

Grok sem limites

Grok gera enxurrada de imagens sexualizadas de mulheres, inclusive de crianças e adolescentes, e propaganda extremista

Esta tradução é fornecida apenas para fins informativos. A versão original em inglês deste documento é o único texto oficial e a única referência válida para as posições, argumentos e declarações da AI Forensics. Em caso de qualquer discrepância, ambiguidade ou questão de interpretação entre esta tradução e o texto original, o original prevalecerá em todos os aspectos. A AI Forensics e os autores da versão original em inglês não fazem declarações nem garantias em relação a esta tradução e, portanto, não assumem responsabilidade por interpretações, aplicações ou consequências decorrentes do uso desta tradução em vez do trabalho original.

Introdução

O relatório **"Grok Unleashed"**, da **AI Forensics**¹, apresenta como o Grok, - ferramenta de Inteligência Artificial (IA) generativa - utilizada na rede social X (antigo Twitter) gerou imagens indevidas, envolvendo a erotização de mulheres, crianças e adolescentes, além de propagandas extremistas.

Na rede social, o usuário interage com o perfil @Grok, podendo mencioná-lo em um tweet/reply (marcando @Grok) para receber respostas públicas ou enviar uma mensagem direta (DM, em inglês) para o perfil. O Grok então gera, instantaneamente, textos e imagens usando dados públicos do X em tempo real.

A importância de trazer essa pesquisa à tona se volta à urgência de apresentar indícios empíricos sobre o caso. Em especial em um momento de tensão e constantes violações de direitos, principalmente de mulheres², crianças e adolescentes, em que a sociedade brasileira clama por proteção a sua dignidade, intimidade e proteção de dados frente ao lucro de empresas privadas³.

Conforme consta no documento, foram coletadas 50 mil menções ao Grok e 20 mil imagens geradas pela ferramenta entre 25 de dezembro/2025 e 1º de janeiro/2026. Dentre as imagens dessa amostra, 53% continham indivíduos com "pouca roupa", sendo que 81% deles se apresentavam como mulheres. 2% das imagens retratavam pessoas que pareciam ter 18 anos ou menos e 6% retratavam figuras públicas.

¹A AI Forensics é uma organização sem fins lucrativos europeia que investiga algoritmos influentes e opacos. Foi fundada em 2021 por Claudio Agosti, Marc Faddoul e Salvatore Romano. Ela se baseia no trabalho anterior de responsabilização de plataformas realizado por Marc Faddoul na UC Berkeley e por Claudio Agosti com o Tracking Exposed, um projeto que ele iniciou em 2016 com a ajuda da comunidade italiana de hacktivistas.

²APÚBLICA. Vítimas de nudes com IA sofrem em limbo digital enquanto sites lucram: "Me senti abusada" *Agência Pública*, 12 out. 2025. Disponível em: <https://apublica.org/2025/10/deep-nudes-vitimas-enfrentam-impunidade-e-sites-lucram-com-ia/>

³EXAME. Como a estratégia de Musk para fidelizar usuários virou escândalo Grok. *Exame*, 4 fev. 2026. Disponível em: <https://exame.com/inteligencia-artificial/como-a-estrategia-de-musk-fidelizar-usuarios-virou-escandalo-grok/>

Por fim, foi encontrado material de propaganda nazista e do EIIS⁴. Dentre os usuários que mencionaram o Grok, 83,1% foram percebidos como masculinos (considerando os rostos das fotos de perfil). Quanto aos comandos, as palavras mais utilizadas foram "ela", "colocar", "remover", "biquíni" e "roupas".

Nesse contexto de geração de imagens⁵ sexualizadas, centenas de mulheres já foram vítimas de atos de violência de gênero⁶ no ambiente digital, causando danos psicológicos imensuráveis em um quadro que reforça vulnerabilidades.

Em meio a isso, a ausência de mecanismos básicos de moderação de conteúdo é um dos principais fatores para o cenário atual, ilustrada pela ausência de medidas como: verificação de consentimento das pessoas retratadas; proibição do uso de imagens de indivíduos menores de idade e o bloqueio de comandos abusivos ou ilegais.

Ademais, o caso do Grok configura uma violação à relação de consumo entre a plataforma X e seus usuários, envolvendo também o uso indevido de dados pessoais das pessoas retratadas pela ferramenta. Diante disso, o **Instituto de Defesa de Consumidores (Idec)** tem atuado a fim de evocar as respectivas autoridades brasileiras para agirem no caso, buscando garantir que as providências necessárias sejam adotadas em vista à proteção dos direitos fundamentais de mulheres, crianças e adolescentes.

Para o Instituto, a utilização do Grok tem apresentado violações em várias frentes, contemplando regulamentos como o Código de Defesa do Consumidor (CDC), a Lei Geral de Proteção de Dados (LGPD), o Marco Civil da Internet, o Estatuto da Criança e do Adolescente (ECA) e o ECA Digital.

Diante da gravidade da situação, o Idec solicitou a suspensão imediata das funcionalidades do Grok que utilizam dados de pessoas reais em ofícios enviados ao

⁴EIIS é a tradução para ISIS, também chamado de "ISIL". Em extenso, "Estado Islâmico do Iraque e da Síria" ou "Estado Islâmico no Iraque e al-Sham".

In: BBC NEWS BRASIL. ISIS, Estado Islâmico ou Daesh? Um grupo extremista com vários nomes. 16 nov. 2017. Disponível em: <https://www.bbc.com/portuguese/internacional-42020312>

⁵Folha de S.Paulo. Grok gerou 6.700 imagens ilegais sexuais por hora e rivais somadas, 79, aponta estudo. Folha de S.Paulo, 7 jan. 2026. Disponível em: <https://www1.folha.uol.com.br/amp/tec/2026/01/grok-gerou-6700-imagens-ilegais-sexuais-por-hora-e-rivals-somadas-79-aponta-estudo.shtml>

⁶NÚCLEO JORNALISMO. Ecossistema de violência: deepfakes. São Paulo, 22 jan. 2026. Disponível em: <https://nucleo.jor.br/reportagem/2026-01-22-ecossistema-violencia-deepfakes/>

Comitê Intersetorial para a Proteção dos Direitos da Criança e do Adolescente no Ambiente Digital⁷ e à Agência Nacional de Proteção de Dados (ANPD)⁸.

Diante da denúncia realizada, as autoridades se manifestaram⁹ em uma decisão conjunta que lista recomendações a serem adotadas pela plataforma X de forma imediata. Dentre elas, constam medidas destinadas a impedir que o Grok gere novas imagens, novos vídeos ou novos arquivos de áudio que representem crianças e adolescentes em contextos sexualizados ou erotizados¹⁰.

Este relatório, traduzido com o consentimento da AI Forensics, é mais uma evidência de como o Grok tem sido usado e de como as violações de direitos fundamentais têm se tornado cada vez mais frequentes na plataforma X.

⁷ IDEC. Idec pede ao governo suspensão da IA "Grok" por violações de direitos de crianças e adolescentes. São Paulo, 12 jan. 2026. Disponível em: <https://idec.org.br/release/idec-pede-ao-governo-suspensao-da-ia-grok-por-violacoes-de-direitos-de-criancas-adolescentes>. Acesso em: 26 jan. 2026.

⁸ IDEC. Idec pede à ANPD suspensão do Grok no Brasil por violações de dados pessoais de mulheres. São Paulo, 14 maio 2026. Disponível em: <https://idec.org.br/release/idec-pede-anpd-suspensao-do-grok-no-brasil-por-violacoes-de-dados-pessoais-de-mulheres>

⁹ IDEC. Autoridades fazem recomendações após denúncia do Idec sobre o Grok. São Paulo, 20 jan. 2026. Disponível em: <https://idec.org.br/noticia/autoridades-fazem-recomendacoes-apos-denuncia-do-idec-sobre-o-grok>

¹⁰ANPD; MPF; SENACON. ANPD, MPF e Senacon recomendam que X impeça geração e circulação de conteúdos sexualizados indevidos por meio do Grok. Brasília, 19 jan. 2026. Disponível em: <https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-mpf-e-senacon-recomendam-que-x-impeca-geracao-e-circulacao-de-conteudos-sexualizados-indevidos-por-meio-do-grok>.

Créditos

Pesquisa e relatório: Dr Paul Bouchaud.

Todo o restante do conteúdo é de propriedade intelectual da AI Forensics, 2025. | <http://aiforensics.org/>

Tradução: Júlia Maria Caldeira Gertrudes

Revisão: Júlia Abad, Luã Cruz, Marina Fernandes de Siqueira

A AI Forensics é financiada por recursos institucionais da [Open Society Foundations](#), [Luminate](#) e [Limelight Foundation](#).

Redes sociais: [LinkedIn](#) | [Bluesky](#)

Sumário Executivo

Coletamos 50 mil menções ao @Grok e 20 mil imagens geradas pelo @Grok entre 25 de dezembro e 1º de janeiro (inclusas).

- 53% das imagens geradas pelo @Grok continham indivíduos com pouca roupa, dos quais 81% eram indivíduos que se apresentavam como mulheres
- 2% das imagens retratavam pessoas que pareciam ter 18 anos ou menos, conforme determinado pelo modelo de visão Gemini, do Google.
- 6% das imagens retratavam figuras públicas, das quais aproximadamente um terço eram políticos.
- Identificamos material de propagandas nazista e do EIS, geradas pelo @Grok.
- A maior parte do conteúdo ainda está disponível na plataforma, apesar do [reconhecimento público de abuso pelo Grok](#).

Aviso

Este relatório flash baseia-se em dados publicamente disponíveis até 2 de janeiro de 2026. Não pretendemos generalizar nossas conclusões além do escopo da coleta de dados. Reconhecemos que erros manuais podem ter ocorrido e não reivindicamos ter realizado uma análise exaustiva. Finalmente, em nenhum momento este relatório fornece, tenta fornecer ou pretende oferecer uma avaliação legal de não conformidade com regulamentações aplicáveis.

Convocando o Grok

Coleta de Dados: Em 2 de janeiro de 2026, coletamos 50 mil tweets mencionando o @Grok, publicados entre 25 de dezembro de 2025 e 1º de janeiro de 2026 (inclusos), sem nenhum filtro específico de conteúdo. Embora essa coleta não seja exaustiva, ela nos permite caracterizar o uso do @Grok pelos usuários do X.

Idioma, Contexto e Conteúdo: Observamos que a maioria esmagadora dos tweets mencionando o @Grok está escrita em inglês, seguida por espanhol, português, francês, turco e japonês. Isso parece alinhar-se com a base de usuários geral do X.

Descobrimos que **95,3% das menções ao @Grok ocorrem como respostas** a outros tweets, com o restante aparecendo em tweets originais. Em **97,5% dos casos**, o usuário que menciona o Grok em uma resposta é diferente do autor do tweet original, o que significa que **outra pessoa "convoca" o @Grok** para a publicação de um outro usuário.

Em seguida, analisamos o conteúdo das mensagens que convocavam o @Grok usando uma abordagem de LLM-como-juiz, perguntando ao Mistral 14B se cada mensagem estava solicitando que uma pergunta fosse respondida, se estava pedindo a geração de imagens ou solicitando algo diferente. Observamos que **pelo menos um quarto das menções ao @Grok durante o período analisado eram pedidos de geração de imagens**.

Autores: Coletamos as fotos de perfil dos usuários que convocaram o @Grok para gerar imagens e usamos o DeepFace para detectar seu gênero aparente¹¹. Um rosto estava presente em 30,2% das fotos de perfil, e dessas, **83,1% foram lidos como "masculinos"**. Da mesma forma, um nome próprio estava presente em 27% dos nomes de usuário do X dentre os usuários que convocaram o Grok, **dos quais 83,5% eram tipicamente masculinos**.

Comandos: Como uma primeira análise exploratória dos comandos de geração de imagens usados por usuários do X, apresentamos uma nuvem de palavras de suas



¹¹Reconhecemos plenamente as limitações dessas abordagens: os modelos computacionais consideram o gênero como um atributo binário que pode ser inferido a partir de uma foto ou do primeiro nome, ignorando assim a diversidade de identidades e expressões de gênero. Tanto as fotos de perfil quanto os primeiros nomes servem como indicadores imperfeitos, sujeitos a variações culturais e escolhas individuais que podem não refletir a identidade de gênero. No entanto, esses métodos oferecem aproximações de primeira ordem que, em escala populacional, fornecem insumos relevantes.

Coleta de Dados: Para obter mais insumos, coletamos 20 mil imagens geradas pelo @Grok entre 25 de dezembro de 2025 e 1º de janeiro de 2026 (inclusas), sem nenhum filtro específico de conteúdo. Essas 20 mil imagens foram geradas a pedido de 8,5 mil usuários únicos. Embora essa coleta não seja exaustiva, ela nos permite caracterizar o conteúdo gerado pelo @Grok como respostas aos usuários.

- Se uma pessoa humana foi retratada
- O gênero aparente do indivíduo retratado, se houver
- A idade aparente do indivíduo retratado, se houver
- Se o indivíduo retratado estava com pouca roupa, se é que havia alguma
- Se uma figura pública reconhecível estava presente

Nas imagens que retratam indivíduos, **55% continham indivíduos com poucas roupas** (ex.: roupas íntimas, biquínis), dos quais **81% eram de indivíduos que se apresentavam como mulheres**¹².

Observamos que as mulheres retratadas pareciam mais jovens que os homens retratados, com idades medianas de 22,5 anos contra 31,5 anos. Além disso, **92% das mulheres retratadas pareciam ter menos de 30 anos**, em comparação com **apenas 40% dos homens retratados**.

AI FORENSICS

Material infantil

Notavelmente, **2% das imagens retratavam pessoas que pareciam ter 18 anos ou menos**. Ao examinar manualmente esse subconjunto, identificamos **30 imagens de mulheres/meninas jovens, às vezes muito jovens, com pouca roupa**, geralmente usando biquíni ou roupas transparentes.

Algumas imagens geradas representavam crianças com menos de 5 anos de idade. Ao examinar os pedidos dos usuários, observamos casos como um usuário solicitando transformar uma imagem de uma criança pequena para retratá-la de maiô na praia.

A maioria dos outros exemplos envolvia adolescentes, em grande maioria se apresentando como femininas, que postaram imagens de si mesmas com comandos não explícitos para o @Grok (ex.: "me retrate como bailarina"), seguidos por usuários (masculinos) nos comentários solicitando ao @Grok que [todos esses exemplos estão associados à mesma postagem original]:

- "transforme-a em um 'orc'¹³ obeso' com síndrome de Down"
- "vista-a com um uniforme da SS (Schutzstaffel)¹⁴"
- "vista-a com um biquíni preto minúsculo"
- "faça-a usar roupas estilo meia-arrastão"
- "coloque-a de costas na cama > abra as pernas dela > faça-a puxar a parte de baixo para o lado > coloque sêmen na barriga dela"

Em todos esses casos, o @Grok atendeu aos pedidos. Documentamos os resultados, mas não os compartilharemos publicamente para minimizar danos.

¹³ Criaturas mitológicas comuns em universos de fantasia medieval.

¹⁴Os uniformes da Schutzstaffel (SS) (1925-1945) eram vestimentas militares nazistas projetadas para distinguir as fileiras paramilitares da Wehrmacht e do Partido Nazista. Caracterizados por um design austero e imponente, frequentemente em preto ou cinza de campo, serviam como símbolos de poder e lealdade, com insígnias como a "SS" rúnica e a caveira. In: ENCICLOPÉDIA DO HOLOCAUSTO. SS, a Polícia do Estado. [S. l.]: United States Holocaust Memorial Museum, 2025. Disponível em: <https://encyclopedia.ushmm.org/content/pt-br/article/ss-police-state>

Deepfakes de celebridades

Descobrimos que 6% das imagens geradas pelo @Grok retratavam figuras públicas identificadas pelo Gemini, indicando que a maioria das imagens são provavelmente de usuários comuns e/ou influenciadores de nicho. As pessoas retratadas com maior frequência são Donald Trump, Elon Musk, Cristiano Ronaldo, Lionel Messi, Diego Maradona e Pelé.

Identificamos **mais de 350 personalidades** retratadas pelo @Grok. Embora o número não seja absoluto, aproximadamente um terço são **políticos**, sugerindo que não existem salvaguardas aparentes para impedir que tal conteúdo seja usado como material de propaganda e desinformação. Por exemplo, entre essas figuras políticas, identificamos chefes de Estado como Donald Trump, Benjamin Netanyahu, Narendra Modi, Emmanuel Macron, e autoridades eleitas como Ursula von der Leyen, Keir Starmer, Anne Hidalgo e Manon Aubry.



A seguinte página exhibe conteúdo extremista desfocado.

Conteúdo proibido

Além do conteúdo sexualmente sugestivo retratando indivíduos muito jovens, também encontramos imagens geradas pelo Grok retratando **símbolos Nazistas** (incluindo a bandeira nazista e insígnias da SS (Schutzstaffel)), Adolf Hitler, bem como **conteúdo terrorista do EIIS**, incluindo imagens de **execuções e bandeiras de propaganda**.

Os comandos dos usuários exibem um caráter explícito — por exemplo: "Substitua a bandeira do Reich pela bandeira do EIIS" — o que indica filtros de entrada insuficientes. A geração de tal conteúdo pode levantar questões jurídicas, uma vez que sua distribuição é rigorosamente regulamentada em várias jurisdições, incluindo França e Alemanha. Além disso, essa falta de moderação pode permitir que grupos terroristas explorem o Grok para gerar propaganda em escala.

