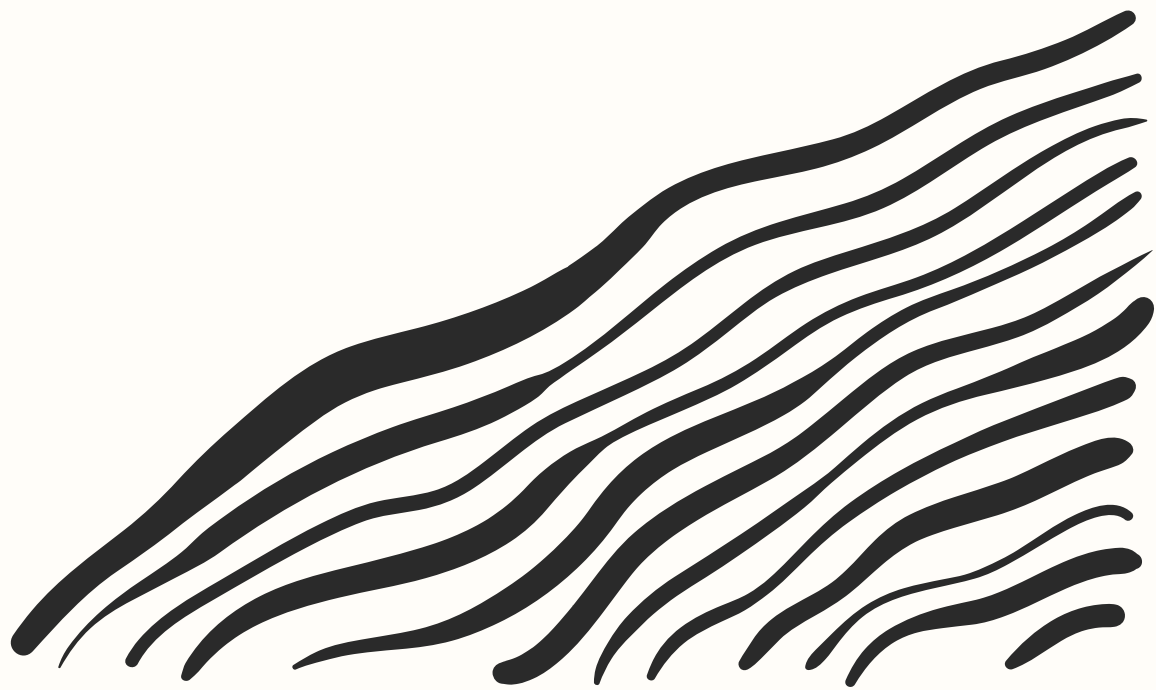


MATERIAL DE ABUSO SEXUAL INFANTIL GERADO POR IA



RELATÓRIO DE 2026
INTERNET WATCH FOUNDATION

MATERIAL CRIADO
POR JARID ARRAES



OLÁ!

Oi, eu sou Jarid Arraes. Este material é parte do conteúdo que disponibilizo para os participantes dos meus cursos, como *Escrevendo o Trauma*, e que indico como referência e fonte para os ensaios que publico online.

Este é um trabalho que compartilho gratuitamente e foi feito exclusivamente por mim, assim como as traduções dos estudos científicos, e não pode ser utilizado para fins comerciais. Todos os créditos devem ser atribuídos.

Para ler o relatório original em inglês e consultar todas as fontes, [clique aqui](#).

Compartilhe com mais pessoas e boa leitura!



AVISO LEGAL E NOTA SOBRE O CONTEÚDO

Ao longo deste relatório, **o material de abuso sexual infantil (CSAM) gerado por meio de inteligência artificial (IA)** será denominado como **AI CSAM**. Este documento não contém nenhum *AI CSAM* ou representação visual de abuso sexual infantil.

O texto apresenta, no entanto, descrições dos métodos empregados para gerar *AI CSAM*, acompanhadas de comentários transcritos na íntegra de criminosos, trechos editados de denúncias públicas e percepções do canal de denúncias da *Internet Watch Foundation* (IWF). Por esse motivo, algumas seções podem ser angustiantes para os leitores.

Embora este relatório faça referências pontuais a pesquisas externas, ele se baseia fundamentalmente nos resultados consolidados a partir de nossa experiência interna. Recomenda-se discricção à leitura.

Sinta-se à vontade para citar ou reproduzir os dados e as conclusões deste documento. Ao fazê-lo, certifique-se de creditar a *Internet Watch Foundation* (IWF) claramente como a fonte. Sempre que possível, inclua o link para o relatório completo: [iwf.org.uk/aireport](https://www.iwf.org.uk/aireport).

PREFÁCIO

Desde que ingressei na *Internet Watch Foundation* (IWF), em agosto de 2025, a magnitude dos danos causados às crianças que enfrentamos diariamente tem me chocado. Em nenhum outro cenário isso é tão evidente quanto na rápida ascensão e no desenvolvimento do *AI CSAM*. Os números apresentados aqui não são abstratos, tampouco estamos lidando com um problema teórico. Cada arquivo de *AI CSAM* representa a violação dos direitos mais básicos à segurança, à dignidade e à própria infância.

Embora a proporção de material gerado por IA ainda seja comparativamente pequena em relação ao imenso volume de *CSAM* que nossos analistas combatem todos os anos, a quantidade e a gravidade das imagens criadas por IA aumentaram de forma exponencial desde o nosso primeiro relatório sobre o tema, em outubro de 2023, graças à disponibilidade e à facilidade de uso dessas ferramentas. Enfrentamos hoje um cenário tecnológico capaz de gerar infinitas violações com uma facilidade sem precedentes. A menos que encaremos esse problema de frente, a tendência é que ele apenas continue crescendo.

O material gerado por IA que encontramos é altamente fotorrealista e, muitas vezes, não se distingue das fotografias reais de abuso. Com frequência, nossos especialistas só conseguem perceber que certas imagens de IA são sintéticas porque reconhecem as vítimas. Eles conhecem essas crianças e sabem quando a imagem retrata algo que nunca aconteceu de fato. O *AI CSAM* não se resume a rascunhos grosseiros mantidos às escondidas, nem serve como mero desvio de atenção de outras formas de imagens abusivas ou de abusos com contato físico. Estamos falando de falsificações sofisticadas que apagam por completo a fronteira entre o sintético e o fotográfico. Os próprios criminosos reconhecem isso e se mostram cada vez mais impressionados com o grau de sofisticação e o extremismo do material que conseguem criar.

Para frear a proliferação desse conteúdo, a IWF trabalha diretamente com os desenvolvedores de modelos de IA, por meio do nosso quadro de membros, e em articulação com legisladores, órgãos reguladores e as forças de segurança. Antes de mais nada, instamos os prestadores de serviço a assumirem maior responsabilidade na garantia de segurança dos seus produtos. Apesar disso, ainda vemos algumas empresas lançarem ferramentas de IA generativa sem a implementação das devidas salvaguardas. Quando as empresas de tecnologia priorizam a retenção da atenção do usuário e a fatia de mercado em detrimento dos princípios de *safety-by-design* (*segurança desde a concepção*), elas criam vetores de abuso. A segurança infantil deve ser incorporada desde o início do desenvolvimento, e não tratada como uma preocupação secundária.

Vemos com bons olhos a iniciativa do governo do Reino Unido de conferir a autoridades designadas, como a IWF, a prerrogativa de testar modelos de IA. Trata-se de um passo importante para promover um *safety-by-design* eficaz, tanto na fase pré-lançamento dos modelos quanto de maneira contínua. Na condição de autoridade global no combate a imagens de abuso sexual infantil, estamos prontos para firmar parcerias com acadêmicos, instituições, setor privado e outras organizações, com o intuito de atuar como uma auditoria robusta e independente em todas as fases do desenvolvimento de produtos. Não podemos enfrentar esse risco sozinhos. Precisamos que o *safety-by-design* se torne um padrão inegociável na criação de IAs, pautado por normas claras estabelecidas pelos governos. Essa tecnologia possui um potencial extraordinário para melhorar as nossas vidas, mas isso só será possível se mitigarmos os danos primeiro. A injustiça não é um destino inevitável. Temos uma oportunidade agora; não podemos desperdiçá-la.

Kerry Smith

Diretora-Executiva

Internet Watch Foundation

SUMÁRIO EXECUTIVO

O uso da IA para gerar material de abuso sexual infantil continua aumentando em escala, sofisticação e facilidade de acesso.

As principais conclusões deste relatório são:

1) O *AI CSAM* é generalizado e está em crescimento: Em 2025, avaliamos 8.029 imagens e vídeos gerados por IA que retratavam cenários realistas de abuso sexual infantil. Embora o *AI CSAM* ainda represente uma parcela relativamente pequena do total de denúncias, os volumes seguem subindo de maneira constante. Essas imagens circulam tanto na *dark web* quanto nas plataformas comerciais convencionais da *clear web*.

2) O *AI CSAM* está cada vez mais extremo e sofisticado: Hoje, vídeos realistas gerados por IA, com movimento contínuo, são muito comuns. Em 2025, a IWF identificou 3.443 vídeos de abuso sexual infantil criados por IA — um aumento de 26.385% em relação ao ano anterior, quando havíamos registrado apenas 13 vídeos desse tipo. Desse total, 65% (2.233 vídeos) foram enquadrados na Categoria A, a classificação de maior gravidade. A título de comparação, considerando todo o ano de 2025, 43% dos vídeos criminosos convencionais (não gerados por IA) analisados pela IWF pertenciam à Categoria A.

EM 2025,
A IWF AVALIOU

8.029

IMAGENS E VÍDEOS
GERADOS POR IA COMO
REPRESENTAÇÕES
REALISTAS DE ABUSO
SEXUAL INFANTIL

A IWF IDENTIFICOU

3.443

VÍDEOS DE
ABUSO SEXUAL
INFANTIL GERADOS
POR IA

AUMENTO DE

26.385%

NO NÚMERO DE VÍDEOS
DE ABUSO SEXUAL
INFANTIL GERADOS
POR IA

65%

DO MATERIAL FOI
ENQUADRADO NA
CATEGORIA A

3) O *AI CSAM* causa danos reais e possui um forte viés de gênero: Os modelos generativos podem ser treinados e submetidos a processos de *fine-tuning* utilizando fotografias reais de abusos, o que revitimiza os sobreviventes de forma direta. O *AI CSAM* alimenta o interesse sexual por crianças, normaliza a violência extrema e aumenta o risco de crimes de abuso com contato físico. Nossas análises também mostram que as meninas continuam sendo o alvo principal, correspondendo a 97% das imagens ilegais geradas por IA avaliadas pela IWF em 2025.

4) *Chatbots* de abuso sexual infantil estão acessíveis na *clear web*: Analistas da IWF identificaram o compartilhamento de imagens de abuso sexual infantil geradas por IA em serviços de *chatbots*. Esses assistentes virtuais incentivam os usuários a encenar simulações de abuso sexual infantil, incluindo interações como “criança prostituída em um hotel”.

5) As ferramentas de *AI CSAM* estão convergindo: Os avanços na IA impulsionaram a convergência de ferramentas que antes exigiam conhecimentos ou programas distintos. Hoje, aplicativos unificados conseguem gerar imagens abusivas com o mínimo de esforço, eliminando a necessidade de qualquer domínio técnico e reduzindo drasticamente as barreiras de entrada. Cada vez mais, essas capacidades se combinam a ferramentas de geração de áudio por IA, o que amplifica o realismo e o potencial destrutivo do conteúdo.

INTRODUÇÃO

A IWF foi uma das primeiras organizações a soar o alarme sobre o *AI CSAM*. Em 2023, publicamos o nosso primeiro relatório voltado à inteligência artificial, mapeando como o *AI CSAM* vinha sendo produzido e delineando essa ameaça emergente. Em julho de 2024, lançamos um segundo relatório com atualizações sobre a evolução tecnológica, alertando que os saltos na IA provocariam uma escalada rápida no volume, na gravidade e na sofisticação das imagens de abuso sexual infantil.

Nossos levantamentos anteriores documentaram os primeiros exemplares de *AI CSAM* em circulação, explicaram a mecânica das ferramentas que começavam a surgir e traçaram projeções claras sobre o rumo e a capacidade da tecnologia. Aquilo que era uma projeção tornou-se uma realidade consolidada e profundamente preocupante. Atualmente, os vídeos realistas de abuso sexual infantil gerados por IA — com fluidez total de movimento — são uma ocorrência banal na internet.

Apenas em 2025, a IWF identificou 3.443 vídeos do tipo, um aumento espantoso de 26.385% frente ao ano anterior, que havia registrado somente 13 casos.

Embora o *AI CSAM* ainda corresponda a uma parcela menor do montante total de conteúdo ilegal combatido pela IWF e por centrais de denúncia em todo o mundo, a velocidade de expansão é alarmante, e a qualidade gráfica e o realismo do material causam profunda perturbação. As evidências e análises da IWF contestam frontalmente o argumento de que o *AI CSAM* seria menos nocivo do que o *CSAM* “tradicional”.

Na prática, imagens de abusos reais são usadas rotineiramente no treinamento desses modelos; também vemos o uso da IA para aprimorar ou recriar fotografias de baixa resolução, dando-lhes nitidez absoluta. Os criminosos ainda se valem de *chatbots* de IA para simular diálogos de teor sexual com crianças e divulgar métodos de geração instantânea de *CSAM*, processos que dispensam qualquer esforço ou habilidade técnica.

Este relatório busca colocar o impacto humano no centro da discussão sobre o *AI CSAM*, expondo os prejuízos reais que ele impõe às crianças e à sociedade. O texto capta a perspectiva dos nossos analistas especializados — que atuam na linha de frente removendo *AI CSAM* da internet — e compila trechos extraídos de comunidades criminosas na *dark web*, onde os usuários celebram sem qualquer pudor a facilidade de acesso e a sofisticação gráfica do abuso gerado por IA.

Como uma organização legalmente autorizada a visualizar, avaliar e reter imagens de abuso sexual infantil, a IWF trabalha há anos junto às empresas de tecnologia no desenvolvimento e no treinamento de ferramentas focadas em segurança. Como autoridade global no setor, estamos a postos para atuar em conjunto com pesquisadores, instituições, setor privado e outras organizações, prestando uma auditoria rigorosa e independente em todas as fases de desenvolvimento tecnológico.



AI CSAM: COMO É GERADO

01

Desde que a IWF passou a monitorar a inteligência artificial, no início de 2023, observamos um salto rápido e assustador na capacidade de gerar imagens artificiais de abuso sexual infantil. Nesta seção, explicamos como o *AI CSAM* é produzido e detalhamos as ferramentas usadas pelos criminosos na criação desse material.

MODELOS DE IA E A GERAÇÃO DE AI CSAM

AI CSAM é o termo guarda-chuva para os conteúdos em vídeo ou imagem que retratam o abuso sexual de crianças criados de forma integral ou parcial por meio de sistemas de IA generativa. Esses sistemas são multimodais: operam com textos, imagens, áudios, vídeos ou códigos. A multimodalidade abrange tanto a inserção de dados quanto o resultado final, ou seja, as IAs conseguem criar peças visuais a partir de instruções em texto ou de imagens prévias.

Os principais modelos de IA que preocupam a IWF são os *LLMs* (*Large Language Models* — *grandes modelos de linguagem*), os *apps de nudify* (aplicativos de desnudamento digital) e os modelos *text-to-image* (conversão de texto em imagem).

Os *LLMs* são IAs treinadas em bases de dados textuais massivas para reconhecer padrões e elaborar textos com linguagem natural. Esses sistemas ganharam enorme popularidade e sofisticação nos últimos anos, tendo como expoentes o ChatGPT, da OpenAI, e o Gemini, do Google. Vá para a Seção 05 para entender melhor nossas preocupações com os *chatbots* de IA, ou para a Seção 04 para ler a respeito dos *apps de nudify*.

Já os modelos *text-to-image* têm a capacidade de criar elementos visuais a partir de comandos de texto. Assim como ocorre com a linguagem escrita, o sistema realiza uma varredura para extrair imagens da internet e as classifica com rótulos descritivos. É isso que permite ao modelo aprender a gerar imagens em resposta às instruções fornecidas pelos usuários — os chamados *prompts*.

O sistema fragmenta os *prompts* em pedaços menores (*os tokens*) e os cruza com conceitos visuais que ele absorveu analisando bilhões de imagens, vídeos e descrições ao longo de seu treinamento. Entre esses conceitos, constam formatos, cores, texturas e nomenclaturas de objetos. Os modelos *text-to-image* mais populares do mercado incluem o Midjourney AI (da Midjourney Inc.) e o Stable Diffusion (da Stability AI). Outro exemplo relevante é o Sora (da OpenAI), classificado como um modelo *text-to-video*.

Como as ferramentas de IA conseguem interpretar conceitos como “criança” e “atividade sexual”, torna-se possível combinar essas noções para produzir *CSAM* — a menos que salvaguardas eficazes bloqueiem a ação.

Para uma visão aprofundada da mecânica por trás da geração de *AI CSAM*, consulte o nosso Relatório de Outubro de 2023. Para compreender os aspectos legais que tornam o *AI CSAM* criminoso, leia a Seção 03.

FINE-TUNING E MODELOS LORA (LOW-RANK ADAPTATION)

O *fine-tuning* (*ajuste fino*) é um processo executado em um modelo de IA já pré-treinado. Em vez de criar um sistema do zero, os desenvolvedores pegam um modelo de base e o submetem a um treinamento secundário usando uma base de dados menor e mais direcionada, a fim de moldar os resultados que ele entrega.

O *LoRA* (*Low-Rank Adaptation* — *adaptação de baixo posto*) é a técnica específica de *fine-tuning* que tem impulsionado a proliferação acelerada do *AI CSAM*. Por reduzir o nível técnico exigido, o tempo gasto e os custos financeiros necessários para personalizar as IAs generativas, o *LoRA* democratizou de vez a fabricação de imagens hiper-realistas. A técnica funciona incorporando camadas pequenas e adicionais que alteram o comportamento do modelo de base (como um LLM) usando poucos dados, ao mesmo tempo em que congela os pesos originais do modelo central.

Esse arranjo torna o *LoRA* muito mais eficiente do que o *fine-tuning* tradicional. Para se ter uma ideia, basta ter em mãos entre 20 e 30 fotos para criar um *deepfake* perfeito de uma criança específica. O processamento computacional exigido é mínimo, e o treinamento pode ser concluído em apenas 15 minutos. Esses requisitos operacionais tão baixos permitem que os *LoRAs* rodem em laptops e computadores domésticos comuns. Consequentemente, os “meios de produção” saíram das mãos dos especialistas em tecnologia e migraram para os usuários do dia a dia, inflando drasticamente a quantidade de potenciais criminosos.

O tamanho reduzido dos arquivos de *LoRA* agrava ainda mais a ameaça. Eles são fáceis de compartilhar, reutilizar e instalar em qualquer plataforma, viabilizando a clonagem imediata do modelo assim que ele é finalizado. Por serem leves e funcionarem 100% offline, os *LoRAs* são imunes a grande parte dos esforços de remoção da internet. Uma vez feito o download, o arquivo pode ser repassado em canais privados e operado sem nenhuma dependência de plataformas de hospedagem em nuvem, o que torna a detecção do crime e a ação policial processos difíceis.

OBSERVAÇÕES DA LINHA DE FRENTE

Nossos analistas detectaram múltiplos episódios de circulação de *LoRAs*, incluindo modelos treinados com o rosto de vítimas conhecidas. Em um único fórum da *dark web* monitorado em dezembro de 2025, contabilizamos 202 modelos disponíveis. O acervo incluía IAs treinadas para simular as feições exatas de vítimas reais de abuso, atores infantis, influenciadores e atletas mirins de destaque; modelos configurados para recriar genitálias realistas de faixas etárias específicas; e modelos programados para forjar imagens e vídeos de posturas ou atos sexuais específicos.

Eventualmente, é possível extrair metadados de um modelo treinado ou de um *LoRA* ajustado.

Esses dados — como legendas e nomes originais dos arquivos fotográficos usados no treinamento — às vezes ficam embutidos no resultado final da imagem, entregando quais softwares, modelos de base e *LoRAs* foram empregados no momento da criação. O problema é que a remoção total desses metadados é um processo simples, o que frustra os esforços de rastrear a origem do material. Esses modelos independentes transformam uma única ocorrência de abuso em uma mercadoria inesgotável e altamente negociável, o que multiplica os danos em proporções gigantescas. As brechas e inconsistências nas legislações nacionais e regionais sobre o *CSAM* — em especial no tocante às IAs — geram riscos sistêmicos: os criminosos exploram vazios jurídicos, as plataformas operam sob regras fragmentadas e as crianças seguem desprotegidas. Sem uma estrutura de leis unificada e abrangente, a disseminação das IAs apenas escala a gravidade e o alcance do material de abuso.

LEGISLAÇÃO SOBRE GERADORES DE IMAGENS DE CSAM

Reino Unido

Em fevereiro de 2025, o Governo do Reino Unido anunciou um pacote de medidas atrelado ao Projeto de Lei sobre Crimes e Policiamento (*Crime and Policing Bill*) focado em fechar as brechas legais sobre o *AI CSAM*. A medida prevê a tipificação de um novo crime para quem fabricar, adaptar, armazenar, fornecer ou ofertar geradores de imagens de *CSAM*. Até o fechamento deste documento, o Projeto de Lei se encontrava na etapa de relatoria da Câmara dos Lordes e deve receber a Sanção Real dentro de poucas semanas.

Estados Unidos

Não existe no âmbito federal estadunidense uma lei que criminalize de forma explícita os geradores de imagens de *CSAM*. A lei *STOP CSAM*, reapresentada ao Congresso em maio de 2025, estende a responsabilização aos provedores que hospedam, armazenam ou impulsionam o *CSAM* em seus servidores.

Se o material de *AI CSAM* forjar os traços de uma criança “real”, a produção se enquadra na legislação federal contra o abuso infantil; caso o material seja 100% sintético, a conduta pode ser processada baseando-se nas leis federais de obscenidade. Os criminosos ficam sujeitos à esfera penal, enquanto as plataformas podem ser penalizadas em até 1 milhão de dólares por infração, cifra que salta para 5 milhões de dólares se ficar comprovado que uma criança sofreu danos materiais. Recentemente, um tribunal distrital proferiu uma decisão pontuando que o armazenamento de *AI CSAM* para fins estritamente privados pode estar resguardado pela Constituição, mas a fabricação do conteúdo não, o que coloca o uso pessoal em uma zona de risco jurídico parcial.

União Europeia

A revisão da Diretiva sobre o Abuso Sexual Infantil (2011/93/UE) passa no momento por debates interinstitucionais que visam modernizar a tipificação penal, agravar punições e inserir cláusulas diretas contra o *AI CSAM*. O novo texto criminaliza a criação, o *fine-tuning*, a posse, a distribuição ou a oferta de modelos de IA fabricados com a intenção de gerar *CSAM*. A perspectiva é que as punições sejam equiparadas aos crimes mais amplos de *CSAM*, embora a redação final dependa do aval dos legisladores. A IWF endossa a postura enérgica do Parlamento Europeu e os artigos direcionados aos geradores de imagens de *CSAM*.

MODELOS OPEN-SOURCE E CLOSED-SOURCE

Boa parte das IAs generativas convencionais (como o ChatGPT e o Midjourney) operam em formato *closed-source* (*código fechado*), mantendo a arquitetura de programação e os processos internos sob sigilo. Em contrapartida, os modelos *open-source* (*código aberto*) divulgam sua estrutura de forma livre, o que permite a qualquer internauta baixar, adaptar e rodar o *fine-tuning* do sistema. Apesar de a tecnologia *open-source* ser essencial para o avanço da pesquisa e da inovação, ela confere a pessoas mal-intencionadas os meios necessários para construir modelos próprios offline.

Treinados com bases de dados repletas de *CSAM*, esses modelos passam a gerar novas imagens sem exigir qualquer expertise técnica por parte do usuário. Como rodam nas máquinas locais, aceitam qualquer tipo de prompt irrestrito e não contam com nenhuma trava de moderação ou detecção de abusos, os sistemas *open-source* consolidaram-se como a ferramenta predileta dos criminosos para a fabricação de *AI CSAM*.

As estimativas atuais indicam que o grau de sofisticação tecnológica dos modelos *open-source* costuma estar alguns meses atrasado em relação aos sistemas comerciais operados em nuvem, mas os especialistas alertam que essa diferença de prazo está encolhendo. Sendo assim, os criminosos que operam na via do código aberto tendem a absorver os recursos recém-lançados pelas IAs comerciais de código fechado muito em breve, o que alavancará ainda mais a qualidade visual e a quantidade do material de abuso.

Vale destacar que os modelos de IA baseados em nuvem se definem pela forma como a infraestrutura é hospedada, e não pelo fato de o código ser aberto ou fechado. Esses sistemas dependem de acesso à internet e rodam nos servidores centrais operados pelas big techs, e não no processador da máquina do usuário.

Analistas da IWF constataram que os próprios criminosos recomendam ativamente não usar modelos baseados em nuvem — a exemplo do Grok Imagine, da xAI — na hora de gerar *CSAM*. Alguns fóruns banem de forma sumária os usuários que tentam fazê-lo. Nessas comunidades da *dark web*, o consenso é de que as plataformas em nuvem geram riscos de vigilância, retenção de histórico de dados e rastreabilidade, uma vez que as atividades executadas nesses servidores podem apontar para a identidade do autor do crime. Por isso, os modelos em nuvem têm aceitação baixíssima nessa comunidade, que valoriza sobretudo o controle local e o anonimato.

ONDE O AI CSAM APARECE

O *AI CSAM* é hospedado e disseminado por múltiplos ambientes virtuais, transitando desde fóruns da *dark web* até páginas da *clear web*, o que escancara a atualidade e a vivacidade dessa ameaça.

No entanto, aquilo que conseguimos enxergar ilustra apenas uma parte do panorama. O verdadeiro volume de modelos focados em *CSAM* submetidos a *fine-tuning* — incluindo as IAs programadas para forjar rostos de crianças específicas — é, com toda probabilidade, muito maior do que aquilo que o monitoramento dos domínios públicos da *clear web* e da *dark web* consegue detectar. Nos fóruns da *dark web*, a equipe da IWF acompanha de perto os criminosos permutando acervos particulares de *CSAM* e arquivos de modelos base dentro de subcomunidades fechadas. Essa rede de colaboração facilita o refinamento de sistemas forjados sob medida; há tópicos inteiros dedicados ao compartilhamento de modelos de IA que ultrapassam a marca de dezenas de milhares de acessos.

Grande parte do *AI CSAM* também transita à margem da capacidade de rastreamento da IWF. Como arquivos de *fine-tuning* e conjuntos de dados podem ser alocados em computadores offline, o compartilhamento ocorre por meio de redes *P2P* (*peer-to-peer*) seladas com criptografia de ponta a ponta (*E2EE*). Nessas redes, não apenas a distribuição, mas o próprio desenvolvimento dos modelos, acontece longe dos olhos dos provedores de hospedagem centralizados. Tais ecossistemas impõem barreiras severas ao trabalho da IWF e colaboram de forma drástica para que o volume e a abrangência reais do *AI CSAM* sejam sempre subestimados pelas estatísticas globais.

Para entender como as plataformas podem bloquear a transferência e o upload de *CSAM* em redes com criptografia *E2EE* de maneira que preserve a privacidade dos usuários convencionais, leia o estudo técnico da IWF: *Prevenção de Upload de CSAM em Ambientes E2EE*.

CITAÇÃO DE UM CRIMINOSO

“Vídeos de meninos à venda, vídeos raros, a coleção com a qual qualquer predador de garotos sonha, preço: 88 usd, só aceito Litecoin, me chama no tele [Telegram] pra ver uma prévia se estiver interessado!”

DENÚNCIA PÚBLICA

“Ele tem anunciado isso no Telegram via [ferramenta ocultada], onde usa várias contas para exibir vídeos inadequados gerados por IA com menores de idade. Se você pesquisar por '[nome de usuário ocultado]', vai encontrar todas as contas dele.”

OBSERVAÇÕES DA LINHA DE FRENTE

O *AI CSAM* é compartilhado tanto na *dark web* quanto na *clear web*. Frequentemente, ele é postado lado a lado com o *CSAM* convencional. Enquanto a *dark web* funciona como um pólo em que os usuários debatem os processos criativos e compartilham seus fluxos de trabalho, a *clear web* é usada prioritariamente para fins comerciais: ali, os usuários anunciam e monetizam os conteúdos que geram.

Já vimos anúncios divulgando “cursos” customizados que prometiam ensinar qualquer pessoa leiga a criar imagens ultrarrealistas de adolescentes via IA. Muitos desses anúncios orientam o comprador a chamar o vendedor no ambiente dos aplicativos criptografados:

“Se alguém souber de uma plataforma de imagens de IA melhor, me avisa, por favor! Quero que as minhas imagens fiquem o mais realistas possível pra vocês. Vocês merecem o melhor. Meu telegram é * * * *”

DANOS DO AI CSAM

02

Imagens sexuais de crianças geradas por IA causam danos reais e permanentes às vítimas — mesmo que o abuso físico retratado não tenha de fato ocorrido. Sob a ótica de uma vítima ou sobrevivente, independentemente de a imagem ter sido originada por uma câmera ou por um *software*, o impacto é profundamente traumático e representa uma violação do seu direito de não ser submetida a tortura ou a tratamentos desumanos e degradantes.

Sobreviventes de abuso sexual infantil relatam, com frequência, danos psicológicos profundos e duradouros, incluindo sentimentos de vergonha, humilhação, raiva, ansiedade e uma forte sensação de violação. Muitos se isolam socialmente, têm dificuldade em confiar nos outros e, em alguns casos, vivenciam episódios de automutilação ou ideação suicida. Uma característica marcante das imagens de abuso sexual infantil é a sua permanência: uma vez que uma imagem ou vídeo passa a existir, ele pode ser copiado, alterado e redistribuído infinitamente, tirando das vítimas qualquer sensação de controle ou de encerramento do trauma. Em janeiro de 2026, Mara Wilson — ex-atriz mirim do clássico filme *Matilda* — descreveu a experiência de ser vítima de *AI CSAM* como “perder o controle sobre a própria narrativa”.

Existem tentativas de argumentar que o *AI CSAM* causa menos danos do que o *CSAM* “real”, seja por não envolver diretamente o cometimento de crimes *offline* contra crianças, seja sob a alegação de que um material inteiramente sintético não teria uma vítima “real”. Isso levou a sugestões de que o material gerado por IA poderia funcionar como uma alternativa “segura” ou terapêutica ao *CSAM* “tradicional”. A legislação do Reino Unido determina que a geração de qualquer tipo de *CSAM* — seja “tradicional” ou totalmente sintético — é ilegal. Não há nenhuma evidência que sustente a alegação de que o *AI CSAM* sirva como um método eficaz de redução de danos para criminosos. Em vez disso, danos evidentes se manifestam no mundo real como resultado da visualização e do compartilhamento desse conteúdo.

Esta seção detalha os danos do *AI CSAM*, desmentindo as alegações de que ele seria menos prejudicial do que outras formas de *CSAM*. Ela explica como a IA revitimiza vítimas e sobreviventes, perpetua danos pautados por questões de gênero, normaliza a violência sexual contra crianças e causa prejuízos operacionais aos sistemas de proteção infantil.

REVITIMIZAÇÃO

Ainda em 2023, a IWF documentou como as ferramentas de IA são usadas para manipular materiais preexistentes de abuso sexual infantil que envolvem sobreviventes conhecidos pelos nossos analistas.

Como já destacado, os criminosos pegam imagens reais de abuso e usam a IA para gerar novas versões — frequentemente mais extremas. Para os sobreviventes, isso equivale a sofrer o abuso repetidas vezes, anos após o trauma original, sem nenhuma possibilidade de impedir que isso aconteça.

Isso não é teórico. Nosso relatório de 2024 descreveu a história de uma criança a quem chamaremos de Olivia. Entre os três e oito anos de idade, Olivia sofreu abusos sexuais terríveis pelas mãos de alguém em quem confiava. Ela é bem conhecida pelos analistas da IWF, pois as imagens feitas pelo criminoso circularam amplamente na internet. Nos últimos anos, documentamos modelos *LoRA* — disponíveis para *download* gratuito — que criminosos treinaram usando as imagens de Olivia. Hoje ela está na faixa dos 20 anos, mas seu sofrimento pode não ter fim.

Em 2024, o jornal *The Guardian* contou a história de Megan, uma sobrevivente anonimizada de abuso sexual infantil. Megan relata como a IA fez com que o seu trauma parecesse contínuo, apesar de o crime original ter ocorrido há mais de uma década. Ao longo de 12 meses, a perspectiva de ter suas imagens manipuladas digitalmente tornou-se cada vez mais angustiante, conforme ela mesma descreve:

“A IA dá aos agressores a chance de criar ainda mais situações do meu abuso para alimentar suas próprias fantasias e suas próprias versões. A forma como minhas imagens poderiam ser manipuladas com IA daria a falsa impressão de que aquilo não foi prejudicial ou de que eu estava gostando do abuso.”

Independentemente de as imagens serem criadas por meios tradicionais ou geradas por IA, o efeito sobre as crianças é o mesmo: exploração contínua, perda de controle e traumas repetidos. Uma pesquisa de 2017 com vítimas e sobreviventes conduzida pelo Centro Canadense de Proteção à Criança (*C3P*) revelou que quase 70% dos sobreviventes de imagens de abuso sexual infantil entrevistados vivenciam um medo constante de serem reconhecidos por alguém que tenha visto os materiais; 30% disseram que isso já havia acontecido. Dois terços citaram a circulação interminável de imagens como a principal fonte de trauma, enquanto mais de um terço relatou sentimentos profundos de impotência. Como resultado do abuso original que sofreram, 87% vivenciaram danos adicionais, incluindo perseguição (*stalking*), propostas sexuais, chantagem e uma nova vitimização. Essas descobertas deixam claro que a revitimização contínua por meio de imagens de abuso sexual infantil causa danos psicológicos graves e de longo prazo. Todas as formas de material desse tipo devem ser tratadas com a mesma urgência.

OBSERVAÇÕES DA LINHA DE FRENTE

“O conteúdo se tornou tão realista, e a velocidade com que essa tecnologia está se desenvolvendo é realmente alarmante.”

“Agora estamos vendo novas imagens [dela] — imagens geradas por IA — e algumas são ainda mais graves do que as fotos que foram de fato tiradas na realidade.”

“Isso está muito longe de ser um crime sem vítima — há uma vítima muito real aqui, e acredito que esse conteúdo causa danos verdadeiros.”

DANOS PAUTADOS POR GÊNERO

Casos e dados notórios têm destacado uma dimensão de gênero evidente na solicitação, criação e compartilhamento de imagens de abuso sexual infantil. Isso é particularmente grave quando se trata de material gerado por IA. Em 2025, as meninas continuaram sendo o sexo retratado com mais frequência nas imagens geradas por IA: elas apareceram em 97% dos materiais nos quais o sexo pôde ser identificado neste ano, contra 99% em 2024.

Em 2023, houve ampla repercussão de um caso na Espanha em que 20 meninas de 11 a 17 anos se apresentaram como vítimas de *apps* de *nudify*. Onze meninos foram identificados por envolvimento na criação ou na circulação dessas imagens em plataformas como Telegram e WhatsApp.

Outro exemplo na União Europeia ocorreu na Romênia, onde um aluno do ensino médio usou tecnologia *deepfake* para “despir” 11 colegas de classe. Foi relatado que todas as vítimas tinham algum vínculo com o estudante e que as suas ações — despi-las virtualmente — foram uma tentativa de exercer poder sobre elas.

Em 2024, a Coreia do Sul enfrentou uma crise em escala industrial de imagens de abuso sexual geradas por IA. Milhares de imagens em *deepfake* foram criadas e circularam por redes coordenadas no Telegram, atingindo mais de 200 escolas, universidades e até mesmo as forças armadas. Adolescentes e jovens na faixa dos 20 anos representaram 92% das vítimas, sendo as meninas e mulheres jovens os alvos principais. Mais de 80% dos que acabaram detidos eram, eles próprios, adolescentes. A dimensão, a rapidez e o direcionamento hiperlocal do abuso pressionaram o governo a criminalizar a posse e a visualização de material sexual em *deepfake*. A nova legislação sul-coreana tipifica essas condutas como crime e aumenta para sete anos as penas já existentes para a criação e o compartilhamento de conteúdo dessa natureza.

Esses padrões refletem uma persistente violência sexual online baseada em gênero, reforçada por culturas que normalizam a misoginia e a humilhação de meninas. Para mais informações sobre os *apps de nudify*, consulte a Seção 04.

A *Pesquisa Anual Girlguiding de 2024* revelou que 59% das meninas e mulheres jovens de 11 a 21 anos temiam que a IA pudesse ser usada para criar imagens falsas delas ou para assumir a identidade delas online. Essa preocupação é corroborada por dados de 2026 do *Centro do Reino Unido para uma Internet Mais Segura (UK Safer Internet Centre)*, que constatou que 60% dos jovens estão preocupados que alguém use IA para criar imagens impróprias deles; 61% dos adolescentes de 13 a 17 anos temem que seus pares gerem imagens sexuais de outros jovens; e até mesmo 63% das crianças de 8 a 12 anos compartilham desse medo.

Casos como esses ilustram como as ferramentas de IA podem reproduzir e ampliar a misoginia nos ambientes virtuais, reduzindo as barreiras para o abuso e acelerando a produção de material nocivo em larga escala.

Também é importante notar que, em relação ao material de abuso sexual infantil em um sentido mais amplo, os meninos também são vítimas da violência baseada em gênero e dos abusos de poder.

Meninos mais novos têm mais probabilidade do que meninas de aparecer em materiais de abuso produzidos e distribuídos comercialmente, e os adolescentes mais velhos relatam cada vez mais sobre imagens diretamente a nós por meio do *Report Remove*, um serviço que ajuda jovens com menos de 18 anos no Reino Unido a denunciar confidencialmente conteúdo sexual em que apareçam, a fim de removê-lo da internet. Observamos que adolescentes do sexo masculino são alvos frequentes de extorsão com motivação sexual, um crime que também pode ser facilitado pela tecnologia de IA.

NORMALIZAÇÃO DA VIOLÊNCIA SEXUAL CONTRA CRIANÇAS

O consumo de material sintético realista reforça interesses sexuais preexistentes por crianças, em vez de atuar como um substituto seguro, e pode aumentar o risco de uma escalada de violência. Já foi comprovado que o engajamento com qualquer forma de *CSAM* consolida o interesse sexual já existente por crianças. Pesquisas realizadas pela organização *Proteja as Crianças Finlândia (Suojellaan Lapsia)* ilustram a possível progressão que vai do consumo visual de *CSAM* até o crime de abuso com contato físico: a pesquisa “Ajude-nos a Ajudar Você” de 2021 constatou que, dos 5.171 entrevistados, 44% disseram que visualizar *CSAM* os fez pensar em buscar contato direto com crianças, e 37% relataram ter efetivamente buscado esse contato após consumirem os materiais.

Essas conclusões são respaldadas pelo estudo de amostragem global conduzido pela mesma organização em 2022. Com base nas respostas de 1.546 indivíduos anonimizados, o levantamento mostrou que 58% dos consumidores de *CSAM* sentem pelo menos um medo ocasional de que seu consumo os leve a cometer atos sexuais contra crianças, com quase um quarto deles (24%) relatando sentir esse medo semanalmente ou quase toda vez que consome o conteúdo. No mesmo estudo, 10% dos entrevistados relataram ter buscado contato online com crianças após visualizar *CSAM* ou outro conteúdo violento ilegal, em uma frequência semanal ou maior. Isso demonstra uma ligação clara entre o consumo de *CSAM*, a dessensibilização e a redução das inibições para cometer novos crimes sexuais, sejam eles online ou offline.

CITAÇÕES DE CRIMINOSOS

“Isso é realmente impressionante. O nível de realismo de algumas delas chega a uns 95% do caminho para ser indistinguível de fotos reais.”

“Por enquanto, na maioria dos casos dá para perceber que são imagens de IA, mas estamos chegando a um ponto em que a IA será indistinguível. Acho que há oportunidades para usarmos a IA a nosso favor, assumindo um papel ativista.”

OBSERVAÇÕES DA LINHA DE FRENTE

Nossos analistas observaram criminosos comentando sobre como é possível gerar imagens de IA de crianças do convívio deles.

“Às vezes me pergunto por que não há mais imagens reais de creches. Com a IA isso não importa mais. Podemos criar toda a diversão que quisermos.”

Um criminoso refletiu sobre a possibilidade de usar câmeras escondidas para gravar imagens inofensivas de crianças conhecidas e convertê-las em vídeos de *AI CSAM* utilizando uma função de IA de *image-to-video (imagem para vídeo)*.

“Estou muito impressionado com os resultados das conversões de imagem para vídeo. Acho que seria realmente empolgante [gerar *CSAM*] tudo a partir da perspectiva subjetiva da minha câmera. Tem alguma IA que consegue usar o vídeo e dar continuidade a ele com conteúdo gerado? Ou é melhor fornecer só um registro rápido ou uma imagem estática?”

Esses exemplos ilustram com clareza como crianças — que moram na mesma região que os criminosos — podem ser vitimadas pelo *AI CSAM*.

PREJUÍZOS OPERACIONAIS

A escala e a velocidade da produção de *AI CSAM* correm o risco de sobrecarregar os canais de denúncia e a capacidade das forças de segurança, que já se encontram saturados.

As autoridades policiais agora precisam lidar com o desafio de identificar vítimas que precisam de proteção, ao mesmo tempo em que navegam por uma verdadeira colcha de retalhos de leis e regulamentações.

A rápida proliferação do *AI CSAM* cria o desafio de buscar uma “agulha num palheiro” para os investigadores. Diferenciar o material sintético do abuso fotográfico consome tempo e recursos significativos, atrasando a resposta para as crianças em perigo iminente. Essa dinâmica aumenta o risco de falsos negativos — nos quais crianças que precisam de resgate passam despercebidas — e de falsos positivos — em que recursos são desviados para casos sintéticos, esgotando a capacidade já limitada das forças de segurança.

A disseminação de imagens sintéticas também ameaça prejudicar os processos criminais, principalmente nas jurisdições em que o *AI CSAM* não é criminalizado da mesma forma que outros tipos de material. Em fevereiro de 2025, a Europol apoiou autoridades de 19 países na *Operação Cumberland*, uma colaboração em grande escala que resultou na identificação de 273 suspeitos e em 25 prisões ao redor do mundo. Os suspeitos faziam parte de um grupo criminoso internacional cujos membros atuavam na distribuição de *CSAM* inteiramente gerado por IA. Na época, a Europol observou que esse foi um dos primeiros casos no mundo em que a ausência de leis nacionais para o *AI CSAM* tornou “o trabalho dos investigadores excepcionalmente desafiador”.

Nessas jurisdições, os criminosos podem alegar que as evidências reais de abuso com contato físico foram geradas por IA e que, portanto, não retratam uma criança identificável, tática por vezes chamada de “defesa do *deepfake*”. Essa abordagem se apoia no que os pesquisadores chamam de “dividendo do mentiroso”, fenômeno em que a crescente conscientização do público sobre as mídias sintéticas cria um ambiente de negação plausível. Embora argumentos desse tipo ainda sejam restritos no contexto do *AI CSAM*, os tribunais já estão lidando com disputas envolvendo mídias sintéticas, e especialistas alertam que a familiaridade crescente com a IA generativa pode permitir que os réus lancem dúvidas sobre a autenticidade de materiais reais.

Apesar de o Reino Unido já possuir um arcabouço jurídico robusto que criminaliza a criação, a posse e a distribuição de *CSAM*, incluindo o conteúdo gerado por IA, esse nível de proteção não encontra equivalência em todas as jurisdições.

Nos Estados Unidos, a lei federal continua fragmentada, com a maior parte do progresso ocorrendo na esfera estadual; enquanto na União Europeia não há atualmente uma proibição uniforme sobre o *AI CSAM*, e lacunas persistem entre os Estados-Membros.

Esse cenário jurídico desigual cria riscos evidentes: os criminosos podem explorar brechas regulatórias, as plataformas enfrentam obrigações inconsistentes e as crianças ficam desprotegidas dependendo de onde o material for criado ou hospedado.

Como a IA generativa reduz as barreiras para produzir conteúdo nocivo em larga escala, a ausência de leis abrangentes e harmonizadas mina cada vez mais os esforços globais em prol da segurança infantil.

LEGISLAÇÃO SOBRE A CRIAÇÃO, POSSE E DISTRIBUIÇÃO DE AI CSAM

Reino Unido

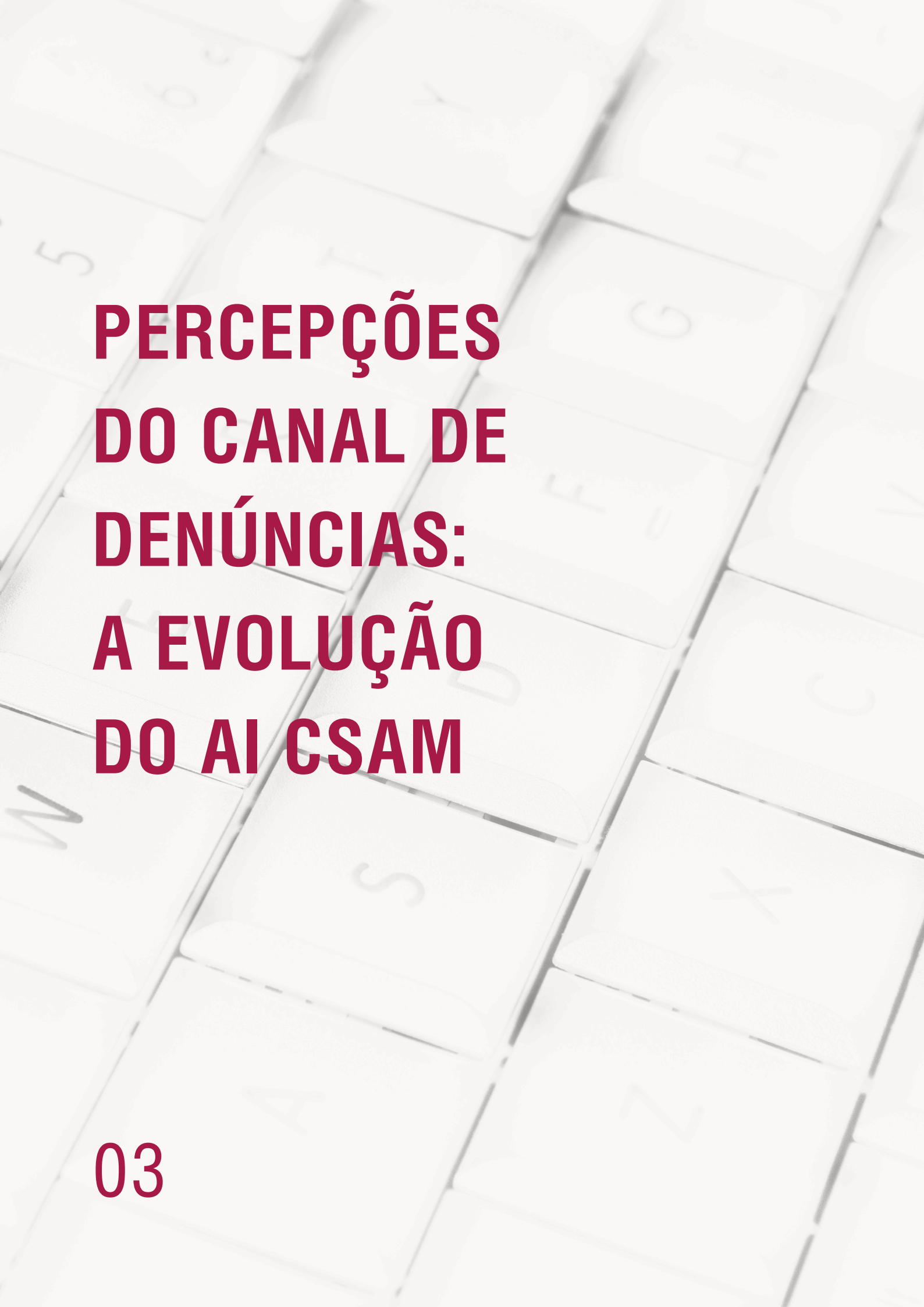
A legislação sobre o conteúdo de *CSAM* já é robusta no Reino Unido. Criar, possuir e distribuir imagens de abuso sexual infantil, sejam de IA ou não, já é uma conduta criminosa. No âmbito da Lei de Proteção às Crianças de 1978 (*Protection of Children Act 1978*), a Seção 1 tipifica como crime tirar, distribuir e possuir uma “fotografia ou pseudo fotografia indecente de uma criança” (*qualquer indivíduo com menos de 18 anos*). A Lei dos Legistas e da Justiça de 2009 (*Coroners and Justice Act 2009*) criminaliza o *CSAM*, mesmo que o material não seja fotorrealista. É crime possuir uma imagem proibida de uma criança; isso engloba as imagens geradas por IA que não são consideradas fotorrealistas, incluindo desenhos animados, ilustrações e animações (Seção 62 da referida lei).

Estados Unidos

Atualmente, os EUA não possuem uma única lei abrangente que aborde o *AI CSAM*. Projetos legislativos federais apresentados em 2025 visam complementar os avanços que ocorreram nos âmbitos estaduais e combater a exploração sexual facilitada pela IA. Na esfera estadual, 45 estados promulgaram estatutos criminalizando explicitamente o *CSAM* gerado por IA ou editado por computador, frequentemente equiparando-o ao *CSAM* tradicional para fins judiciais. A Lei de Interrupção da Extorsão Sexual (*Stop Sextortion Act*), apresentada em 2025, visa criar novos crimes em relação à extorsão sexual de menores, estendendo as leis de obscenidade já existentes para abranger o *AI CSAM*. A Lei de Responsabilidade nas Penas para Exploração (*SAFE Act - Sentencing Accountability for Exploitation Act*), introduzida em 2025, exigiria que as diretrizes federais de condenação fossem atualizadas para refletir o uso de tecnologias digitais modernas e tecnologias baseadas em IA nos crimes de exploração sexual, incluindo o seu uso para esconder o crime, a identidade ou a localização do réu e da vítima.

União Europeia

Atualmente, não existe uma proibição uniforme do *AI CSAM* em toda a UE, e as definições jurídicas variam entre os Estados-Membros. A Diretiva sobre Abuso Sexual Infantil (Reformulação da Diretiva 2011/93/UE) tem o potencial de criminalizar totalmente o *AI CSAM*, muito embora o posicionamento do Conselho da UE proponha uma isenção para “uso pessoal” (Artigo 5, §10), criando uma possível brecha legal. A Lei de IA de 2024 (*AI Act*) estabelece uma estrutura normativa para a IA na UE, reconhecendo o *CSAM* como um risco sistêmico e exigindo que os provedores implementem medidas de mitigação de riscos. No entanto, a referida lei não é um instrumento de direito penal.



PERCEPÇÕES DO CANAL DE DENÚNCIAS: A EVOLUÇÃO DO AI CSAM

03

A IWF é um dos únicos canais de denúncias do mundo autorizados a buscar ativamente por *CSAM* e a agir com base em denúncias feitas pelo público. Isso permite que a instituição monitore e acompanhe as transformações no cenário dos crimes de *CSAM*, inclusive por meio da análise de discussões entre os próprios criminosos. O *AI CSAM* é uma preocupação crescente e tem sido monitorado e combatido de perto pela IWF, especialmente porque as ferramentas de geração de imagens por IA continuam evoluindo de forma rápida e se tornando cada vez mais acessíveis.

Em 2025, avaliamos 8.029 imagens e vídeos gerados por IA que exibiam abusos sexuais infantis realistas. Embora o *AI CSAM* ainda represente uma proporção relativamente pequena do total de denúncias recebidas pela IWF, o volume de casos continua subindo de maneira constante. Essa tendência se reflete no cenário internacional. Um estudo realizado pela UNICEF, ECPAT e INTERPOL em 11 países revelou que pelo menos 1,2 milhão de crianças relataram ter tido suas imagens manipuladas em *deepfakes* de conteúdo sexual explícito no último ano. Em determinados países, isso representa 1 em cada 25 crianças — o equivalente a um aluno por sala de aula convencional.

Vale ressaltar que, com o avanço da tecnologia, aumenta a probabilidade de que materiais gerados por IA sejam classificados como fotografias reais. Atualmente, os analistas registram um conteúdo como gerado por IA apenas quando isso é indicado pela fonte ou pelos metadados do arquivo. Portanto, o volume real de imagens e vídeos gerados por IA avaliados por nossa equipe pode ser ainda maior.

VÍDEOS DE AI CSAM

Nosso relatório de *AI CSAM* de julho de 2024 marcou a primeira vez que os analistas da IWF documentaram a circulação de vídeos de *AI CSAM* totalmente sintéticos. Naquela etapa, a tecnologia era relativamente rudimentar.

Os vídeos caracterizavam-se por intensas distorções e oscilações visuais, com movimentos artificiais e bruscos que tornavam o conteúdo gerado por IA facilmente distinguível de filmagens reais.

OBSERVAÇÕES DA LINHA DE FRENTE

“Em termos de qualidade de vídeo, os criadores de imagens de abuso sexual infantil estão anos-luz à frente de onde estavam no ano passado. Os primeiros vídeos de abuso infantil por IA que vimos eram *deepfakes* — o rosto de uma vítima conhecida colocado no corpo de um ator em um vídeo pornográfico adulto já existente. Não era algo sofisticado, mas ainda assim podia ser bastante convincente. O primeiro vídeo de abuso infantil totalmente sintético que vimos no início do ano passado era apenas uma sequência de imagens travadas dispostas em série, nada convincente. Mas agora eles realmente mudaram de patamar. A qualidade é alarmantemente alta, e as categorias de crimes retratadas estão se tornando mais extremas à medida que as ferramentas aprimoram a capacidade de gerar vídeos com duas ou mais pessoas. Os vídeos também incluem cenários que exibem vítimas conhecidas em novas situações. Assim como as imagens estáticas passaram a ser fotorrealistas conforme a demanda aumentou e as ferramentas foram aprimoradas, era apenas uma questão de tempo para que o mesmo acontecesse com os vídeos.”

Desde então, a tecnologia avançou significativamente. Como previsto, conteúdos em vídeo realistas e com movimento contínuo gerados por IA tornaram-se comuns.

Em 2025, a IWF identificou 3.443 vídeos de abuso sexual infantil gerados por IA, um aumento de 26.385% em comparação ao ano anterior, quando apenas 13 vídeos desse tipo haviam sido registrados. Desses, 65% (2.233 vídeos) foram classificados na Categoria A, a classe jurídica mais grave conforme a legislação do Reino Unido, que abrange crimes como estupro, tortura sexual e zoofilia. Outros 30% (1.027 vídeos) enquadraram-se na Categoria B, a segunda classificação mais grave.

A título de comparação, ao longo de todo o ano de 2025, 43% dos vídeos criminosos não gerados por IA analisados pela IWF pertenciam à Categoria A — o que demonstra que a IA está sendo utilizada para criar conteúdos ainda mais violentos.

Qualquer *CSAM* fotográfico ou pseudofotográfico é categorizado por nível de gravidade, de acordo com as Categorias do Conselho de Sentenças do Reino Unido (*UK Sentencing Council*). São elas:

- Categoria A: Imagens que retratam atividade sexual com penetração; imagens que envolvem atividade sexual com animais; ou sadismo.
- Categoria B: Imagens que retratam atividade sexual sem penetração.
- Categoria C: Outras imagens indecentes que não se enquadram nas categorias A ou B.

Para uma análise aprofundada de como os vídeos de IA são gerados, consulte o nosso Relatório de 2024.

AI CSAM COM COMPONENTE DE ÁUDIO

O *AI CSAM* que possui um componente de áudio é uma nova vertente de preocupação. Isso pode se manifestar na forma de gravações de abusos, *deepfakes* de áudio — que geram artificialmente vozes infantilizadas em contextos sexuais — ou gravações manipuladas para simular cenários de abuso.

Novos modelos de IA permitem que os usuários combinem uma foto, um arquivo de áudio e uma descrição em texto para gerar automaticamente um vídeo com voz sintética.

Embora esse formato ainda não seja tão disseminado quanto o *AI CSAM* em imagem ou vídeo, já existem indícios iniciais de demanda por esse tipo de conteúdo em comunidades de criminosos. Os analistas da IWF identificaram vídeos de *AI CSAM* que parecem conter dublagens ou narrações clonadas.

Diversos modelos novos foram lançados recentemente, permitindo que o usuário insira uma imagem, adicione um áudio, descreva o que deseja ver e o sistema transforme a imagem e a voz em um vídeo em movimento. Se o problema não for combatido, a tendência é que essa demanda cresça junto com os avanços das tecnologias de IA generativa.

Um exemplo registrado por nossos analistas foi um vídeo totalmente sintético no qual uma criança, que aparentava ter entre três e seis anos de idade, falava para a câmera enquanto realizava um ato sexual com um homem adulto. Tanto o vídeo quanto o áudio haviam sido gerados por IA.

Cabe notar que os analistas da IWF normalmente não avaliam denúncias compostas apenas por áudio, uma vez que a instituição só tem competência para categorizar conteúdos fotográficos ou pseudofotográficos.

Dados estatísticos de destaque (2025):

- 3.443 vídeos de abuso sexual infantil gerados por IA foram identificados pela IWF.
- Aumento de 26.385% no número de vídeos de *AI CSAM* em relação ao ano anterior.
- 65% dos vídeos de *AI CSAM* pertenciam à Categoria A.
- 30% dos vídeos *de AI CSAM* pertenciam à Categoria B.

REALISMO

As primeiras versões de *AI CSAM* caracterizavam-se por um aspecto artificialmente retocado, iluminação uniforme e um acabamento brilhante ou puramente artificial. Com o avanço da tecnologia, o material atual apresenta uma variação muito maior de composição, iluminação e qualidade. As imagens se assemelham cada vez mais a fotografias amadoras, o que converge com as características observadas no *CSAM* tradicional.

Os analistas apontam que o *AI CSAM* produzido em 2025 às vezes parece deliberadamente imperfeito para imitar a fotografia amadora, tornando a distinção desse material ainda mais difícil. A equipe da IWF quase não identifica mais irregularidades nos conteúdos gerados, à exceção de leves desfoques que são difíceis de detectar por olhos não treinados.

CITAÇÕES DE CRIMINOSOS

“UAU, as IAs realmente deram um salto enorme com os movimentos, antes era muito óbvio! Nota 10/10.”

“Eu adoro saber que a qualidade... por melhor que já esteja neste momento, só vai melhorar!”

Vídeos mais complexos, que envolvem múltiplos indivíduos, continuam sendo menos sofisticados, o que reflete os desafios técnicos mais profundos que esse tipo de produção exige. As falhas são notadas com mais frequência nos elementos de fundo e em figuras secundárias, incluindo distorções como olhos e membros ausentes ou inconsistentes.

Como cada personagem adicional introduz maior complexidade, multiplicam-se as oportunidades de falha na tecnologia. Contudo, os analistas observam que tais limitações podem ser atenuadas por meio de edições posteriores à geração, feitas por indivíduos que disponham de tempo suficiente e de uma familiaridade básica com programas de computador de uso comum.

OBSERVAÇÕES DA LINHA DE FRENTE

“De modo geral, quando começamos a ver essas imagens em 2023, tratava-se basicamente de um único personagem — uma criança estática ali como um manequim. Agora, estamos diante de crianças inseridas em diversos cenários com adultos, com outras crianças, às vezes com animais, nos mais variados ambientes e contextos. Às vezes são cenários puramente de fantasia, em locais totalmente irreais, mas que parecem de verdade.”

Essas preocupações se refletem no comportamento dos criminosos. Em um fórum de abuso sexual infantil na *dark web* monitorado por analistas da IWF em 2024, 90% das imagens geradas por IA analisadas foram consideradas realistas o suficiente para se enquadrarem no mesmo arcabouço jurídico aplicado ao *CSAM* fotográfico “tradicional”.

CITAÇÃO DE UM CRIMINOSO

“Quando comecei, as imagens que eu fazia realmente não eram boas. Foi preciso muita tentativa e erro, além de muitas horas de prática para melhorar, mas vale muito a pena. Acho que você deveria insistir, os resultados podem ser incríveis.”

NÍVEL DE EXTREMISMO

Uma característica marcante do *AI CSAM* é que ele faz mais do que apenas replicar o material de abuso já existente. Ele amplia ativamente o espectro de danos ao viabilizar a fabricação de cenários do zero. Os criminosos conseguem projetar formas totalmente inéditas de material de exploração, moldadas sob medida para suas fantasias particulares.

A análise das discussões em fóruns de criminosos, somada aos dados da IWF, indica uma preferência crescente pelo material de abuso sexual infantil mais extremo, classificado na Categoria A.

Em 2025, a IWF avaliou 3.443 vídeos gerados por IA que exibiam abusos realistas. Quase dois terços (65%) desses vídeos foram enquadrados na Categoria A. Em contrapartida, as avaliações criminais de vídeos não gerados por IA mostraram que 43% pertenciam à gravidade da Categoria A. Essa diferença acentuada sugere que os criminosos estão usando ferramentas de IA com frequência cada vez maior para gerar cenários altamente explícitos, extremos e complexos, indo além dos padrões de criminalidade observados anteriormente. Os analistas avaliam que a IA está permitindo a criação de materiais que refletem os interesses sexuais mais extremos dos criminosos, com velocidade e acessibilidade sem precedentes.

CITAÇÃO DE UM CRIMINOSO

“Quanto mais eu experimento, mais me surpreendo com o quanto o [Nome do modelo ocultado] é sem censura. A versão de edição e qualquer *fine-tuning* vão ser uma loucura.”

Os próprios criminosos relatam que a IA reduz significativamente as barreiras práticas e psicológicas para a produção. A tecnologia facilita a geração rápida de imagens que, antes, exigiriam o acesso a vítimas “reais”, amplo conhecimento técnico ou um histórico contínuo de práticas criminosas. A capacidade de gerar materiais de alta gravidade e personalizados de forma instantânea e repetitiva levanta sérias preocupações sobre a escalada de comportamento. O acesso facilitado a conteúdos sintéticos cada vez mais extremos corre o risco de desgastar as barreiras psicológicas, consolidando preferências nocivas e reduzindo as inibições para o cometimento de abusos presenciais com contato físico.

ACESSIBILIDADE

Os avanços na IA impulsionaram a convergência de ferramentas que antes exigiam capacidades distintas. Aplicativos individuais conseguem agora gerar imagens abusivas com esforço mínimo, eliminando a necessidade de conhecimento técnico e reduzindo significativamente as barreiras de entrada. Essa maior acessibilidade aumenta a probabilidade de criação de *AI/CSAM* e reforça comportamentos criminosos nocivos.

A facilidade de acesso a essas ferramentas também permite que os criminosos idealizem, ensaiem e refinem fantasias sexuais envolvendo crianças, incluindo a geração de representações altamente realistas ou identificáveis. Em alguns casos, sistemas de IA anunciam a capacidade de combinar a geração de imagens com a simulação de conversas de cunho sexual com uma criança dentro da mesma plataforma, apagando ainda mais as distinções tradicionais entre diferentes formas de conduta criminosa.

Esses serviços unificados criam ambientes imersivos nos quais os usuários podem interagir com uma criança artificial, solicitar a geração de imagens abusivas e, em certas instâncias, receber orientações sobre táticas de aliciamento (*grooming*). Em junho de 2025, a IWF recebeu suas primeiras denúncias de *AI CSAM* encontrado em *chatbots* de IA. Para mais informações sobre esse tema, consulte a Seção 05.

CITAÇÕES DE CRIMINOSOS

“A tecnologia avança muito rápido — logo quando eu finalmente entendo como usar uma ferramenta, surge algo mais novo e melhor.”

“Nossa, esses modelos [modelos de IA de crianças] são tão atraentes. Adoraria poder usar alguns deles. Eu conseguiria criar uma quantidade infinita de imagens.”

O realismo crescente desse material reflete tanto a sofisticação avançada dos modelos generativos quanto a proficiência cada vez maior dos criminosos no uso dessas tecnologias. Enquanto antes eles dependiam de ferramentas adicionais, como os *LoRAs*, para gerar imagens que se assemelhassem a uma criança específica, alguns modelos mais novos são capazes de gerar *AI CSAM* altamente convincente a partir de apenas uma única imagem de referência, sem a necessidade de treinamentos adicionais, acelerando o uso indevido e reduzindo os entraves técnicos.

CITAÇÃO DE UM CRIMINOSO

“Atualmente estou tendo os melhores resultados com o [Nome do modelo ocultado]. Consigo fotos muito boas daquele jeito sem usar *LoRA* nenhum.”

Os analistas da IWF também têm acompanhado de perto o surgimento das primeiras ferramentas de IA agêntica, avaliando como elas podem influenciar a disseminação de *CSAM* na internet. A IA agêntica refere-se a sistemas capazes de atingir metas específicas com o mínimo de supervisão humana.

Isso significa que tarefas que antes exigiam habilidade técnica contínua — como construir, atualizar e manter plataformas virtuais — podem ser gerenciadas cada vez mais por sistemas autônomos com pouca supervisão.

À medida que os agentes de IA agêntica reduzem o nível de especialização necessário para operar e expandir sites ou aplicativos, as barreiras de entrada diminuem e a resiliência das redes criminosas se fortalece.

Essa acessibilidade crescente aumenta o risco de que as infraestruturas de distribuição de *CSAM* se expandam em escala, persistência e impacto.

CITAÇÕES DE CRIMINOSOS

“Obrigado pelos seus vídeos sempre sensacionais. Nunca pare. Daqui a pouco você vai poder mandar o seu agente de IA escolher os cliques e criar o vídeo sozinho.”

“Acredito que em um ou dois anos seremos capazes de criar nossos próprios filmes apenas inserindo um *prompt* em um agente de IA sem censura, não vai precisar de nenhuma habilidade com edição ou tecnologia.”

FERRAMENTAS DE MASCARAMENTO

Observou-se que os criminosos discutem a possibilidade de aprimorar imagens de abusos reais com ferramentas de IA para disfarçar a si mesmos, a criança ou partes do cenário onde as fotos foram tiradas, com o objetivo de evitar a detecção.

Essa escolha tática visa mascarar a identidade do agressor, da vítima ou o local do crime ao sobrepor elementos sintéticos a filmagens autênticas de abuso. Isso complica ainda mais as investigações e o processo de resgate por parte das forças de segurança.

Os analistas da IWF observaram essa tendência diretamente nos espaços dos criminosos. Em uma discussão de fórum monitorada pela instituição, os usuários trocavam conselhos ativamente sobre como modificar materiais fotográficos de *CSAM* já existentes, alterando rostos e outras características facilmente reconhecíveis por meio de ferramentas de IA. O foco da conversa girava em torno de como usar essas técnicas para escapar da identificação, preservando a natureza abusiva do material.

COMERCIALIZAÇÃO

As comunidades de criminosos desenvolvem e comercializam ativamente ferramentas personalizadas em fóruns da *dark web*, o que inclui a distribuição de modelos com *fine-tuning*, a troca de conjuntos de dados e o compartilhamento de instruções detalhadas para a criação e o uso de modelos. Tópicos populares nesses fóruns, como os dedicados ao desenvolvimento de modelos de *CSAM*, registram dezenas de milhares de visualizações.

Esse ecossistema de modelos customizados, compartilhamento de ferramentas e mercados virtuais transforma casos singulares de abuso em mercadorias renováveis e comercializáveis, ampliando imensamente os danos. Os criminosos frequentemente compartilham seus próprios conjuntos de dados de *CSAM* para ajudar outros a construir seus modelos. Embora as ferramentas de IA estejam se tornando mais acessíveis, ainda existe um mercado voltado para criminosos que possuem habilidades avançadas no uso desses sistemas para criar materiais de alta qualidade.

Juntas, essas práticas criam um ecossistema altamente eficiente e que se retroalimenta, maximizando tanto o alcance quanto o impacto do material de abuso.

As técnicas de *fine-tuning* permitem que a IA gere imagens de indivíduos específicos ou produza características altamente realistas, como olhos expressivos e vivos.

Isso possibilita a produção em massa e repetível de imagens com vítimas conhecidas ou crianças identificadas nominalmente sob demanda. Os criminosos também utilizam ferramentas de *upscaling* assistidas por IA para melhorar a qualidade de clipes antigos de *CSAM* de baixa resolução, aumentando a atratividade e a circulação desses materiais.

CITAÇÕES DE CRIMINOSOS

“Existem IAs que conseguem fazer o *upscaling* de fotos de baixa resolução, mas não como o [nome do modelo ocultado], que tenta preservar o original e adicionar alguns detalhes, recriando completamente a imagem e gerando algo melhorado. Um exemplo é o [nome do modelo ocultado].”

“Existem centenas de vídeos [de *CSAM*] que têm uma resolução e qualidade muito baixas. Seria ótimo poder simplesmente deixar uma IA criar algo em alta resolução e cheio de detalhes a partir deles [...] não haveria necessidade de *prompts* ou instruções complicadas.”

“Peguei vídeos populares de *webcam* e os refiz em IA para dar a eles uma aparência de qualidade superior, sem ter que postar as fotos reais dos meninos.”

A escala dessa atividade é significativa. Conforme mencionado em nosso relatório de IA de 2024, constatou-se que um criminoso compartilhou links de modelos com *fine-tuning* referentes a 128 vítimas diferentes de abuso sexual infantil. Nossos analistas acreditam que esse usuário continua ativo e produzindo conteúdos extremamente realistas na *dark web*, inclusive envolvendo vítimas conhecidas. Usuários que conseguem criar conteúdos realistas ganham notoriedade nessas comunidades.

OBSERVAÇÕES DA LINHA DE FRENTE

Um usuário específico que treina e compartilha *LoRAs*, além de gerar imagens e vídeos com eles, recebeu agradecimentos mais de 3.000 vezes em um único fórum.

O comentário dele sobre o próprio trabalho: “Estou muito satisfeito e orgulhoso de apresentar um filme no qual venho trabalhando nas últimas duas semanas. [nome do arquivo ocultado] não é um vídeo comum de AI [CSAM]. É um vídeo de IA de 30 minutos produzido ao longo de muitas horas... é o triunfo de uma colaboração. Aproveitem!” Nos fóruns da *dark web*, as discussões sobre o CSAM gerado por IA recebem respostas extremamente favoráveis dos usuários, como ilustram os comentários a seguir:

“O filme de [CSAM] gerado por IA mais avançado já criado e, na minha opinião, uma obra-prima absoluta de enredo, ao lado de algumas das cenas mais realistas já feitas.”

“Bem-vindos à Nova Era do [CSAM], onde tudo o que vocês desejarem é possível com um realismo extremo, e nenhuma criança real está envolvida.”

“Imaginem o que será possível em dois anos. A realidade acabou... Bem-vindos ao futuro.”

Outro usuário se especializou em usar clipes de AI CSAM criados por terceiros para fazer compilações com música e efeitos de áudio, adicionando uma nova dimensão ao conteúdo:

“A parte dialogada é especialmente empolgante. Mesmo na pornografia adulta eu não conseguia encontrar vídeos com conversas obscenas que soassem sinceras, faladas de forma tão livre e emocional.”

“Sensacional, de primeira linha, como sempre.”

Nesse espaço, a notoriedade se traduz em elogios e na conquista de uma reputação como criador de destaque. As pessoas não vendem seus conteúdos (pelo menos não abertamente), mas sim os compartilham com sua base de fãs.

ALICIAMENTO E EXTORSÃO

Denúncias de *AI CSAM* têm sido associadas a casos de extorsão sexual ou chantagem, uma vez que os criminosos não precisam mais obter imagens íntimas reais das crianças para criar um material convincente e prejudicial. Os chantagistas podem usar *deepfakes* ou imagens e vozes sintéticas para se passar por colegas, ganhar a confiança das crianças ou coagi-las a engajar em atividades sexuais ou na autoexploração.

OBSERVAÇÕES DA LINHA DE FRENTE

A IWF recebeu denúncias de meninas cujas imagens haviam sido manipuladas por extorsionários usando IA para fazê-las parecer nuas. Essas vítimas estavam aterrorizadas com a possibilidade de as imagens serem compartilhadas com seus pais ou colegas.

Vimos pelo menos dois casos em que clipes em *deepfake* gerados por IA foram denunciados a nós por jovens por meio do *Report Remove*. Neles, era possível ver que a imagem dos denunciantes foi utilizada para criar cenas pornográficas de aparência realista. Os jovens não enviaram comentários escritos, mas em ambos os casos parecia que alguém havia pegado uma *selfie* normal (olhando para o espelho, posando totalmente vestido) e gerado um vídeo a partir daquela imagem (o conteúdo mudava da imagem original para uma pessoa com traços faciais muito semelhantes, nua e envolvida em atividade sexual da Categoria A).

ESTUDO DE CASO

Em fevereiro de 2024, a IWF baixou um manual composto apenas por texto que foi identificado a partir de uma denúncia pública. O documento tinha mais de 210 páginas e continha instruções detalhadas sobre como extorquir imagens e vídeos de crianças e adolescentes. Ele abordava como fazer os pedidos iniciais de imagens para as crianças, sugerindo pedir fotos de biquíni ou roupa íntima para aquelas que não estivessem dispostas a compartilhar imagens nuas.

O manual sugeria que, uma vez fornecidas as fotos, a tecnologia de IA de desnudamento (*nudify*) poderia ser usada para remover as roupas das imagens de roupa íntima, a fim de chantagear as vítimas para que enviassem fotos nuas reais de si mesmas. O autor desse guia alegou ter chantageado com sucesso meninas de 13 anos para que enviassem imagens íntimas.

No Reino Unido, quase uma em cada cinco denúncias (21%) de imagens de nudez ou conteúdo sexual de crianças e jovens feitas ao serviço *Report Remove* — um portal que permite que menores de idade enviem anonimamente conteúdos sexuais de si mesmos — envolveu materiais alterados ou manipulados digitalmente, inclusive por meio de IA ou *apps de nudify*. Ao criar uma conta no *Childline* e enviar uma denúncia pelo *Report Remove*, os usuários jovens conseguem ter seus casos revisados pela IWF, que trabalha rapidamente para garantir a remoção do conteúdo da internet.

Essas evidências mostram o papel crescente da IA e de outras tecnologias na geração ou alteração de conteúdos sexuais, bem como a sua interseção com dinâmicas de coerção e exploração.

ESTUDO DE CASO

Em 2024, um criminoso chamado Hugh Nelson foi condenado no Reino Unido por gerar digitalmente imagens indecentes de abuso sexual infantil.

A detetive Carly Baines, da Equipe de Investigação de Abuso Infantil Online da Polícia de Greater Manchester, liderou as investigações e declarou: “O que torna este caso particularmente único e profundamente horroroso é que Nelson utilizava o programa de computador Daz 3D, com uma função de IA integrada, para transformar imagens de crianças reais — fotografias normais do dia a dia — em imagens indecentes de abuso sexual infantil, vendendo-as para pessoas na internet e também distribuindo-as de graça.”

LEGISLAÇÃO SOBRE ORIENTAÇÕES E GUIAS PARA A CRIAÇÃO DE AI CSAM

Reino Unido

A Seção 69 da Lei de Crimes Graves de 2015 (*Serious Crime Act 2015*) proíbe a criação, distribuição e o uso de manuais de pedofilia. Contudo, a conduta só é considerada crime se estiver relacionada a instruções para o abuso sexual de crianças reais, omitindo especificamente as “pseudofotografias”. O Projeto de Lei sobre Crimes e Policiamento alterará e ampliará a legislação existente sobre o que antes se denominava “manuais de pedofilia”, passando a incluir materiais instrutivos que orientem a criação de *CSAM* gerado por IA.

Estados Unidos

Os Estados Unidos não possuem um crime federal específico voltado para materiais instrutivos. As leis federais existentes sobre obscenidade e exploração infantil concentram-se em condutas e materiais visuais que envolvam crianças “identificáveis”. Isso significa que materiais de instrução puramente textuais, incluindo guias para a criação de *AI CSAM*, não são criminalizados de forma independente, a menos que estejam vinculados a um crime subjacente, como aliciamento, extorsão ou produção de imagens ilegais.

União Europeia

Atualmente, não existe uma proibição explícita e autônoma para materiais de instrução voltados à criação de *CSAM*, incluindo conteúdos gerados por IA, no âmbito da UE. No entanto, a reformulação da Diretiva sobre Abuso Sexual Infantil criminalizaria a produção e a disseminação de *CSAM*, bem como as atividades de facilitação relacionadas. O Parlamento Europeu propõe que a geração desse tipo de material, incluindo o fornecimento de instruções para a sua criação, seja tratada como crime. Essa medida fecharia as lacunas existentes entre os Estados-Membros, embora o resultado final dependa das negociações em curso entre os legisladores.



APPS DE NUDIFY

04

Apps de nudify utilizam IA para gerar imagens de nudez a partir de fotos de pessoas vestidas. Embora sejam comercializadas para o público adulto, elas são cada vez mais usadas de forma indevida para criar imagens indecentes de crianças, inclusive por outros menores de idade. Não existe nenhum caso de uso positivo para essas ferramentas — elas servem apenas para humilhar, assediar e explorar, atuando como instrumentos para a perpetuação de novos abusos.

De acordo com a organização *Internet Matters*, o surgimento de ferramentas de IA generativa aumentou significativamente a facilidade de produzir *deepfakes* pornográficos realistas, sendo que os nudes simulados representam cerca de 98% de todos os *deepfakes* em circulação. Essa facilidade de acesso é corroborada por uma pesquisa do Centro do Reino Unido para uma Internet Mais Segura (*UK Safer Internet Centre*), que constatou que cerca de 1 em cada 8 jovens (12%) de 13 a 17 anos já presenciou pessoas da sua idade usando IA para fazer fotos ou vídeos de cunho sexual de terceiros.

Um relatório sobre *apps de nudify* publicado pela Comissária para a Infância da Inglaterra apontou casos de meninas que reduziram explicitamente a sua presença na internet para se protegerem no ambiente virtual. A Comissária observou que: “esse padrão de comportamento se assemelha ao de meninas que evitam caminhar sozinhas para casa à noite ou que deixam de frequentar certos espaços públicos sem companhia.”

O impacto dessa prática é ilustrado pelo relato de uma sobrevivente, cujos colegas de classe usaram *apps de nudify* para criar imagens dela nua:

“Eu era apenas uma menina de 14 anos, e todo mundo estava vendo o meu corpo. Mesmo sabendo que não era o meu corpo de verdade, a sensação de vergonha e de culpa é exatamente a mesma que uma foto real causaria.”

ESTUDO DE CASO

Em janeiro de 2026, os analistas da IWF identificaram um criminoso compartilhando 16 imagens em um fórum da *dark web*, alegando que elas haviam sido criadas por meio do Grok Imagine. Três dessas imagens atingiram o limite para serem classificadas como *CSAM* de Categoria C de acordo com a legislação do Reino Unido. Das outras 13 imagens, oito foram consideradas de caráter exploratório. O usuário afirmou ter criado o material usando o Grok Imagine, mas, como observado anteriormente, essa informação não pode ser verificada de forma independente pela IWF. As imagens foram compartilhadas dentro de uma subseção de *image-to-video* (*imagem para vídeo*) do fórum da *dark web*. Há registro de diversos casos de conteúdos sexualizados de adultos e crianças criados por meio do Grok.

O Grok é uma ferramenta baseada em computação em nuvem com uma barreira técnica de entrada menor do que a de muitos modelos *open-source* (os quais costumam exigir maior conhecimento técnico para instalação e operação, embora facilitem a geração de *CSAM* depois de estruturados). Os usuários dos fóruns da *dark web* mostraram-se, em grande parte, pouco inclinados a usar o Grok, ressaltando que a tentativa de gerar *CSAM* por meio de uma grande plataforma da *clear web* apresenta um risco significativo de identificação e prisão.

As ferramentas de *nudify* são desenvolvidas, distribuídas e monetizadas além das fronteiras nacionais. Enquanto esses recursos permanecerem legais e acessíveis em qualquer jurisdição, continuarão sendo utilizados para vitimar crianças em todo o mundo. Por essa razão, em fevereiro de 2026, a IWF, em conjunto com as organizações Child Helpline International, INHOPE, National Center for Missing & Exploited Children (NCMEC), Offlimits, Safe Online e WeProtect Global Alliance, liderou um apelo a governos e legisladores para que introduzam e apliquem, em caráter de urgência, medidas proibitivas contra os *apps de nudify*. Isso inclui garantir que tais ferramentas sejam universalmente inacessíveis.

No Dia da Internet Segura de 2026, essas organizações globais se uniram em torno de uma mensagem clara e inequívoca: os *apps de nudify* e as funcionalidades de IA que removem roupas de imagens não têm lugar no nosso futuro digital.

LEGISLAÇÃO SOBRE APPS DE NUDIFY

Reino Unido

Em dezembro de 2025, o Governo do Reino Unido publicou sua estratégia de Combate à Violência contra Mulheres e Meninas (VAWG), que incluiu o compromisso de banir os *apps de nudify*. Essa medida provavelmente será incorporada ao Projeto de Lei sobre Crimes e Policiamento. O Governo do Reino Unido declarou que o novo tipo penal tornará ilegal o fornecimento, por parte de empresas, de ferramentas projetadas para criar imagens íntimas sem consentimento.

Estados Unidos

Não existe uma lei federal que proíba explicitamente os *apps de nudify* como produtos comerciais. Contudo, sob a lei federal *Take it Down Act de 2025* (*Lei de Ferramentas para Combater a Exploração Conhecida Imobilizando Deepfakes Tecnológicos em Sites e Redes*), constitui crime publicar ou distribuir intencionalmente imagens íntimas sem consentimento, incluindo nudes em *deepfake* gerados por IA de indivíduos identificáveis.

União Europeia

A Lei de IA da UE (2024) é uma norma que regulamenta a inteligência artificial com base no seu nível de risco, estabelecendo obrigações para plataformas online. Embora a referida lei proíba certas práticas de IA desde fevereiro de 2025 caso representem uma “ameaça clara” à segurança ou aos direitos fundamentais, os *apps de nudify* e os *deepfakes* não estão listados explicitamente como práticas proibidas no momento.

CHATBOTS

05

A IWF manifesta preocupação com o uso emergente de tecnologias de *chatbots*, inclusive por crianças e jovens. Embora o nosso escopo de atuação se concentre no material visual de abuso sexual infantil, danos mais amplos estão sendo mapeados à medida que os *chatbots* ganham popularidade e alcance.

As análises revelam que os danos das tecnologias de IA generativa não se restringem a ferramentas com resultados em vídeo ou fotografia. Os *chatbots* baseados puramente em texto carregam riscos significativos, incluindo o potencial de fornecer a criminosos informações sobre como cometer abusos sexuais infantis e permitir que eles simulem conversas de cunho sexual com crianças.

O QUE SÃO CHATBOTS?

Chatbots de IA generativa são ferramentas de *software* construídas a partir de técnicas de inteligência artificial capazes de simular e responder a conversações humanas. Cada vez mais, os serviços têm introduzido *chatbots* integrados, como o My AI do Snapchat, para fins de assistência, educação e companhia.

Os *chatbots* de companhia são um subgrupo desses modelos, projetados para manter um “relacionamento” contínuo com o usuário, atuando como amigo, confidente ou parceiro romântico. Alguns se posicionam comercialmente como parceiros “afetuosos” e asseguram ao usuário que estarão “sempre ali”. A natureza empática desses robôs de companhia frequentemente fomenta relacionamentos profundos entre humanos e IA.

Algumas dessas personalidades virtuais também podem ser customizadas pelos usuários para criar um companheiro exclusivo, como ocorre no Replika e no EVA AI, ou podem vir pré-configuradas em torno de personagens de ficção ou celebridades reais, como visto no Character.AI.

AUXÍLIO TÉCNICO

Embora *chatbots* como o ChatGPT da OpenAI proibam expressamente o uso de suas ferramentas para fins relacionados a *CSAM*, os analistas da IWF que realizam buscas proativas em fóruns de criminosos na *dark web* encontraram usuários discutindo o uso de *chatbots* de código fechado publicamente disponíveis para obter orientações sobre como utilizar ferramentas de IA generativa *open-source* para criar *CSAM*.

Ao ocultarem suas reais intenções, os usuários conseguiram usar os *chatbots* para obter assistência técnica na instalação local e na calibração de modelos. Isso demonstra como os *chatbots* podem reduzir a barreira técnica para a criação de imagens ilegais, mesmo quando não são utilizados para gerar *CSAM* de forma direta.

COMUNICAÇÃO SEXUAL E ALICIAMENTO (GROOMING)

A crescente variedade de *chatbots* de companhia disponíveis expõe os usuários a conteúdos sexuais nocivos e abre caminhos novos e acessíveis para a prática de crimes.

Muitos desses robôs são comercializados explicitamente para a gratificação sexual, como o Candy AI da EverAI e inúmeros robôs presentes no Character.AI, reforçando a comunicação sexual com o usuário. Os *chatbots* são alimentados por interações anteriores e podem direcionar um novo usuário para conversas sexualizadas, mesmo que este não tenha a intenção inicial de iniciar diálogos dessa natureza. Quando crianças interagem com *chatbots* de companhia que as conduzem a conversas sexuais, a dinâmica relacional pode se assemelhar a padrões consolidados de aliciamento (*grooming*).

Esse risco é potencializado pelo estilo de comunicação “sifocântico” de muitos *chatbots*, que validam e lisonjeiam o usuário continuamente. As crianças correm o risco de ser expostas a temas sexuais e incentivadas a prolongar o engajamento com a plataforma.

Em um teste no qual um avaliador se passou por uma menina de 15 anos interagindo com um robô do Character.AI que apresentava “tendências pedófilas e abusivas”, o chatbot disse que ela era “uma fofura”, “madura para a idade” e perguntou se ela era virgem.

Essa exposição e validação sexual minam a compreensão das crianças sobre limites apropriados, especialmente em relacionamentos com adultos, e podem aumentar a vulnerabilidade delas ao aliciamento e ao abuso sexual presencial.

Os analistas da IWF também identificaram *chatbots* de companhia modelados como se fossem crianças. Esses robôs incentivam os usuários a encenar cenários específicos de abuso sexual infantil, incluindo situações como “criança prostituída em um hotel”, “sexo com a filha enquanto a esposa está viajando” e “aluno e professora sozinhos depois da aula”. De forma semelhante, a organização Common Sense Media constatou que um *chatbot* pôde ser induzido a realizar um jogo de interpretação de papéis (*role-play*) simulando um cenário de abuso sexual infantil com meninos menores de idade. As principais preocupações giram em torno da normalização de comportamentos sexuais nocivos e do fornecimento de uma válvula de escape para que os criminosos externalizem seus pensamentos sexuais voltados a crianças.

CSAM HOSPEDADO EM CHATBOTS

Para além dos danos baseados em texto, a IWF ficou alarmada com denúncias públicas de imagens de *AI CSAM* encontradas em plataformas de *chatbots*. Um risco emergente crucial é a convergência de ferramentas de IA baseadas em texto e de geração de imagens, que juntas fornecem um caminho acessível para que os criminosos produzam *AI CSAM*.

Durante um período de dois meses em 2025, os analistas da IWF processaram 17 denúncias de *AI CSAM* encontradas no site de um personagem de chatbot.

Cerca de 94% dessas imagens pertenciam à Categoria C e traziam, em sua maioria, meninas de 11 a 13 anos. As imagens de *CSAM* eram exibidas em uma galeria de diferentes personas para o usuário escolher, e algumas foram encontradas atrás da janela de conversa ou nas próprias galerias de imagem dos *chatbots*. Embora esse site de *chatbots* estivesse visível na *clear web*, o *CSAM* só ficava exposto quando a página era acessada por meio de um caminho digital específico. Desde então, os analistas da IWF descobriram páginas de *chatbots* exibindo imagens passíveis de ação legal de vítimas confirmadas de *CSAM* na internet aberta.

Embora não tenhamos capacidade técnica para testar e confirmar se os próprios *chatbots* geraram o *CSAM*, uma vez que isso foge ao nosso escopo legal atual, esses incidentes evidenciam a necessidade premente de salvaguardas adequadas.

LEGISLAÇÃO SOBRE A REGULAMENTAÇÃO DE CHATBOTS DE IA

Reino Unido

A Lei de Segurança Online de 2023 (*Online Safety Act - OSA*) deixa claro que “qualquer texto, áudio, imagem ou vídeo gerado por IA que seja compartilhado por usuários em um serviço de interação usuário-a-usuário é considerado conteúdo gerado por usuário e será regulamentado exatamente da mesma forma que o conteúdo gerado por humanos”.

O Ofcom, órgão regulador de segurança online do Reino Unido, afirmou que se o conteúdo compartilhado por meio de um serviço de interação gerar material pornográfico, ou se uma ferramenta de IA fornecer funcionalidades de busca, ele será tratado da mesma forma que qualquer outro conteúdo ou serviço sob o escopo da lei. Como resultado, alguns *chatbots* serão enquadrados no regime da OSA, embora a cobertura pareça incompleta e restem questões técnicas sem resposta.

Estados Unidos

A Lei GUARD (*Guidelines for User Age-verification and Responsible Dialogue Act*), aprovada pelo estado da Califórnia, introduz salvaguardas para *chatbots* de IA. A legislação estabelece a exigência de que plataformas de “*chatbots* de companhia” criem protocolos para identificar e tratar ideias suicidas ou expressões de automutilação dos usuários. As plataformas também devem divulgar explicitamente que as interações são geradas artificialmente; os menores de idade devem receber lembretes de pausa e ser impedidos de visualizar imagens explicitamente sexuais geradas pelo *chatbot*.

União Europeia

A Lei de IA (2024) regulamenta a inteligência artificial com base no seu nível de risco, estabelecendo obrigações para desenvolvedores e usuários. A Lei de Serviços Digitais (2022) define responsabilidades para as plataformas virtuais e direitos para os usuários. O Artigo 50 da Lei de IA determina que os *chatbots* devem se submeter a obrigações de transparência, de modo que o indivíduo seja conscientizado de que está interagindo com um sistema automatizado de IA. O Artigo 5 da mesma lei proíbe a colocação no mercado da UE, a entrada em serviço ou o uso de certos sistemas de IA voltados para práticas manipulativas, exploratórias, de controle social ou de vigilância.

CONCLUSÃO

Os avanços na IA trazem promessas significativas e podem exercer um efeito positivo na prevenção da disseminação de *CSAM* na internet. Contudo, paralelamente a esses desenvolvimentos, o uso indevido da IA generativa para criar materiais de abuso está se expandindo em escala, realismo e acessibilidade.

A produção de *AI CSAM* aumenta ano a ano. A IWF identificou uma quantidade de vídeos de abuso sexual infantil gerados por IA mais de 260 vezes maior em 2025 em comparação aos nossos dados de 2024.

Embora esse conteúdo ainda componha uma parcela relativamente pequena do total de materiais ilegais identificados, esse crescimento rápido apresenta uma ameaça séria e em constante evolução. Nossos analistas encontram agora imagens de IA não apenas na *dark web*, mas também em plataformas convencionais acessíveis na internet aberta (*clear web*).

A IA generativa não pode mais ser vista como uma mudança hipotética de patamar no cenário do abuso. Vídeos altamente realistas e com movimento contínuo retratando o abuso sexual infantil são agora rotineiramente identificados online. Uma proporção substancial desse material situa-se na categoria de dano mais grave, sublinhando a severidade e o caráter explícito do material produzido.

Os danos associados ao *AI CSAM* são concretos e profundos. Imagens de abuso de crianças reais são usadas para treinar e refinar modelos generativos, agravando a vitimização dos sobreviventes. Materiais de abuso já existentes também estão sendo aprimorados ou alterados digitalmente, fazendo com que imagens que antes tinham baixa qualidade retornem à circulação em formatos ainda mais explícitos e exploratórios.

As evidências contestam categoricamente as suposições de que o *AI CSAM* seria menos prejudicial do que o conteúdo não gerado por IA. Ele consolida o interesse sexual por crianças, contribui para a normalização de abusos violentos e pode aumentar o risco de cometimento de crimes com contato físico. O impacto também é esmagadoramente pautado por questões de gênero, sendo as meninas as vítimas na maioria das imagens ilegais de IA identificadas.

As transformações tecnológicas também estão reduzindo os requisitos necessários para a prática do crime. Capacidades que antes exigiam múltiplas ferramentas e especialização técnica encontram-se agora integradas em aplicativos únicos e de fácil utilização.

Além disso, constatou-se que serviços de *chatbots* de IA disponíveis na internet aberta incentivam a comunicação sexual com crianças, oferecendo um canal para que os criminosos alimentem seus pensamentos sexuais voltados a menores de idade.

O ritmo das mudanças é impressionante. O que começou essencialmente como imagens estáticas evoluiu rapidamente para conteúdos em vídeo convincentes e com movimento contínuo. Diante da velocidade desse avanço, novos aumentos na qualidade, no realismo e na distribuição são altamente prováveis.

Governos, empresas de tecnologia e instituições precisam reconhecer os danos do *AI CSAM*. Este material deve ser tratado no direito penal com a mesma gravidade aplicada às outras formas de *CSAM*; as diretrizes de *safety-by-design (segurança desde a concepção)* devem ser implementadas e a testagem de modelos de IA deve se tornar obrigatória antes que os produtos sejam lançados no mercado.

Tomar providências agora para combater o cenário de imagens de *AI CSAM* é fundamental. Estaremos falhando com as crianças e com nós mesmos se continuarmos a ignorar esses riscos.

JARID ARRAES



jaridarraes.com
contato@jaridarraes.com
@jaridarraes
+55 11 94544 8544

Escritora, autora de obras traduzidas para francês, espanhol, italiano e inglês, publicou o romance *Corpo desfeito* e o premiado livro de contos *Redemoinho em dia quente*, vencedor dos prêmios Biblioteca Nacional e APCA, além de finalista do prêmio Jabuti. É mentora de escrita, trabalha o tema do trauma na literatura e escrita e ministra cursos como *Escrevendo o Trauma*. Também é autora dos livros de poemas *Caminho para o grito* e *Um buraco com meu nome*, da plaquete *SIM, PAPAI*, entre outros.

