reais e figuras totalmente artificiais, ampliando os desafios de prevenção e combate a esse crime.

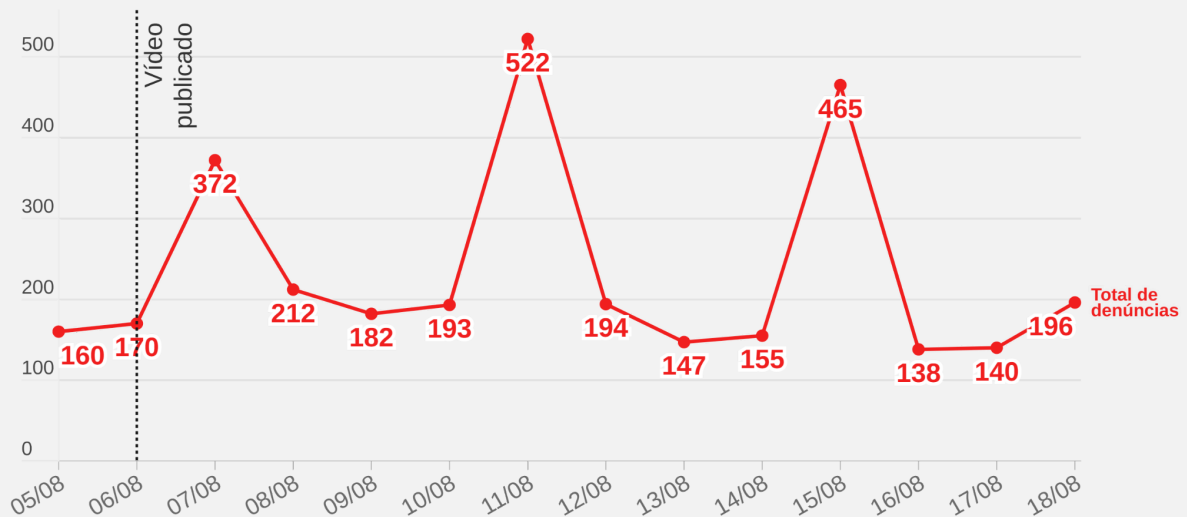
Aumento de denúncias após vídeo-viral sobre exploração sexual-comercial de crianças e adolescentes

Em 6 de agosto, o influenciador Felipe Bressanim, conhecido como Felca, publicou o vídeo **“Adultização”**. Nele, ele discute o papel de influenciadores que, direta ou indiretamente, divulgaram conteúdos ou estímulos relacionados à erotização, sexualização, exploração sexual e adultização de crianças e adolescentes — conceito que se refere à atribuição precoce de comportamentos ou estéticas típicas do universo adulto a crianças e adolescentes, seja por meio da incorporação de traços culturais ou de condutas próprias de adultos. O vídeo alcançou métricas recordes no Brasil, tornando-se um dos conteúdos mais assistidos nas redes sociais. Até 19 de agosto, somava 46 milhões de visualizações e mais de 270 mil comentários.

Entre 1º e 18 de agosto, período de coleta de dados para este relatório, a SaferNet Brasil recebeu 6.278 denúncias anônimas sobre conteúdos digitais envolvendo suspeitas de abuso ou exploração sexual de crianças e adolescentes. **Pouco mais da metade (52%) dessas notificações ocorreu após a divulgação do vídeo “Adultização”**. Desde a publicação do conteúdo, o canal de denúncias registrou 3.246 novas ocorrências, com picos diários de até 500 denúncias em alguns dias.

Denúncias explodem desde o vídeo “Adultização”

Mais da metade das denúncias de recebidas pela SaferNet em agosto (52%) ocorreu após o vídeo-viral, somando 3.246 registros.



A reação no Congresso Nacional

Embora não realize ações de *advocacy* junto ao Poder Legislativo, limitando-se a contribuir tecnicamente em audiências públicas, CPIs (Comissões Parlamentares de Inquérito) e Grupos de Trabalho, quando formalmente convidada, a SaferNet Brasil tem acompanhado pela imprensa e por grupos da sociedade civil a reação dos Deputados e Senadores ao vídeo-viral do Felca. Nos primeiros dias foram apresentados dezenas de [novos projetos de lei](#)¹ para abordar temas como exploração sexual e comercial infantil, além da regulamentação do uso de redes sociais por crianças e adolescentes.

O principal projeto de lei em debate no momento de construção deste documento é o PL 2.628/2022, conhecido como Lei de Proteção de Crianças e

¹ Levantamento feito pelo veículo Núcleo Jornalismo ([nucleo.jor.br](https://docs.google.com/spreadsheets/d/1pb31JC8IV9r9qa39tM3avdMuXXnuM4gn9VVDh6ux5fo/edi?t?ref=nucleo.jor.br&gid=1030177449#gid=1030177449)), acesso em <https://docs.google.com/spreadsheets/d/1pb31JC8IV9r9qa39tM3avdMuXXnuM4gn9VVDh6ux5fo/edi?t?ref=nucleo.jor.br&gid=1030177449#gid=1030177449>

Adolescentes no Ambiente Digital². A proposta estabelece regras para produtos e serviços de tecnologia da informação voltados ou acessíveis a crianças e adolescentes, impondo aos provedores e plataformas digitais o dever de adotar medidas de prevenção, de proteção, de informação e de segurança, sempre observando o melhor interesse da criança e do adolescente, e sua proteção integral, especial e prioritária.

O texto prevê a prevenção e o bloqueio de conteúdos relacionados à exploração e ao abuso sexual infantil, à violência física, ao bullying virtual, ao assédio, ao incentivo a comportamentos suicidas ou autodestrutivos, à promoção de jogos de azar, de produtos nocivos como álcool, tabaco e drogas, e a práticas publicitárias predatórias.

Em sua versão original aprovada no Senado, o texto determina que as configurações padrão de privacidade e segurança sejam as mais protetivas possíveis, restringindo a coleta e o uso de dados pessoais, proibindo o perfilamento para fins publicitários e limitando sistemas de recomendação e geolocalização. O projeto também obriga a disponibilização de ferramentas de controle parental acessíveis e eficazes, capazes de restringir comunicações, controlar o tempo de uso e evitar recursos que incentivem o uso compulsivo.

Além disso, proíbe caixas de recompensa (*loot boxes*) em jogos eletrônicos para crianças, impõe regras de verificação de idade, cria mecanismos para denúncias e remoção de conteúdos violadores sem necessidade de ordem judicial e prevê sanções que vão de advertência a multas e proibição de atividades, destinando os valores arrecadados ao [Fundo Nacional para a Criança e Adolescente](#).

A tabela a seguir compara as recentes obrigações impostas às plataformas digitais a partir do [julgamento](#) do Supremo Tribunal Federal sobre o Art. 19 do Marco Civil da Internet (MCI) — que definiu responsabilidades das plataformas quanto à remoção de conteúdo ilícito mediante notificação — e o previsto no teor do [substitutivo](#) do PL 2.628/2022 apresentado pelo deputado Jadyel Alencar.

² Tramitação do projeto de lei e seu teor original, acesso em <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2477340>

Tema	Rol Taxativo do STF (RE 1.057.258 / Tema 533)	PL 2.628/2022 (art. 6º e correlatos)
Atos antidemocráticos	Condutas previstas nos arts. 286, parágrafo único, 359-L, 359-M, 359-N, 359-P e 359-R do Código Penal.	Não se aplica.
Terrorismo	Crimes de terrorismo ou preparatórios (Lei nº 13.260/2016).	Não se aplica.
Suicídio e automutilação	Induzimento, instigação ou auxílio (art. 122 do Código Penal).	Conteúdos que incentivam comportamentos suicidas e transtornos mentais (art. 6º, III).
Discriminação	Incitação à discriminação por raça, cor, etnia, religião, procedência nacional, sexualidade ou identidade de gênero (arts. 20, 20-A, 20-B, 20-C da Lei 7.716/1989).	Não trata de forma expressa, mas pode estar implicitamente coberto como “conteúdo nocivo” (art. 8º, III).
Violência contra a mulher	Crimes contra a mulher em razão do sexo feminino, inclusive ódio às mulheres (Lei Maria da Penha e Código Penal).	Não se aplica.
Crimes sexuais contra vulneráveis	Pornografia infantil, crimes sexuais contra vulneráveis, crimes graves contra crianças/adolescentes (Código Penal e ECA).	Exploração e abuso sexual de crianças e adolescentes (art. 6º, I; art. 20).
Tráfico de pessoas	Art. 149-A do Código Penal.	Não há menção explícita ao tráfico de pessoas.
Violência física e bullying	Não há previsão expressa.	Violência física, bullying virtual e assédio (art. 6º, II).
Jogos e vícios digitais	Não há previsão.	Padrões de uso que incentivem vício ou transtornos mentais (art. 6º, III).
Produtos nocivos	Não há previsão.	Promoção e venda de jogos de azar, tabaco, álcool, narcóticos (art. 6º, IV).
Publicidade nociva	Não há previsão.	Práticas publicitárias predatórias, injustas ou enganosas (art. 6º, V).

Como funcionaria a moderação de conteúdos com o PL 2.628/2022?

No contexto do PL 2.628/2022, a moderação de conteúdos prevista no Artigo 23 do [substitutivo](#) estabelece que o provedor, diante de uma denúncia, deve decidir entre remover o conteúdo ou mantê-lo. Essa forma de moderação de conteúdo segue o modelo conhecido como notice and take down — em português, notificação e retirada — já consolidado no Brasil e em regiões como a União Europeia, **especialmente em casos de abuso sexual contra crianças e adolescentes**. Nessa prática, as plataformas recebem a notificação, bloqueiam o acesso ao conteúdo e preservam as evidências armazenadas em seus servidores para apoiar investigações, conforme previsto³ na Lei 11.829/08 e no Marco Civil da Internet.

Contudo, ainda que amplamente testado, o modelo continua sendo complexo. Isso ocorre porque envolve uma rede de atores distintos: canais de denúncia, analistas, autoridades policiais, plataformas, serviços de hospedagem com datacenters localizados em outras jurisdições, conselhos de ética e outros órgãos. A ordem dos procedimentos varia de país para país, e a diversidade de sistemas utilizados pelos diferentes participantes cria uma lacuna entre a percepção pública sobre a sua eficácia. Também pesa o fator capacidade, uma vez que nem todas as plataformas ou serviços de hospedagem dispõem da mesma estrutura técnica, humana ou operacional. Essa desigualdade afeta diretamente o tempo de resposta e, em última instância, a velocidade da remoção do conteúdo.

Além do modelo de notificação e retirada (no inglês, *notice and take down*), existem outras abordagens para a moderação de conteúdos online, bem como variações na aplicação do mesmo modelo.

No contexto do [Digital Services Act \(DSA\)](#), o arcabouço jurídico da União Europeia que define regras e obrigações para redes sociais, plataformas digitais e empresas de tecnologia, destacam-se dois termos centrais: *notice and take down*, em que a plataforma deve remover integralmente conteúdos ilegais, e

³ No contexto brasileiro o *notice and take down* está fundamentado no ECA, atualizado pela Lei 11.829/08, que dispõe: art. 241-A §2º (desabilitar acesso após notificação oficial), ECA art. 241-B §2º, III e §3º (posse temporária e sigilo até a entrega à autoridade) e no Marco Civil da Internet art. 15 §2º (preservação cautelar de registros).

*notice and take action*⁴, que permite respostas proporcionais e específicas, podendo ou não incluir a remoção. Esses modelos diferem quanto ao procedimento e ao alcance das medidas que as plataformas podem adotar, oferecendo uma base para respostas⁵ mais flexíveis e graduadas diante de diferentes tipos de infração.

Modelo	Descrição	Ação obrigatória	Exemplo	Medidas possíveis
Notice and Take Down	Modelo tradicional: ao receber notificação sobre conteúdo ilegal, a plataforma deve removê-lo completamente.	Exclusão definitiva ou bloqueio total do conteúdo.	Vídeo contendo abuso sexual infantil é denunciado; a plataforma o retira do ar e preserva evidências para investigação.	Remoção total do conteúdo.
Notice and Take Action	Modelo mais abrangente: ao receber notificação, a plataforma adota uma resposta adequada, proporcional e específica.	A resposta varia conforme o tipo e gravidade da infração; não exige necessariamente remoção total.	Post com informação falsa sobre saúde recebe aviso e redução de alcance, sem remoção total, exceto quando há risco iminente de dano à saúde.	Restrição de acesso por região, desmonetização, retirada de funcionalidades, inclusão de avisos ou rótulos contextuais, limitação de compartilhamento, remoção total (quando necessário).

⁴ Os detalhes sobre a implementação do *notice and take action* previsto no artigo 17(1) (a)-(d) do DSA Europeu estão disponíveis em: <https://transparency.dsa.ec.europa.eu/page/documentation#information-on-the-type-of-restrictions-imposed-on-the-territorial-scope-and-the-duration-of-the-restriction>

⁵ A partir de 1º de julho de 2025, todo serviço online que opere na União Europeia deverá começar a elaborar seu Relatório de Transparência do DSA (TR) utilizando o novo modelo padronizado emitido pela Comissão Europeia. Esse modelo introduz uma taxonomia fixa com 17 categorias de infração e 82 subcategorias. O primeiro relatório anual completo (referente ao ano de 2025) utilizando esse modelo deverá ser entregue até fevereiro de 2026.

Por que a UE adota um modelo diferente do proposto no PL 2.628/2022?

A União Europeia adota o modelo de notificação e ação proporcional por três razões principais. Primeiro, pela flexibilidade, já que nem todo conteúdo nocivo, prejudicial ou indesejável exige remoção completa. Segundo, pela proporcionalidade, ao permitir que a resposta seja ajustada ao tipo e à gravidade da infração. Por último, este modelo, segundo a análise da SaferNet, visa a compatibilidade com os direitos fundamentais, buscando equilibrar a necessidade de segurança com a salvaguarda da liberdade de expressão.

O modelo de notificação e retirada apresenta, no entanto, limitações quando aplicado a conteúdos além daqueles estritamente ilegais. Embora a exclusão rápida de material — como imagens de abuso sexual infantil — ofereça alívio imediato, ela não resolve os fatores que levaram à criação do conteúdo. Problemas sistêmicos, comportamentos ou motivações subjacentes permanecem, tornando provável a recorrência da situação. Por isso, medidas de longo prazo devem abordar as causas, prevenindo novos episódios.

As plataformas podem adotar diferentes estratégias, aplicadas individualmente ou combinadas, com gravidade proporcional à infração ou ao dano causado. Nesse sentido, os remédios alternativos à simples e mera remoção, mediante notificação, de certos conteúdos prejudiciais ou indesejáveis envolve geralmente a aplicação, pelas plataformas, das seguintes medidas⁶:

- **Ações sobre contas:** a) encerrar conta; b) suspender conta; c) impedir novas postagens; d) *shadowban* e/ou *downrank* (medidas para limitar alcance e engajamento de contas); e) remover de buscas.
- **Ações sobre o conteúdo:** a) indisponibilidade; b) remoção das buscas e algoritmos de recomendação; c) redução de alcance/visibilidade; d) exibição apenas para usuários logados; e) desativação de funcionalidades no perfil (ex: comentários, compartilhamentos, etc); e f) inclusão de avisos/etiquetas e notas da comunidade.
- **Consequências monetárias:** a) confiscar os ganhos, suspender e/ou encerrar a monetização; b) multa ao autor (quando previsto em contrato ou

⁶ Para uma análise detalhada e exemplos de aplicação, ver: Goldman, Eric, *Content Moderation Remedies* (2021). *28 Michigan Technology Law Review* 1 (2021), Santa Clara Univ. Legal Studies Research Paper. Disponível em: SSRN: <https://ssrn.com/abstract=3810580> ou <http://dx.doi.org/10.2139/ssrn.3810580>

em decorrência de decisão judicial e/ou administrativa); c) dever de indenizar terceiros prejudicados.

→ **Outras medidas de natureza não punitiva:** a) contra-narrativas; b) e educação midiática; c) Justiça restaurativa e serviço comunitário.

Do ponto de vista normativo, essas medidas refletem alternativas legais e contratuais, mais amplas que a simples remoção de conteúdo, como mostra a tabela a seguir.

Comparativo de medidas de moderação de conteúdo entre notificação e ação

Medida	NTD	DSA	PL 2628/22
Remover conteúdo	✓	✓	✓
Suspender conteúdo ou acesso	✓	✓	✓
Bloqueio por região	✗	✓	✓
Editar ou redigir	✗	✓	✗
Avisos / etiquetas	✗	✓	✓
Desativar comentários	✗	✓	✗
Encerrar conta	✗	✓	✗
Suspender conta ou postagem	✗	✓	✓
Selos / nível de serviço	✗	✓	✗
Shadowban / rebaixamento de alcance	✗	✓	✗
Remover dos resultados de busca	✗	✓	✗
Desativar autocompletar	✗	✓	✗
Bloqueio etário / verificação de idade	✗	✓	✓
Exibir apenas para usuários logados	✗	✓	✗
Monetização: confiscar, encerrar ou suspender	✗	✓	✗
Educação de usuários	✗	✓	✓
Advertência / aviso	✗	✓	✗
Notificar autoridades policiais	✓	✓	✓
Lista de bloqueio (usuário ou conteúdo)	✗	✓	✗

Uso da inteligência artificial generativa no abuso infantil digital

Para além do aumento de denúncias após o vídeo “Adultização”, os dados coletados até o momento indicam um crescimento das ocorrências envolvendo conteúdos de abuso e exploração sexual de crianças e adolescentes gerados por ferramentas de inteligência artificial generativa. Neste ano, a SaferNet Brasil observa um avanço preocupante no uso indevido dessa tecnologia, tanto para produzir conteúdos sintéticos ilegais a partir de imagens reais de crianças — incluindo influenciadores mirins, atrizes e pessoas anônimas — quanto para criar representações totalmente artificiais de crianças e adolescentes.

A inteligência artificial é um tipo de sistema computacional projetado para criar novos conteúdos — como imagens, vídeos, áudios ou textos — a partir de grandes volumes de dados usados em seu treinamento. Ao contrário dos sistemas tradicionais, que apenas classificam ou analisam informações, os modelos generativos produzem material inédito, imitando padrões humanos e sintéticos com notável realismo. Durante o processo de aprendizado, alguns tipos de IA analisam cada ponto de uma imagem — o píxel — e aprendem como esses pontos se organizam: onde há sombra, como se forma o contorno de um rosto, qual é a posição típica dos olhos em relação ao nariz. Ao absorver milhões desses padrões visuais, ela se torna capaz de gerar imagens novas que seguem as mesmas regras — sem copiar nenhuma imagem específica. O resultado é uma criação única, mas visualmente convincente.

Modelos de geração de imagem, como Midjourney (Midjourney Inc.), DALL·E (OpenAI), Flux (Black Forest Labs) e Stable Diffusion (Stability AI), são treinados com extensas bases de imagens associadas a descrições. Ao aprender as relações entre palavras e formas visuais, esses sistemas “traduzem” comandos textuais em imagens realistas. As versões disponibilizadas ao público, conhecidas como modelos-base, já conseguem produzir imagens altamente detalhadas, muitas vezes com aparência quase fotográfica.

O que são esses conteúdos?

A expressão “material de abuso ou exploração sexual infantil” se refere ao termo anteriormente utilizado, “[pornografia infantil](#)”, hoje em desuso na comunidade de

pesquisa e defesa dos direitos humanos. A mudança ocorre para evitar a normalização ou banalização do abuso sexual cometido contra crianças e adolescentes — uma vez que o termo “pornografia” costuma ser associado a conteúdo produzido com consentimento entre adultos. Dessa forma, **no caso de crianças e adolescentes, é necessário pontuar que há sempre uma relação de violência, exploração ou coerção, incompatível com qualquer ideia de consentimento ou sexualidade legítima.**

No caso da inteligência artificial generativa, esse tipo de material criminoso pode envolver fotografias, vídeos, desenhos, animações, textos e áudios que retratem ou simulem atos sexuais com crianças e/ou adolescentes, mesmo que não envolvam vítimas reais (que as imagens não sejam baseadas em uma pessoa natural), como no caso de criações artificiais.

Isso envolve, entre outros exemplos:

- Fotografias reais de crianças ou adolescentes manipuladas digitalmente para aparentarem nudez ou estarem em situações de cunho sexual;
- Geração de imagens hiper-realistas de crianças geradas por inteligência artificial em atos sexuais;
- Animações (*deep fakes*, no inglês) envolvendo vozes e rostos de crianças e adolescentes inseridos em contextos pornográficos;
- Criação de textos eróticos envolvendo personagens infantis com aparência ou idade infantil;
- Treinamento de modelos de IA com bases de dados ilegais contendo material de abuso sexual infantil com contato físico.

As dimensões dos danos

Na avaliação da SaferNet Brasil, é um erro grave subestimar o impacto do material de abuso sexual infantojuvenil gerado por inteligência artificial, tratando-o como uma forma menor de violência. **Os danos psicológicos e secundários causados pelo conteúdo sintético são tão reais e devastadores quanto o do abuso com contato físico.** Para a criança ou adolescente que tem sua dignidade sexual violada — seja por manipulação de uma foto ou pela criação de um deepfake —, o impacto psicológico, a humilhação e o risco de extorsão são igualmente graves.

A facilidade de criação e a velocidade de disseminação proporcionadas pela IA agravam exponencialmente o potencial de dano. Ferramentas que produzem conteúdo inédito e hiper-realista estão cada vez mais acessíveis, permitindo que o abuso seja não apenas replicado, mas também escalado industrialmente. Frequentemente, esse ciclo de violência ocorre no próprio ambiente escolar, onde tanto vítimas quanto autores são adolescentes.

A disseminação desse conteúdo cria um ciclo de revitimização, a imagem ou vídeo continua circulando indefinidamente em redes sociais, aplicativos de mensagem e sites com pouca ou nenhuma moderação, fazendo com que a vítima reviva o trauma a cada nova visualização ou compartilhamento. Além do impacto individual, a proliferação desse material sintético tem um efeito social corrosivo. Essa prática contribui para dessensibilizar a sociedade e normalizar atos que são, em sua essência, uma violência grave e inaceitável.

Impactos no Brasil

O uso da IA para gerar conteúdos de abuso sexual contra crianças e adolescentes já vinha sendo sinalizada no Brasil pela SaferNet desde o ano passado. Em outubro de 2024, um [relatório](#) nosso já destacava o uso crescente das redes sociais e aplicativos de mensagens — mais especificamente o Telegram — no país como ambiente para produção e comercialização de conteúdos ilícitos⁷. Muitas vezes, esses materiais estão lado a lado com evidências de vítimas reais de abuso sexual infantil e imagens sexuais não consentidas de mulheres adultas.

Conforme apontado no documento, usuários aproveitam a plataforma de desenvolvimento de robôs do Telegram para intermediar a criação e venda desses conteúdos criminosos, facilitando o contato direto com compradores no próprio aplicativo, inclusive com transações financeiras mediadas por chaves PIX — o que poderia ser rastreado por autoridades brasileiras e o Banco Central. Um dos canais mencionados como exemplo no relatório vendia mídias artificiais de celebridades brasileiras — maiores de 18 anos e, em alguns casos, jovens

⁷ EM SUAS PRÓPRIAS PALAVRAS: COMO o Telegram tem sido usado no Brasil como um espaço de comércio virtual por criminosos sexuais, acesso em https://arquivos.safernet.org.br/press/Safernet_Telegram.pdf

que acabavam de atingir a maioria — em contextos sexuais por valores entre R\$19,90 e R\$25 naquele momento. Ao momento da produção do relatório, em setembro de 2024, o grupo tinha 46 mil participantes.

Mas o cenário atual vai além dos grupos e canais fechados em aplicativos de mensagens instantâneas. **Atualmente, existem dezenas e centenas de vetores para criação e disseminação desses conteúdos criminosos.** Entre os alvos analisados pela SaferNet estão sites de pornografia e conteúdo adulto com pouca ou nenhuma moderação, aplicativos de “nudificação” — que removem roupas de fotografias sem qualquer verificação de idade — e geradores de imagem baseados em sistemas de código aberto. O problema se torna ainda mais grave no caso dos aplicativos de nudificação, já que a maioria não dispõe de mecanismos eficazes para checar a idade dos usuários ou das pessoas retratadas. Com isso, fotos de crianças e adolescentes podem ser facilmente manipuladas e transformadas em material sexualmente explícito.

Exemplos analisados pela SaferNet Brasil

Um exemplo recente que reacendeu o debate sobre medidas eficazes de coibição foi a ação judicial movida pela Meta, empresa dona de redes sociais como Instagram e Facebook, contra a desenvolvedora do aplicativo CrushAI⁸. A ferramenta oferece recursos de “nudificação” e criação de *deepfakes* sexuais a partir de fotos comuns, permitindo gerar imagens de nudez artificiais sem o consentimento das pessoas retratadas.

A empresa é acusada de ter veiculado anúncios desses aplicativos nas próprias plataformas da Meta, de forma dissimulada. Para evitar os filtros automáticos de detecção de nudez e burlar as regras de publicidade, os responsáveis recorreram a uma estratégia de mascaramento: em vez de exibir diretamente o conteúdo explícito ou sugestivo, utilizavam imagens neutras e aparentemente inofensivas, que não despertavam a atenção dos sistemas de verificação. Esses anúncios funcionavam como uma espécie de isca, direcionando o usuário, após o clique, a páginas externas ou serviços que ofereciam as ferramentas de geração de nudez sintética e *deepfakes* sexuais. Na prática, tratava-se de uma operação planejada para explorar brechas tecnológicas de moderação,

⁸ *Meta sues maker of explicit deepfake app for dodging its rules to advertise AI ‘nudifying’ tech*, CNN, acesso em: <https://edition.cnn.com/2025/06/12/tech/meta-sues-explicit-deepfake-app-crushai>

dificultando a detecção imediata pelas plataformas e ampliando a disseminação dos aplicativos ilegais.

Além desses casos, a SaferNet Brasil identificou outro padrão preocupante: **sites de pornografia e conteúdo adulto que, por falhas ou ausência de mecanismos de moderação eficazes, acabam hospedando material de abuso e exploração sexual de crianças e adolescentes criado por inteligência artificial.**

Entre 2024 e 2025, o canal de denúncias da ONG recebeu 95 registros que continham a palavra-chave “inteligência artificial” nos comentários dos denunciantes — 12 deles relacionados a um único site, cujo nome não é divulgado por razões de segurança. Embora o número não seja absoluto, dada a limitação da pesquisa a uma única palavra-chave, ele funciona como um alerta e é citado apenas para fins didáticos. As tecnologias de IA Generativa permitem produção em escala industrial e gera imagens cada vez mais realistas, dificultando a detecção e a denúncia por parte do público. É provável que esses registros representam apenas a ponta de um iceberg muito maior. Além disso, é importante considerar que esse tipo de crime envolve múltiplas vítimas: em casos já noticiados em escolas, por exemplo, dezenas de adolescentes foram atingidos, e a capacidade produtiva da IA generativa multiplica esse impacto ao criar milhares de imagens em larga escala.

O site citado nesse exemplo, que em tese permite a publicação de vídeos próprios por usuários, vem sendo usado para hospedar produções geradas por IA com cenas de abuso sexual infantil. O mesmo espaço também abriga gravações reais de violência e imagens de nudez não consensual de mulheres adultas — padrão de violência de gênero já documentado pela SaferNet em pesquisas sobre o mercado de crimes sexuais no Telegram.

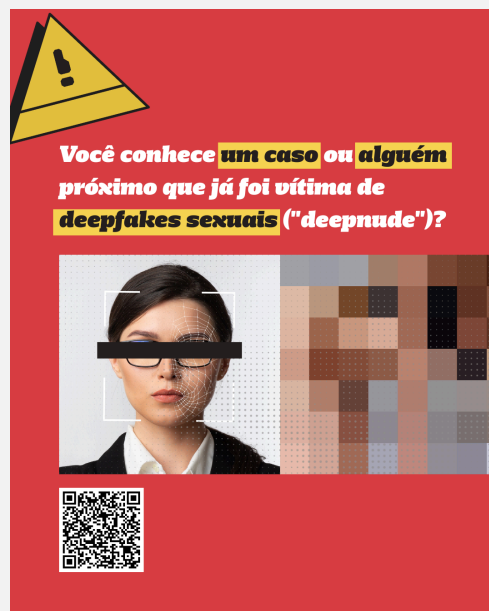
Pesquisa inédita em andamento

Esse é um problema que já tem atingido escolas por todo o país, com vários casos sendo noticiados pela imprensa. Os riscos trazidos pela inteligência artificial generativa para a segurança e proteção de crianças e adolescentes em ambientes digitais geram desafios complexos, principalmente porque tanto as vítimas quanto os autores são adolescentes.

Diante da dificuldade de mensurar com precisão o volume de conteúdo de abuso gerado por inteligência artificial, e considerando que os números disponíveis subestimam a dimensão do problema, a SaferNet Brasil iniciou uma pesquisa inédita no país, com apoio do fundo internacional Safe Online⁹. O estudo busca ir além de dados quantitativos, reunindo evidências qualitativas a partir da perspectiva dos jovens, para oferecer um panorama mais fiel do uso indevido dessa tecnologia.

O objetivo é fundamentar decisões e medidas de proteção de crianças e adolescentes que transcendam o pânico moral e abordagens estritamente punitivas, priorizando ações de conscientização e estratégias eficazes de prevenção desse tipo de crime.

Para Juliana Cunha, diretora da SaferNet e responsável pela pesquisa, trata-se de um fenômeno recente e em rápida evolução — o que abre uma janela de oportunidade para propor respostas mais eficazes, baseadas em evidências. “O resultado desta pesquisa é a proposição de recomendações acionáveis para diferentes atores: empresas de tecnologia, que devem prevenir e mitigar riscos; o poder público, responsável por regular e fiscalizar; e a sociedade civil, que precisa atuar na conscientização e na denúncia.”



⁹ Iniciativa de fundo investimento global focada exclusivamente na segurança infantil no mundo digital. O projeto foi contemplado nesta chamada pública que selecionou quatro propostas de pesquisa: <https://safeonline.global/genai-child-online-safety-shaping-a-safer-digital-future-together/>

Chamada pública para a pesquisa

O fenômeno da criação de imagens não consensuais com inteligência artificial é recente e está em rápida evolução. Isso significa que cada história compartilhada ajuda a construir um panorama real do impacto dessa tecnologia sobre adolescentes.

Ao entrar em contato:

- ★ Você dá voz a quem muitas vezes é invisibilizado, ajudando a mapear a dimensão do problema que ainda não é totalmente conhecida.
- ★ Sua experiência contribui para que a SaferNet Brasil e outros atores — empresas de tecnologia, poder público e sociedade civil — possam desenvolver medidas de prevenção, conscientização e proteção mais eficazes, baseadas em evidências e não apenas em pânico moral.
- ★ Você ajuda a criar recomendações práticas e acionáveis que podem proteger outros adolescentes, influenciar políticas públicas e aprimorar protocolos de denúncia e cuidado.

Cada relato é valioso. É por isso que, **se você é um(a) adolescente ou conhece um(a) adolescente que já foi vítima da criação de imagens não consensuais com inteligência artificial, saiba que a SaferNet Brasil quer ouvir sua história.**

Entre em contato conosco através da central de atendimento e denúncias ou, se preferir, preencha [este](https://forms.gle/G8QUrsHksmvTKCyK6) formulário (acessível também pelo link: <https://forms.gle/G8QUrsHksmvTKCyK6>).

Sobre a SaferNet Brasil

A SaferNet, ONG brasileira referência na promoção dos direitos humanos na internet, completará 20 anos em dezembro de 2025. Ao longo de sua trajetória, sempre adotou uma abordagem multissetorial para enfrentar crimes cibernéticos contra os direitos humanos, acolher vítimas de violência online e desenvolver programas de educação, prevenção e conscientização.

Mantém o Canal Nacional de Denúncias (www.denuncie.org.br), conveniado ao Ministério Público Federal, e o Canal de Ajuda (www.ajuda.org.br), o Helpline, voltado às vítimas de violência e outros problemas no ambiente digital. Também promove o uso seguro da internet por meio de iniciativas como a [disciplina de Cidadania Digital](#).

Reconhecida internacionalmente, a SaferNet opera o [quinto hotline](#) mais atuante do mundo na cooperação com canais de denúncia de outros países para combater a proliferação de imagens de abuso e exploração sexual infantil.

Imprensa

Se você faz parte da imprensa ou da sociedade civil e deseja entrar em contato com a SaferNet sobre este documento, envie um e-mail para o assessor de comunicação Marcelo Oliveira em comunicacao@safernet.org.br.

Expediente

Realização: SaferNet Brasil

Apoio: Fundo internacional Safe Online

Coordenação da pesquisa: Juliana Cunha

Pesquisa e levantamento de dados: Sofia Schurig

Gráficos: Sofia Schurig

Redação e edição: Sofia Schurig e Juliana Cunha

Revisão: Thiago Tavares

Diagramação: Haide Sousa

Imprensa: Marcelo Oliveira

Data de publicação: 20 de Agosto de 2025