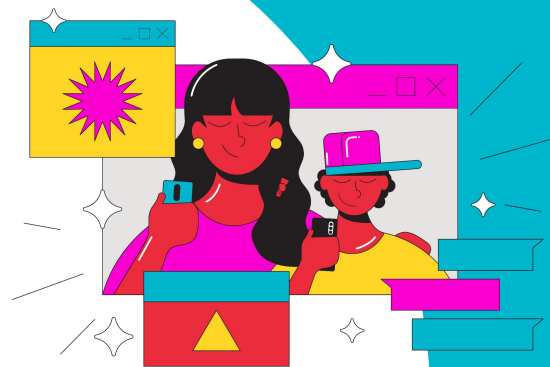


Rostos fabricados, danos reais: como a IA generativa cria conteúdo sexual sintético

Sofia Schurig – Pesquisadora na Safernet
UFBA, EnSI 2026
01/10/2026





NOSSA
PROPOSTA
É EDUCAR
PARA UMA
NAVEGAÇÃO
LIVRE E SEGURA.

É CONSCIENTIZAR
PARA BOAS
ESCOLHAS
ONLINE.

1

Canal de ajuda

helpline

Canal de Ajuda para
vítimas de violência
online

canaldeajuda.org.br

+39,5 mil pessoas
atendidas desde 2007

2

Canais de denúncia

hotline

Central Nacional de
Denúncia de Crimes
Cibernéticos

denuncie.org.br

+4,9 milhões de
denúncias desde 2006

ESCOLA SEGURA

Central de
denúncias Escola
Segura (parceria com
Ministério da Justiça)

gov.br/escolasegura

+12 mil denúncias

3

Educação e prevenção

Cursos e ações de
educação e
conscientização

 cursos online

Disciplina de
Cidadania Digital

cidadaniadigital.org.br

Dia Mundial da
Internet Segura



diadainternetsegura.org.br

+150 mil cursistas EAD

4

Engajamento de jovens

Educação entre
pares,
embaixadores e
mobilizadores

CIDADÃO
DIGITAL



cidadadigital.org.br

+300 mil adolescentes e
jovens engajados

Alguns de nossos parceiros

Secretaria de Estado
da Educação



**GOVERNO
DA PARAÍBA**

Secretaria de
Educação



**GOVERNO DO ESTADO
RIO DE JANEIRO**



GOVERNO DE
**SANTA
CATARINA**
SECRETARIA DA EDUCAÇÃO

MPF
Ministério Público Federal



CONANDA
Conselho Nacional dos Direitos
da Criança e do Adolescente

CONJUVÉ
CONSELHO NACIONAL DA JUVENTUDE

 **UK Government**



MINISTÉRIO DA
EDUCAÇÃO

MINISTÉRIO DOS
DIREITOS HUMANOS
E DA CIDADANIA

SECRETARIA DE
COMUNICAÇÃO SOCIAL

MINISTÉRIO DA
JUSTIÇA E
SEGURANÇA PÚBLICA

GOVERNO FEDERAL
BRASIL
UNIÃO E RECONSTRUÇÃO



Sofia Schurig

Safernet Brasil

Atuação Atual

Pesquisadora na Safernet Brasil desde 2024

Trajetória Profissional

Foi jornalista investigativa na área de tecnologia por quatro anos, cobrindo crimes cibernéticos contra crianças e adolescentes. Foi fellow do Pulitzer Center e repórter no Núcleo Jornalismo

Formação Acadêmica

FACOM, UFBA

Nossa agenda de hoje:



01 **Parte 1:** como a IA generativa cria imagens

02 **Parte 2:** o que de ruim pode acontecer?

03 **Parte 3:** exemplos mapeados

04 **Parte 4:** impacto + perguntas

O que é inteligência artificial generativa?

Inteligência artificial generativa é um subconjunto de sistemas especializados cuja função é criar conteúdo sintético — como textos, imagens, vídeos e áudios — por meio da análise e processamento de grandes volumes de dados.

Modelos de IA generativa — como GPT-4, Llama 3 e Gemini 1.5 (treinados com dezenas de trilhões de *tokens**), Claude 3 (centenas de bilhões a trilhões de *tokens*) e Stable Diffusion (bilhões de pares imagem-texto) — foram desenvolvidos utilizando grandes volumes de dados massivos acumulados até 2024.

**Tokens: o volume de dados de texto usado no treinamento*

Como sistemas de IA aprendem?

01

Consumo

(Dados brutos)

Coleta massiva de textos, fotos e vídeos públicos.

02

Treino

(Análise repetida)

Mapeamento de conexões estatísticas entre palavras e pixels.

03

Ajuste

(Pesos neurais)

Recalibragem de bilhões de números internos a cada erro de previsão.

04

Padrões

(Regularidades)

Internalização de curvas de rostos, iluminação e texturas.

05

Geração

(Respostas prováveis)

Cálculo da matriz de dados mais provável ao receber um prompt.

Nota técnica: A IA não possui consciência ou empatia; ela sintetiza relações matemáticas aprendidas.

O que de ruim pode acontecer e está acontecendo?

“Alimentar os sistemas de IA com a beleza, a feiura e a crueldade do mundo, mas esperar que eles reflitam apenas a beleza é uma fantasia”

Ruha Benjamin, professora na Universidade de Princeton

A origem dos dados e o reflexo do mundo real



01

Bases massivas de raspagem

Ferramentas de IA são treinadas com bilhões de imagens coletadas da internet sem filtro ético rigoroso ou curadoria prévia (ex: banco de dados LAION-5B).

AI Largest Dataset Powering AI Images Removed After Discovery of Child Sexual Abuse Material

SAMANTHA COLE · DEC 29, 2023 AT 7:00 AM

The model is a massive part of the AI-ecosystem, used by Stable Diffusion and other major generative AI products. The removal follows discoveries made by Stanford researchers, who found thousands instances of suspected child sexual abuse material in the dataset.

RELEASING RE-LAION-5B: TRANSPARENT ITERATION ON LAION-5B WITH ADDITIONAL SAFETY FIXES

by: LAION e.V., 30 Aug, 2024

Today, following a [safety revision procedure](#), we announce Re-LAION-5B, an updated version of LAION-5B, that is the first web-scale, text-link to images pair dataset to be thoroughly cleaned of known links to suspected CSAM.

A origem dos dados e o reflexo do mundo real

01

Bases massivas de raspagem

Ferramentas de IA são treinadas com bilhões de imagens coletadas da internet sem filtro ético rigoroso ou curadoria prévia (ex: banco de dados LAION-5B).

02

Espelhando a web

A web já era historicamente hipersexualizada, racista e misógina. Ao "ingerir" esses dados brutos, a IA assimila esses padrões e reproduz essas estruturas.

03

Sistematização de vieses

A tecnologia não é neutra; ela reproduz, em escala industrial, as vulnerabilidades e os preconceitos estruturais embutidos nos seus dados de origem.

O problema do código aberto

Modelos corporativos (fechados)

Ferramentas como DALL-E, Gemini e Adobe possuem *guardrails* (filtros de segurança). Se você pedir para gerar "nudez" ou "violência", o sistema bloqueia o comando.

Modelos de código aberto (pesos abertos)

Arquiteturas como *Stable Diffusion* podem ser baixadas gratuitamente e modificadas pela comunidade.

A raiz do problema

Com modelos abertos, agressores podem desativar completamente os filtros de segurança e rodar a inteligência artificial em servidores próprios ou em robôs / geradores de imagem independentes, ou até mesmo criar produtos dedicados somente à criação de conteúdos criminosos.

Quais os perigos da IA generativa na criação de conteúdo visual?

1. Desinformação e fake news

Criação de imagens e vídeos hiper-realistas para forjar acontecimentos, manipular a opinião pública e disseminar narrativas falsas.

2. Golpes e usurpação de identidade

Uso de simulação de rostos e biometria para realizar fraudes financeiras, criar perfis falsos e enganar familiares ou instituições.

3. Uso não consentido de imagem

Manipulação de fotos reais de pessoas sem autorização para difamação, falsas atribuições ou assédio moral.

4. Amplificação de vieses e preconceitos

Reprodução e automatização de estereótipos de gênero, raça e classe presentes na base de dados dos algoritmos.

Parte 2

Falamos de forma mais profunda de alguns desses riscos nesse material:



Olá!

**Vamos conversar
sobre o futuro?**

Ele já é nosso presente

Criar e educar no mundo de hoje é incrível, mas também assusta, não é? A Inteligência Artificial (IA) chegou mudando tudo: como estudamos, como criamos e até como nos divertimos. Mas, junto com a novidade, vêm as dúvidas e alguns riscos sérios.

Neste território novo e vasto, é normal se sentir um pouco perdido. Por isso, este guia foi criado para ser o seu mapa de navegação.

Mais do que um manual técnico, ele traça as rotas seguras e sinaliza os buracos na estrada (como os riscos e impactos negativos). O objetivo é colocar em suas mãos os recursos necessários para que você saiba exatamente onde está pisando e possa guiar crianças e adolescentes com ética, segurança e sem medo.

Respire fundo, consulte o mapa e conte com a gente nessa jornada!



<https://arquivos.safernet.org.br/pdfs/guiaetramentoia.pdf>

Parte 2

Qual dessas imagens foi gerada por IA?

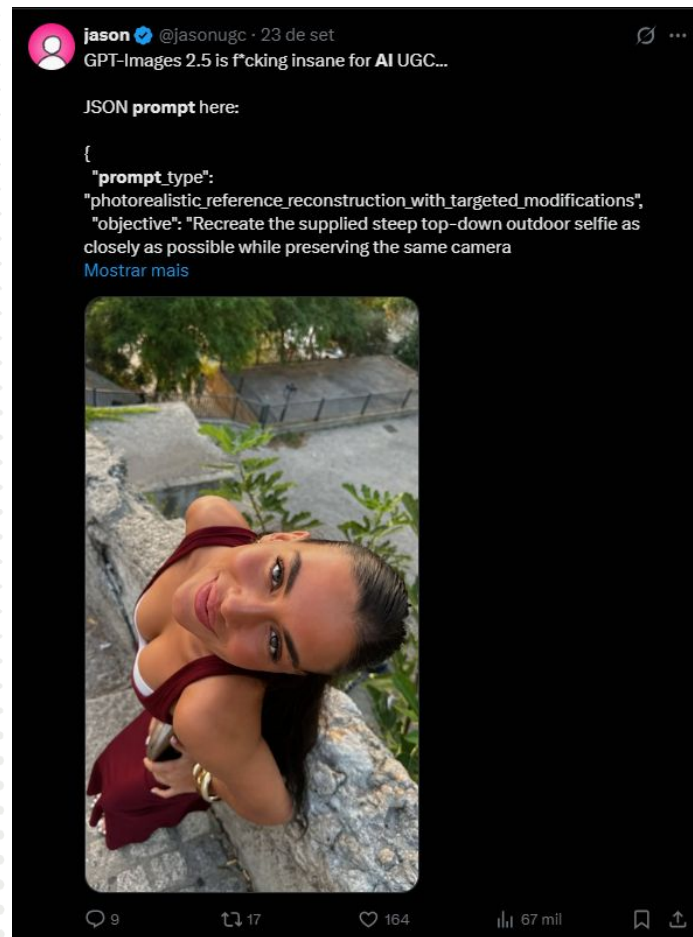
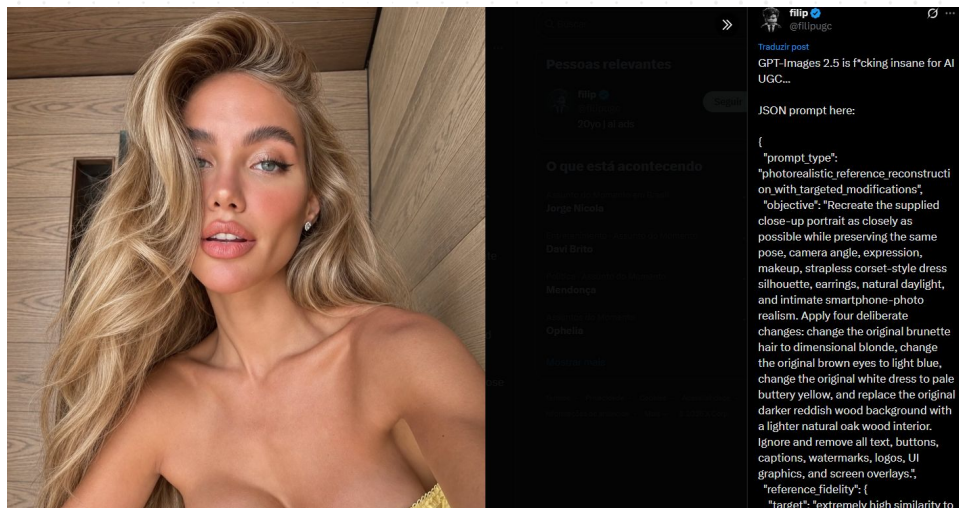
Imagem A:



Imagem B:



Ambas foram geradas com IA!



Qual das imagens — a imagem A, a imagem B ou ambas — foi gerada por inteligência artificial?

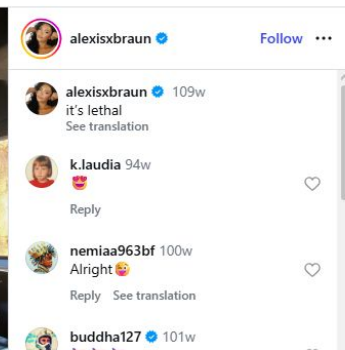
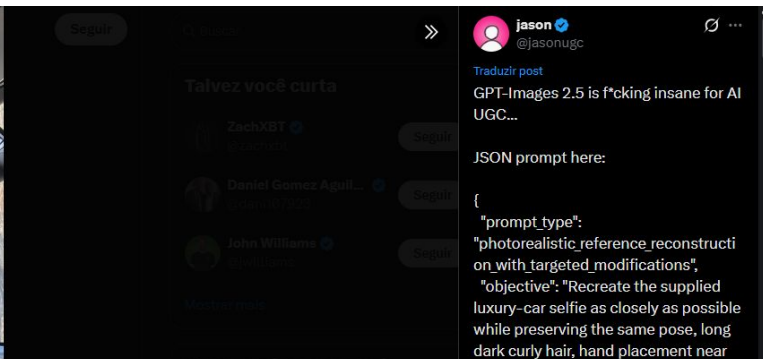
Imagem A:



Imagem B:



A imagem B foi gerada a partir da criadora de conteúdo real da imagem A!



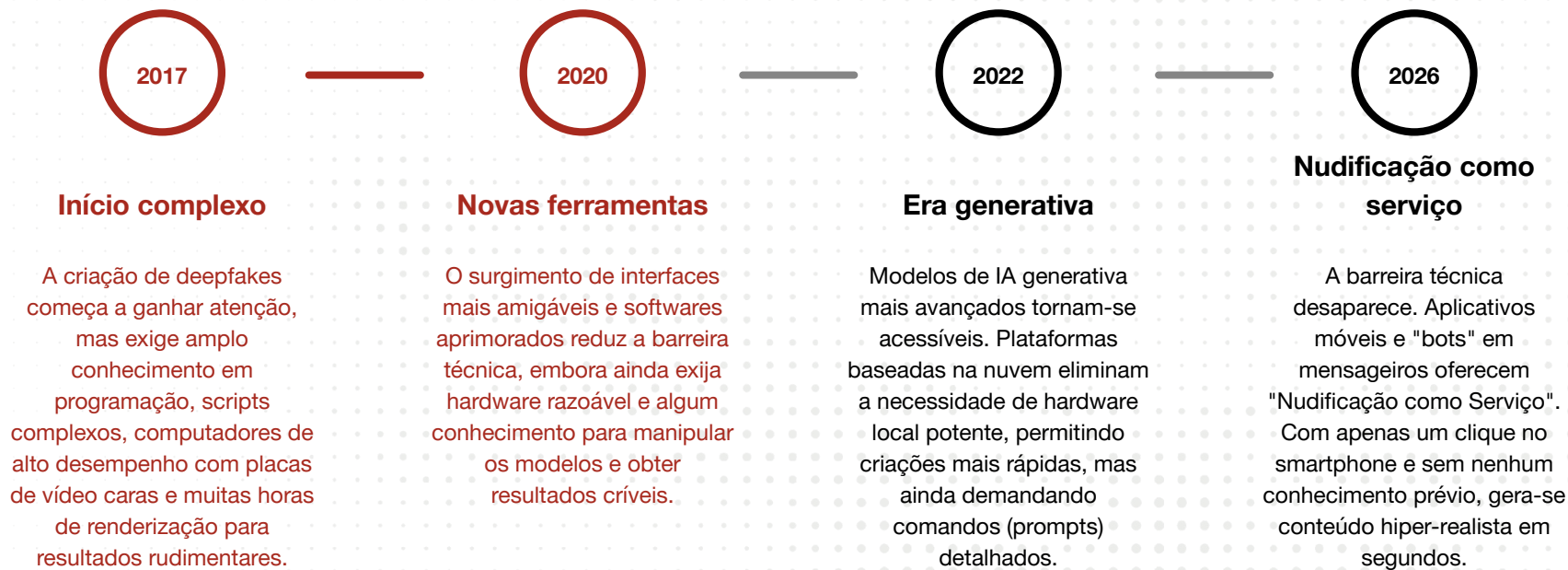
Quais os perigos da IA generativa na criação de conteúdo visual?

Foco da SaferNet — Ponto 3: Uso não consentido de imagem

Como IAs generativas vêm sendo usadas na criação de conteúdo de nudez e violência sexual

Impactando prioritariamente **mulheres, crianças e adolescentes** através da manipulação não autorizada e geração de imagens sintéticas.

A queda da barreira técnica: da complexidade à “Nudificação como serviço”



Quais são os conteúdos que estamos falando?

Conceito 01

Deepfake (Geral)

Síntese ou manipulação digital de áudio, vídeo ou imagem por Inteligência Artificial.

Substituição autêntica de rostos, vozes ou expressões em contextos genéricos. **Exemplo: a imagem da influenciadora que vimos em slides anteriores.**

Conceito 02

Deepfake Sexual

Uso de IA para montagem ou criação de nudez não consentida.

Superposição deliberada do rosto da vítima sobre conteúdos íntimos sem autorização.

Conceito 03

Conteúdo de Violência

Geração sintética de imagens retratando abusos, exploração ou agressões.

Violação direta de leis de proteção humana e políticas globais de uso.

O cenário no Brasil: o que diz a lei?

01

Estatuto da Criança e do Adolescente

Criminaliza a produção e distribuição de material íntimo simulado ou sintético envolvendo menores de 18 anos.

02

Código Penal

Art. 216-B e 218-C: Punição para o compartilhamento não consentido de nudez.

Jurisprudência enquadra manipulação por IA como crimes de difamação e injúria grave.

03

Lei nº 15.022/2026

Tipifica criação e envio de deepfakes íntimos, inclusive para maiores.

Exige marcação clara em conteúdos de IA e impõe multas a plataformas omissas.

Como deepfakes (gerais) são criados?

01

1. Mapeamento

O usuário insere o vídeo original no software de IA, que mapeia a geometria e os movimentos do rosto na cena.

02

2. Análise Facial

O utilizador fornece fotografias de um segundo rosto. A rede neural analisa esta nova identidade.

03

3. Fusão quadro a quadro

A IA sobrepõe o novo rosto aos pontos de ancoragem do vídeo original, calculando a fusão continuamente.

04

4. Ajuste natural

O algoritmo recalcula iluminação, sombras e expressões para criar uma máscara digital perfeita.

Parte 3

Quais exemplos reais temos de que existem danos reais? Exemplo 1: mapeamento de casos em escolas



Ao longo de 7 meses (ago/2025 a fev/2026), o mapeamento da **SaferNet Brasil** identificou graves casos de violência nas escolas.

Vítimas: 100% do gênero feminino (estudantes de 12 a 17 anos e professoras).

Infratores: 100% do gênero masculino na mesma faixa etária.

17

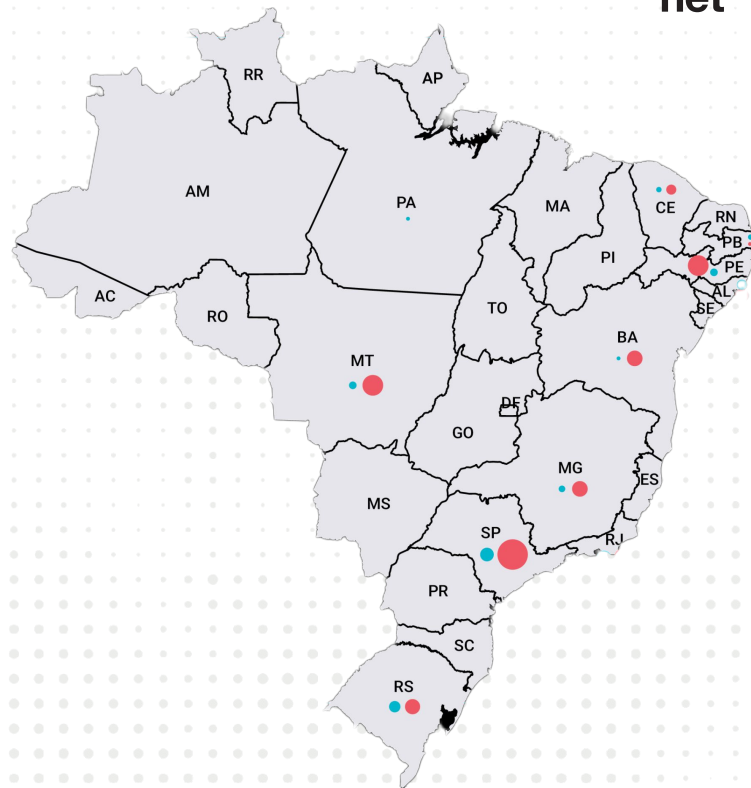
Episódios

61

Infratores

222

Vítimas



Tendências do mapeamento em escolas brasileiras

01

Vitimização feminina

Em nenhum dos 17 casos a vítima era do gênero masculino, mostrando que essa tendência é parte de um problema de violência de gênero estrutural, direcionada sistematicamente contra mulheres e meninas no ambiente escolar.

02

Professoras como vítimas

Casos onde educadoras são expostas por seus alunos corroem a autoridade pedagógica e desestabilizam o ambiente de ensino. A tecnologia é mobilizada deliberadamente como instrumento de humilhação profissional.

03

Comercialização econômica

A venda de imagens manipuladas em canais privados do Telegram ou sites adultos adiciona uma dimensão econômica alarmante ao fenômeno, aproximando a prática de modelos graves de exploração sexual comercial.



Exemplo 2: robôs de nudificação

ESPECIAIS

Telegram diz priorizar segurança de crianças, mas seus robôs contam outra história

// Investigação do Núcleo em parceria com o Pulitzer Center descobriu ao menos 23 bots no Telegram capazes de gerar material explícito de abuso sexual



ONU News
Perspectiva Global Reportagens Humanas

Q. PESQUISAR

Unicef alerta para aumento de imagens sexualizadas de crianças geradas por IA

do pela Folha aponta

4 de dezembro de 2023, 13:02

Telegram permite bots de IA que geram 'deep nudes' e até pornografia infantil

VIOLÊNCIA ONLINE

AGU notifica Google para desindexar sites que criam nudes falsos com IA

30 anos

piuí

questões da internet

A "NUDIFICAÇÃO" DE TODOS NÓS

Se nos anos 2010 as redes sociais tornaram todos alcançáveis, nos anos 2020 a inteligência artificial está tornando abusáveis



TECH

● Ao vivo Política Money WW Agro Infra IA

Apple e Google são obrigadas a derrubar apps de 'nudificação' nos EUA

Procurador de São Francisco quer impedir que empresas lucrem com conteúdo prejudicial ao usuário

O que são robôs de nudificação e como eles funcionam?

O ato de “nudificar” pode ser feito através de produtos personalizados, robôs embutidos em plataformas, aplicativos e websites, mas sempre envolve uma pessoa real como base.

01

Imagem da vítima

Uma foto comum, em que a pessoa aparece vestida, é coletada (normalmente através de perfis em redes sociais) sem consentimento ou conhecimento da vítima, que raramente descobre o uso.



02

Comando do usuário

O agressor envia a imagem e, com um comando simples, pede uma versão sem roupas. Nenhuma habilidade técnica é exigida — a barreira para o abuso é mínima.



03

Foto nudificada

A IA gera um nu falso, mas convincente. O dano é real: humilhação, extorsão sexual e trauma. **Se a vítima for menor de 18 anos, equivale juridicamente a material de exploração ou abuso sexual infantil.**

Parte 3

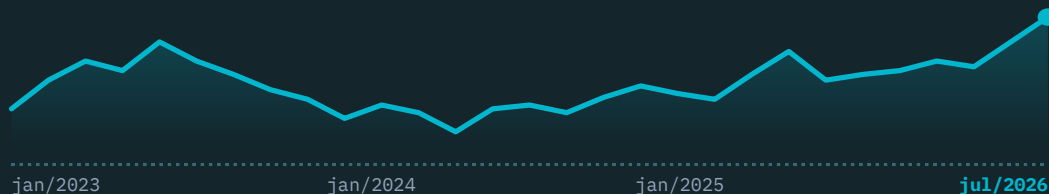


O Telegram como um vetor de risco e espaço para proliferação conteúdo de violência sexual infantil

(a) Evolução mensal das denúncias

por data da denúncia (jan/2023 até jul/2026)

587 denúncias em jul/2026



(b) Registros por mês e ano

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		261	285	221	358	219	197	183	211	2.721
2026	278	273	234	261	265	396	587	—	—	—	—	—	2.294*

■ < 100 ■ 100-200 ■ 200-300 ■ 300-400 ■ > 400 ■ 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.

O Telegram como um vetor de risco e espaço para proliferação conteúdo de violência sexual infantil



Percepção de impunidade

Reputação de resistência a pedidos governamentais de acesso a dados transmite sensação de segurança aos criminosos.



Contas descartáveis

Números de telefone reutilizáveis criam e recriam contas sem registro permanente — rastreabilidade quase impossível.



Escala dos canais

Facilidade para criar comunidades e canais com até **200.000 usuários**.



Acesso em camadas

Canais públicos são vitrines que direcionam o tráfego a **comunidades privadas pagas**, onde a troca de fato ocorre.

Parte 3

E por que o Telegram nos preocupa tanto?

Pagamento facilitado por Stars

Moeda interna do Telegram usada para pagar por bots e conteúdos digitais.

Conversão e saque das Stars

Após 21 dias, podem ser convertidas em Toncoin (criptomoeda da blockchain TON).

Modelo de negócio das Stars

O Telegram retém entre 15% e 30% de cada saque, lucrando indiretamente com as transações.

Impacto e risco do uso das Stars

Cria um circuito financeiro com entrada convencional e saída em ativos de difícil rastreamento.

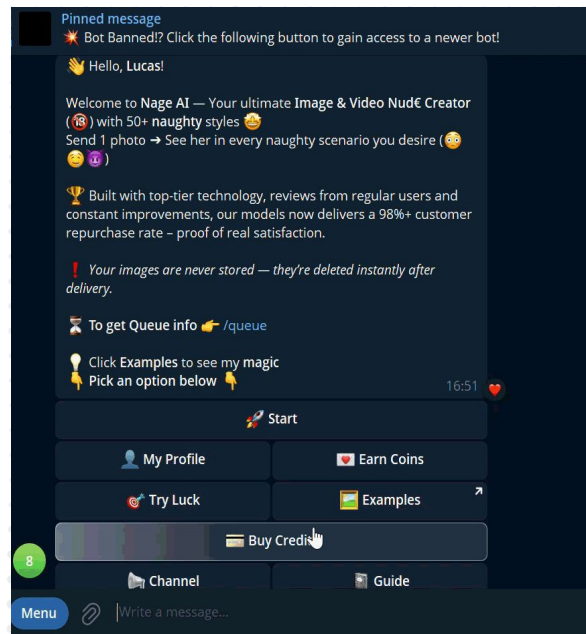


Imagem acima: exemplo de robô de nudificação encontrado pela Safernet durante a análise

Antes de irmos para as perguntas, um toque sobre os problemas desses conteúdos



Perda de controle

A perda de controle sobre a própria identidade gera um sofrimento profundo diante de imagens não autorizadas.



Estigma

O estigma isola a vítima, que frequentemente é forçada a abandonar suas redes sociais e relações próximas.



Morte social

Ruptura drástica na rotina, forçando a mudança de aparência física e até a troca de instituição de ensino.



Saúde mental

Gatilhos para quadros severos de TEPT, depressão profunda, autolesão e ideação suicida recorrente.

E vamos para as perguntas!
Muito obrigada pela sua atenção.

Perguntas, pedidos de materiais e afins também via **sofia@safernet.org.br**.

Precisa de ajuda? canaldeajuda.org.br

Precisa fazer uma denúncia? denuncie.org.br