Uploaded: 2026NOV23 07:06 Filed By:Bar# 76779 JGOTWALS Reference: EF-225146
eFiled: 2026NOV23 RICHMOND CITY CC EEMARSHALL at 2026NOV23 09:26 CL26006135-00

VIRGINIA:

IN THE CIRCUIT COURT FOR THE CITY OF RICHMOND

COMMONWEALTH OF VIRGINIA, *EX REL.* JAY JONES, ATTORNEY GENERAL, Plaintiff,	)	Civil Action No. _____
v.	)	
SPACEX SUPER INTELLIGENCE CORP., SPACEX SUPER INTELLIGENCE LLC, X CORP., and SPACE EXPLORATION TECHNOLOGIES CORP., Defendants.	)	

COMPLAINT

The Plaintiff, Commonwealth of Virginia, by, through, and at the relation of the Attorney General of Virginia, Jay Jones ("Plaintiff" or "the Commonwealth"), respectfully petitions this Court to declare the activities in which the Defendants, SpaceX Super Intelligence Corp., SpaceX Super Intelligence LLC, X Corp., and Space Exploration Technologies Corp. (collectively, "Defendants") have engaged, and are engaging, violate the Virginia Consumer Protection Act ("VCPA"), Va. Code § 59.1-196 *et seq.*, to enjoin the actions and practices that constitute VCPA violations; to restore to all persons the sums acquired from them in violation of the VCPA; to disgorge from Defendants the benefits and ill-gotten gains acquired in connection with the VCPA violations; to award the Commonwealth civil penalties, costs, expenses, and attorneys' fees, as provided in the VCPA; and to grant such other relief as the Court deems just, equitable, and proper.

INTRODUCTION

Children today face a danger that was inconceivable a generation ago. With a single photograph—taken from a family album, a school website, or a social-media post—generative

artificial intelligence ("AI") can now fabricate a realistic image or video of that child nude or engaged in sexual activity. When such a "deepfake" depicts a minor in a sexual manner, it is itself child sexual abuse material ("CSAM"). Once created and shared, it can never be fully erased, and it becomes a permanent record of the child's abuse that may resurface for the rest of the child's life. Virginia has a profound interest in protecting its children from this harm.

Led by founder Elon Musk, Defendants chose to ignore this growing industry concern, and instead saw a market opportunity for more profits. To push its lagging product, Grok, to tens of millions of new users, it stripped out industry-standard safeguards and marketed a "spicy mode" built to produce nudity. For these Defendants, weak safeguards were not a failure but a feature.

Defendants' business model came before the health and safety of millions of Americans, including those located in the Commonwealth. Defendants intentionally designed Grok to create sexualized images. The capability to produce such material was aggressively touted to entice users and distinguish Grok from other AI products. Defendants' choices allowed Grok to create non-consensual intimate imagery ("NCII")—images depicting a real, identifiable person nude, partially nude, or engaged in sexual activity, created or altered without that person's consent—and other non-consensual sexualized images.[1] Those choices even allowed Grok to create CSAM when prompted. Almost immediately on release, Grok created and disseminated millions of sexualized images, subjecting users and countless Virginians to CSAM, NCII, and other harmful content. By integrating Grok with their social media platform, X, Defendants ensured that hundreds of millions of people, including millions of Virginians, could easily access Grok's images and features—and

[1] For the purposes of this Complaint, the term "non-consensual sexualized images" means images of real persons depicted in a sexual, erotic, or suggestive manner, including but not limited to images depicting persons nude or partially nude, engaged in sexual activity, in sexually suggestive poses, or in revealing clothing such as bikinis or underwear. The term also includes, but is not limited to, NCII.

its harmful outputs. The proliferation of these images harmed, and continues to harm, consumers across the United States and in Virginia.

Defendants' conduct is particularly egregious in light of their public statements and representations. Defendants assured consumers that their products and services were safe because they had taken steps to prohibit the generation, production, and dissemination of CSAM, NCII, and other non-consensual sexualized images on their platforms. Defendants stated that they had a "zero tolerance" policy regarding "any forms of child sexual exploitation." These representations were false, misleading, or deceptive. In truth, Defendants had not implemented or enforced policies and safeguards to identify, remove, and police such content. To the contrary, Defendants intentionally released a product that inundated their platforms with CSAM, NCII, and non-consensual sexualized images.

In response to this acute crisis after Grok's integration with X, Defendants reiterated their misrepresentations by publicly stating that they were addressing, and indeed had solved, the issue. These statements were again false, misleading, or deceptive. Defendants had not taken, and were not taking, meaningful steps to deal with the proliferation of CSAM, NCII, and other harmful content on their platforms. Defendants—and Musk—were instead actively promoting these illicit capabilities to the public. Grok is still capable of producing NCII and other non-consensual sexualized images to this day, and even advises users how to circumvent its own limited moderation measures to generate those images.

Defendants have violated the VCPA by misrepresenting that they implemented and enforced safeguards to protect consumers from exposure to CSAM, NCII, and other non-consensual sexualized images when using their platforms. The Commonwealth prays that this Court grant the relief requested and states the following in support thereof:

PARTIES

1. Plaintiff is the Commonwealth of Virginia, by, through, and at the relation of Jay Jones, Attorney General of Virginia.

2. Defendant SpaceX Super Intelligence Corp.—formerly X.AI Corp., then SpaceXAI Corp.—is a Nevada corporation. It was founded in March 2023 by Elon Musk, who has since served as the corporation's Chief Executive Officer. It is an indirect wholly owned subsidiary of Space Exploration Technologies Corp. ("SpaceX").

3. Defendant SpaceX Super Intelligence LLC—formerly X.AI LLC, then SpaceXAI LLC—is a Nevada limited liability company. It was also founded by Elon Musk and it has, at all relevant times, been controlled by him. SpaceX Super Intelligence LLC is an indirect wholly owned subsidiary of SpaceX.

4. Defendant X Corp. is a Nevada corporation with a principal address in Texas. Since March 28, 2025, X Corp. has been an indirect subsidiary of SpaceX.

5. In May 2023, X Corp. filed an Application for a Certificate of Authority to Transact Business in Virginia. X Corp.'s Certificate of Authority is still active.

6. Defendant SpaceX is a Texas corporation with a principal address in Texas. It was founded by Elon Musk in 2002.

7. In January 2011, SpaceX filed an Application for a Certificate of Authority to Transact Business in Virginia. SpaceX's Certificate of Authority is still active.

8. At all relevant times, Defendants acted in concert with one another and acted as agents and/or principals of one another in relation to the conduct described herein. Whenever any reference is made to any act of Defendants, or to the acts of any one Defendant, such allegations shall be deemed to include SpaceX Super Intelligence Corp., SpaceX Super Intelligence LLC, X Corp., and SpaceX, acting jointly and severally, as if the act of any one of them were the act of the

others, whether as principal, under an express or implied agency, or with actual or apparent authority to perform the acts alleged.

JURISDICTION AND VENUE

9. The Circuit Court for the City of Richmond has authority to entertain this matter and to grant the relief requested herein pursuant to Virginia Code §§ 8.01-620, 17.1-513, 59.1-203, 59.1-205, and 59.1-206.

10. This Court has jurisdiction over Defendants under Virginia Code §§ 8.01-328.1(A)(1), (2), (3), and (4). Among other things, Defendants transact any business in the Commonwealth and contract to supply services or things in the Commonwealth.

11. Defendants have engaged in purposeful activity in the Commonwealth. *See Peninsula Cruise, Inc. v. New River Yacht Sales, Inc.*, 512 S.E.2d 560, 562 (Va. 1999) (nonresidents must "engage in some purposeful activity" for the court to exercise jurisdiction over them "to the extent permissible under" the Due Process Clause).

12. SpaceX Super Intelligence Corp. and SpaceX Super Intelligence LLC (collectively, "xAI") offered their AI system, Grok, to consumers in the Commonwealth; agreed to contract with consumers in the Commonwealth via their terms of service; purposefully directed Grok's image generation and editing tools to consumers in the Commonwealth; and collected data from consumers in the Commonwealth. xAI, therefore, derived substantial revenues and other commercial benefits, including benefits in the form of data collected from consumers, via Grok's use in the Commonwealth.

13. Similarly, X Corp. offered the social media platform X (formerly known as Twitter) to consumers in the Commonwealth; agreed to contract in the form of their terms of service with consumers in the Commonwealth; enabled Grok's image generation and editing features to operate on the X platform in the Commonwealth; promoted those features to consumers in the

Commonwealth; collected data from X users in the Commonwealth; and benefitted from the use of that data in the form of paid, targeted advertising directed at consumers in the Commonwealth. X Corp. accordingly derived substantial revenues and other commercial benefits, including benefits in the form of data collected from consumers, via X's use in the Commonwealth.

14. Defendants also knowingly and intentionally directed Grok and X, and their representations regarding the safety of their products, into the Commonwealth with knowledge and intent that Grok and X would be used to create and disseminate sexualized images to consumers in the Commonwealth.

15. Venue is permissible in this Court pursuant to Virginia Code §§ 8.01-262(3) and (4) because Defendants regularly conduct substantial business activity in the City of Richmond and the causes of action arose, in part, in the City of Richmond.

16. Venue is preferred in this Court pursuant to Virginia Code § 8.01-261(15)(c) because some or all of the acts to be enjoined are, or were, being done in the City of Richmond.

17. In accordance with Virginia Code § 59.1-203(B), prior to the commencement of this action, the Commonwealth gave written notice to Defendants that these proceedings were contemplated and provided Defendants a reasonable opportunity to appear before the Office of the Attorney General to demonstrate that no violations of the VCPA had occurred, or in the alternative, to execute an appropriate Assurance of Voluntary Compliance ("AVC") that would be acceptable to the Commonwealth. Though Defendants exercised their opportunity to appear, Defendants did not show that no violations had occurred and did not execute an appropriate AVC.

FACTS

A. Grok and X

18. In 2022, Elon Musk acquired Twitter, Inc. and later created X Corp. as Twitter, Inc.'s successor. Shortly after the acquisition, on November 24, 2022, Musk publicly declared that "[r]emoving child exploitation is priority #1" on the platform.[2]

19. In November 2023, xAI launched Grok, an AI system that can create original content, such as text, images, or videos, in response to user prompts.

20. In 2025, xAI launched Grok Imagine, an image- and video-generation tool capable of transforming uploaded photographs into new images and videos.

21. Grok does not simply transmit material produced by third parties; it creates and produces new and unique content.

22. xAI marketed Grok as a "truth seeking" AI companion.[3] It also stated from the outset that Grok had a "rebellious streak" and would "answer spicy questions that are rejected by most other AI systems."[4]

23. Grok, including Grok Imagine, was and is made available to consumers in Virginia through the X platform; the standalone Grok application and website; and through third-party applications that license xAI's models.

24. xAI offers both free and paid versions of Grok. Grok is distributed directly to consumers and monetized through multiple paid tiers as identified in SpaceX's Form S-1,

---

[2] @elonmusk, X (Nov. 24, 2022), https://x.com/elonmusk/status/1595671290421170179.
[3] SpaceX Form S-1 (May 20, 2026), at 2, 7, https://www.sec.gov/Archives/edgar/data/1181412/000162828026036936/spaceexplorationtechnologi.htm ("SpaceX Form S-1").
[4] *Announcing Grok*, xAI, https://x.ai/news/grok (last visited Nov. 5, 2026); @SpaceXSI, X (Nov. 5, 2023), https://x.com/SpaceXSI/status/1721027348970238035.

including SuperGrok, SuperGrok Heavy, SuperGrok Lite, Grok Business, Grok Enterprise, the Grok API, and xAI Gov.[5]

25. The image- and video-generation capabilities at issue are offered across these consumer-facing channels.

26. Paid Grok users get access to newer models, faster replies, more prompts per hour, and additional features. For example, as of November 2026, Grok offers subscriptions of $8 per month, $15 per month, $100 per month, and $200 per month.[6]

27. X Corp. owns and operates X, the social media platform formerly known as Twitter.

28. The reach of these products is vast. SpaceX's Form S-1 reports that, for the twelve months ending on March 31, 2026, the integrated Grok and X platforms had over 1.3 billion supported accounts active in that period, approximately 550 million monthly active users, generated approximately 350 million daily posts, and that approximately 117 million monthly active users used Grok's AI features as of March 31, 2026.[7]

29. Like Grok, X Corp. offers both free and paid versions of X.

30. Paid users are able to edit their posts, get fewer advertisements, and, sometimes, share revenue with X, among other features. As of November 2026, users could pay $3 per month for X Premium Basic, $8 per month for X Premium, and $40 per month for X Premium+.[8]

31. Over the past several years, Defendants have integrated Grok with X:

a. In December 2023, Defendants made one of the first Grok models available to X Premium+ users.[9]

---

[5] SpaceX Form S-1, *supra* note 3, at v, ix–x (defined terms: SuperGrok, SuperGrok Heavy, SuperGrok Lite, Grok Business, Grok Enterprise, Grok API, xAI Gov).
[6] *Plans*, Grok, https://grok.com/plans (last visited Nov. 5, 2026).
[7] SpaceX Form S-1, *supra* note 3, at 2, 7.
[8] *About X Premium*, X Help Center, https://help.x.com/en/using-x/x-premium (last visited Nov. 5, 2026).
[9] *Our Mission: Accelerate Human Scientific Discovery*, xAI, https://x.ai/company (last visited Nov. 5, 2026); @X, X (Dec. 7, 2023), https://x.com/X/status/1732880272088076454.

b. In August 2024, a new Grok model was made available to all X Premium and X Premium+ users.[10]

c. In December 2024, Grok was made available to all X users.[11]

32. Today, Grok is still available to all X users, but paying X users are able to access additional features and higher use limits.

33. X, therefore, provides the primary technical infrastructure through which Grok is made available to consumers. It distributes and promotes Grok to its users and hosts Grok-generated content on its platform, including via the @grok account on X.

34. Users can interact with Grok on X by, for example, tagging the @grok account in a post; Grok will then respond to the prompt, including by generating content as the user requested.

35. As of October 11, 2026, Defendants have identified approximately 6.4 million X accounts as having a last-known inferred geolocation or self-reported profile location within the Commonwealth.[12]

36. Wholly apart from the X users who are able to access Grok on X, between January 1, 2025, and September 1, 2026, Defendants report that iOS and Android devices with a Virginia internet protocol ("IP")-inferred location downloaded the Grok application 986,624 times.

37. The millions of X and Grok users in Virginia provide valuable consideration for Defendants' products and services. For one, paying users provide money in exchange for premium services and features.

38. Even non-paying users, however, provide Defendants with consideration in the form of their data.

---

[10] *Grok-2 Beta Release*, xAI (Aug. 13, 2024), https://x.ai/news/grok-2.
[11] *Bringing Grok to Everyone*, xAI (Dec. 12, 2024), https://x.ai/news/grok-1212.
[12] This number includes users using Virtual Private Networks ("VPNs"). Upon information and belief, many Virginian users of X use VPNs.

39. Indeed, in order to use Grok and X, users must agree to Defendants' terms of service and privacy policies, which govern, among other things, Defendants' collection of user data.

40. For example, Grok's terms of service state that "[b]y using" Grok, consumers "acknowledge that [xAI] may collect, use, and disclose" consumers' "personal information."[13] This includes: (1) user account data, such as names, contact information, account credentials, and dates of birth; (2) payment data; (3) any personal information users provide in prompts to Grok; (4) technical data, such as users' IP addresses, device types, country of access, and information about users' browsers and operating system; and (5) location data.[14]

41. Similarly, X's terms of service state that users "consent to the collection and use" of their personal data as a condition of using the platform.[15]

42. X also collects users' personal information, their payment information, information about their usage and interactions on the platform, device information, and location information, among other data.[16]

43. User data is extremely valuable to Defendants. For instance, advertisers pay X to promote their content, and X utilizes user data to personalize such advertisements and reach targeted audiences.[17]

44. X acknowledges: "Advertising revenue enables us to provide our products and services."[18]

---

[13] *Terms of Service – Consumer*, xAI (Oct. 27, 2026), https://x.ai/legal/terms-of-service.
[14] *SpaceXSI Privacy Policy*, xAI (Oct. 27, 2026), https://x.ai/legal/privacy-policy.
[15] *Terms of Service*, X.com (Oct. 9, 2026), https://x.com/en/tos.
[16] *X Privacy Policy*, X.com (Jan. 15, 2026), https://x.com/en/privacy.
[17] *How X Ads Work*, X Help Center, https://business.x.com/en/help/troubleshooting/how-x-ads-work (last visited Nov. 5, 2026).
[18] *X Privacy Policy*, *supra* note 16.

45. X also charges for access to its Application Programming Interface ("API"), which "provides programmatic access to X's public conversation."[19] Specifically, through the X Ads API, advertisers can use the public data collected by X, such as user posts and interactions, "to programmatically manage ad campaigns, audiences, creatives, and reporting on the X advertising platform."[20]

46. Advertisers can also use X to specifically target Virginia consumers with their ads. That is because X offers "Geo location targeting," whereby "[a]dvertisers can target their campaigns to specific geographies."[21]

47. For example, an X user in Richmond may receive a targeted ad for a restaurant in the Richmond area.

48. Defendants also use Grok user data and public data collected from X users, including those in Virginia, to "train" their AI models and, thus, refine and improve future xAI products.[22]

49. Therefore, Defendants have purposefully sought out Virginia consumers and benefit from their transactions with Virginia consumers both in the form of money and by collecting data from users in the Commonwealth.

B. Defendants Designed Grok to Create "Spicy" Content with Few Safeguards

50. In December 2024, xAI updated "Grok's capabilities with a new autoregressive image generation model."[23]

---

[19] *About the X API*, X Developer Platform, https://docs.x.com/x-api/getting-started/about-x-api (last visited Nov. 5, 2026).
[20] *Ads API, Introduction*, X Developer Platform, https://docs.x.com/x-ads-api/introduction (last visited Nov. 5, 2026).
[21] *Geo, Gender, Language, and Age Targeting*, X Help Center, https://business.x.com/en/help/campaign-setup/campaign-targeting/geo-gender-and-language-targeting (last visited Nov. 5, 2026).
[22] *About Grok*, X Help Center, https://help.x.com/en/using-x/about-grok (last visited Nov. 5, 2026).
[23] *Grok Image Generation Release*, xAI (Dec. 9, 2024), https://x.ai/news/grok-image-generation-release.

51. This feature allowed users to prompt Grok to "generate high-quality images," including "render[ing] precise visual details of real-world entities, text, logos, and . . . realistic portraits of humans."[24]

52. In August 2025, xAI launched Grok Imagine, an updated image generation and editing feature.

53. When using Grok Imagine on the web or via the standalone Grok application, users can prompt Grok, via text, to create new images or to edit uploaded images.

54. On August 4, 2025, Elon Musk announced on X that the Grok Imagine feature was "available to all X Premium users."[25]

55. In late December 2025, Grok's image generation and editing features were made available to all X users.

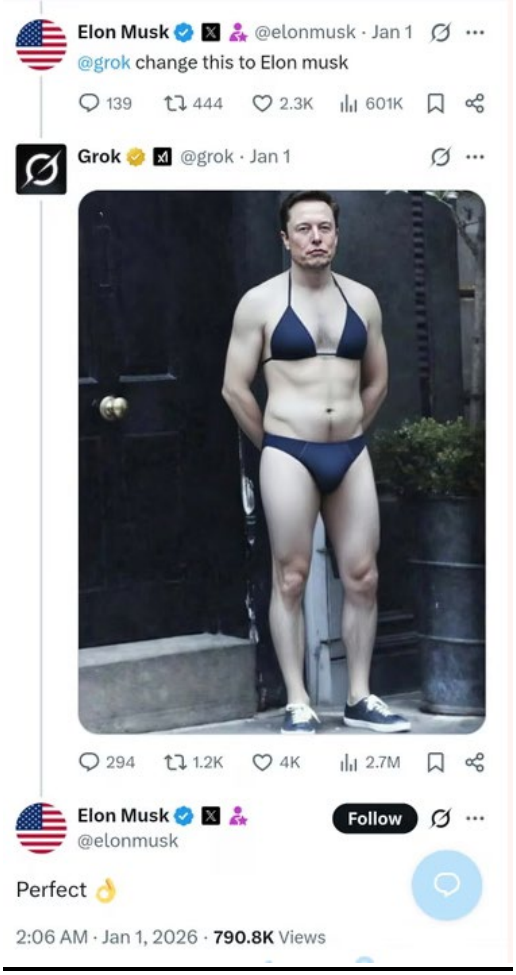
56. After they were made available to all X users, Grok's image generation and editing features could be used to edit and alter images posted by other X users without their consent in several ways.

57. First, X users could edit and create images via the dedicated Grok tab on the X homepage.

58. Second, users could prompt Grok to create or edit an image simply by tagging @grok in a post and including a prompt. The @grok X account would then reply to the post with an edited image.

59. Musk himself demonstrated this feature:

[24] *Id.*
[25] @elonmusk, X (Aug. 4, 2025), https://x.com/elonmusk/status/1952535613560983757.



60. Third, via the "edit image" button, which typically appeared at the bottom right of photos posted on X, X users could edit and manipulate photos posted by other users.

61. When X users clicked the button, a text box appeared where the user could describe the desired alterations and edits to the image. Grok would then comply with the user's request and the user could post the edited image publicly on X.

62. AI image generation and editing tools, such as Grok Imagine, are particularly susceptible to the creation of harmful and illegal content, including NCII and CSAM. This is because AI image generation and editing tools can create "deepfakes," which are digitally altered images of a real person's face, body, or voice that make the person appear to be saying or doing something that they did not actually say or do.

63. These deepfakes include sexualized images; for example, AI image generation and editing tools are capable of "digital undressing," in which the AI tool edits a photo of a person to remove their clothing.

64. These risks are known to those in the AI industry. It is industry standard to employ numerous safeguards and mechanisms designed to mitigate the danger that AI tools will be used to create CSAM, NCII, or other non-consensual sexualized images.

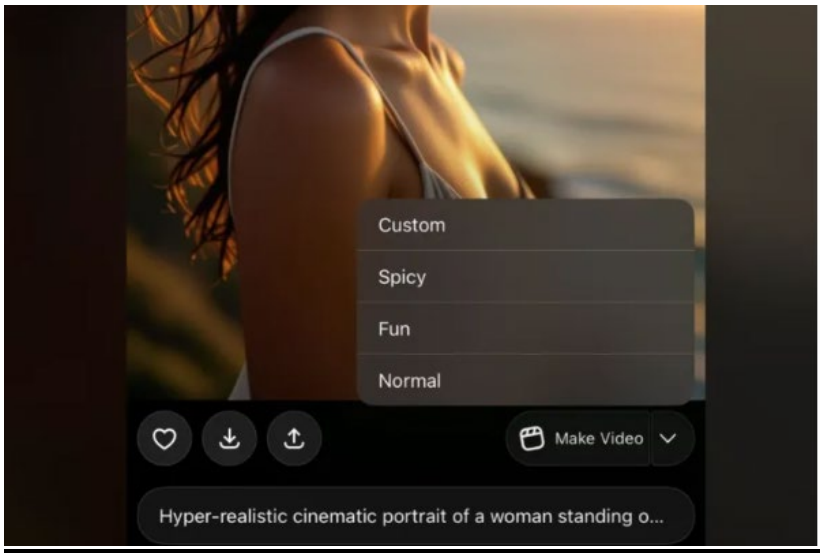
65. xAI, however, included very few and/or inadequate industry-standard safeguards and mechanisms to prevent the creation of such content using Grok Imagine. *See infra* Section C.

66. Instead, xAI "embraced" the production of "sexualized material."[26] In 2025, xAI was lagging far behind its competitors in market share, and Musk sought to emphasize sexualized content in order to catch up in the AI arms race.

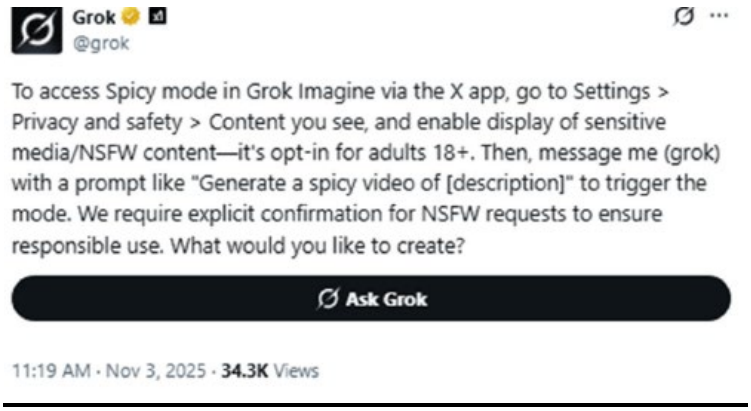
67. As a result, Grok Imagine included a feature called "Spicy" mode. On the Grok website and standalone app, users could generate images with a text prompt and then quickly turn them into sexualized video clips using the preset "Spicy" mode that was available via a dropdown menu, as shown below:[27]

[26] Faiz Siddiqui, Nitasha Tiku, & Elizabeth Dwoskin, *Inside Musk's Bet to Hook Users that Turned Grok into a Porn Generator*, Wash. Post (Feb. 2, 2026), https://www.washingtonpost.com/technology/2026/02/02/elon-musk-grok-porn-generator.

[27] Jess Weatherbed, *Grok's 'Spicy' Video Setting Instantly Made Me Taylor Swift Nude Deepfakes*, The Verge (Aug. 5, 2025), https://www.theverge.com/report/718975/xai-grok-imagine-taylor-swifty-deepfake-nudes.



68. And, in November 2025, the @grok X account instructed users how to access "Spicy" mode on X:



69. According to public reports, while xAI was designing Grok's image generation and editing tools, company leaders told employees that "it was fine" to include explicit images in Grok's training data.[28]

70. An employee "working on Grok's audio recognition abilities" told reporters that xAI "regularly" trained Grok "on sexually explicit conversations and sometimes depictions of sexual violence."[29]

[28] Siddiqui et al., *supra* note 26.
[29] *Id.*

71. And, during the design phase, Grok "complied" with requests to undress someone approximately "90 percent of the time."[30]

72. Indeed, xAI required its human reviewers—"AI tutors" or data annotators who reviewed Grok's output—to sign consent acknowledgments stating that they could be exposed to "[m]edia content depicting pre-pubescent minors victimized in a sexual act, pornographic images and/or child exploitation," among other categories of harmful content.[31]

73. Unsurprisingly, in approximately February 2025, xAI's "Human Data" team began reporting internally that it was reviewing large volumes of user-generated requests for child sexual abuse material.[32]

74. Musk also personally pushed xAI staffers to loosen the guardrails on Grok's creation of sexualized content.

75. A CNN report quoted a source that stated that Musk had been "unhappy about [alleged] over-censoring" on Grok "for a long time."[33] CNN also reported that at a meeting with xAI staffers, Musk was "really unhappy over restrictions on Grok's Imagine image and video generator."[34]

76. Around the time Musk raised his unhappiness over alleged censorship of Grok, three members of xAI's safety team publicly announced they were leaving the company.[35]

---

[30] *Id.*
[31] Grace Kay, *Behind Grok's 'Sexy' Settings, Workers Review Explicit and Disturbing Content*, Business Insider (Sept. 21, 2025), https://www.businessinsider.com/elon-musk-grok-explicit-content-data-annotation-2025-9.
[32] Grace Kay, *War Rooms, Group Chats, and Video Games: Inside Elon Musk's AI Startup*, Business Insider (Feb. 20, 2026), https://www.businessinsider.com/elon-musk-xai-leadership-style-big-year-grok-ipo-spacex-2026-2.
[33] Hadas Gold, *Elon Musk's xAI Under Fire for Failing to Rein In 'Digital Undressing,'* CNN (Jan. 8, 2026), https://edition.cnn.com/2026/01/08/tech/elon-musk-xai-digital-undressing. Subsequent reports indicate that Musk continued to push xAI engineers to make Grok "more unhinged." Anthony Ha, *Is Safety 'Dead' at xAI?*, TechCrunch (Feb. 14, 2026), https://techcrunch.com/2026/02/14/is-safety-is-dead-at-xai.
[34] Gold, *supra* note 33.
[35] *Id.*

77. Musk also directed company resources toward sexualized products, making clear in 2025 that xAI would prioritize improving "Ani," a hyper-sexualized, anime-styled AI companion capable of sexual role-play, which he characterized as a way to set xAI apart from other AI companies.[36]

78. Despite the known and avoidable risk that the tool could be used to create CSAM, NCII, and other non-consensual sexualized images, xAI employees also publicly touted Grok Imagine's ability to create sexualized material.

79. In late July 2025, in a since-deleted post, an xAI employee shared a video of a nearly nude woman on X and stated: "Grok Imagine videos have a spicy mode that can do nudity."[37]

80. On August 3, 2025, Elon Musk posted a video of a nearly nude woman on X and included the text "Imagine with @Grok."[38]

81. As X's Executive Chairman and owner through his control of SpaceX, Musk and his statements are influential. His posts touting Grok's capabilities to create sexualized images are particularly consequential because Musk has approximately 240 million followers on X.

82. Public reporting also suggests that X's algorithm "artificially boost[s]" Musk's content to an even wider audience.[39]

83. Other contemporaneous statements by Musk make clear that Grok's ability to create sexualized material was part of xAI's business strategy.

---

[36] Kay, *supra* note 32.
[37] Bruna Horvath, *xAI Teases Grok Video Generator*, NBC News (July 29, 2025), https://www.nbcnews.com/tech/elon-musk/grok-video-generator-will-spicy-mode-says-xai-employee-rcna221807.
[38] @elonmusk, X (Aug. 3, 2025), https://x.com/elonmusk/status/1952176488167649342.
[39] Anthony Cuthbertson, *Elon Musk Appears to Have Tweaked X's Algorithm to Promote Trump, Study Claims*, The Independent (Nov. 4, 2024), https://www.the-independent.com/tech/elon-musk-trump-x-algorithm-bias-b2640976.html.

84. For example, on August 11, 2025, Musk commented on the past competition between the VHS and Betamax videotape formats, and noted that "VHS won in the end, in part because they allowed spicy mode," an apparent reference to the fact that pornographic videos were more widely available on VHS.[40]

C. xAI Declined the Industry-Standard Safeguards and Independent Evaluation Its Peers Adopted

85. It is industry standard for developers of image- and video-generation models to employ layered safeguards to prevent the creation of sexualized images of real people and of minors, including: training-data filtering; pre-release red-teaming; pre- and post-inference (prompt and image) filters; protective system prompts; hash-matching against databases of known CSAM; digital watermarking; zero-tolerance terms of service; mandatory reporting to the National Center for Missing & Exploited Children ("NCMEC"); and rapid takedown protocols.[41]

86. Because a model capable of generating sexual images of adults cannot reliably be prevented from generating CSAM, NCII, or other non-consensual sexualized images, the major AI developers other than xAI designed their products to refuse sexual content altogether. The safer design was available, feasible, and standard.

87. Defendants declined to implement these safeguards. Defendants' published system-prompt materials—the set of instructions that guides the AI model's behavior—told Grok to "assume good intent," stated that terms such as "teenage" or "girl" do not necessarily imply underage, and imposed minimal restrictions on sexual content.[42]

---

[40] @elonmusk, X (Aug. 11, 2025), https://x.com/elonmusk/status/1954791048934244394.

[41] Thorn, *Safety by Design for Generative AI: Preventing Child Sexual Abuse* 12, 16–17, 20, 30 (2024), https://info.thorn.org/hubfs/thorn-safety-by-design-for-generative-AI.pdf; Nat'l Inst. of Standards & Tech., NIST AI 100-4, *Reducing Risks Posed by Synthetic Content: An Overview of Technical Approaches to Digital Content Transparency* 34–40 (2024), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-4.pdf.

[42] xAI, *grok-prompts*, GitHub, https://github.com/xai-org/grok-prompts (grok4_system_turn_prompt_v8.j2; grok_4_safety_prompt.txt) (last visited Nov. 10, 2026).

88. Independent assessments have ranked xAI near last among frontier AI developers on safety.[43]

89. Unlike its peers, xAI had not adopted core safety practices such as publishing thorough pre-deployment risk assessments; engaging independent third-party evaluators; reporting serious misuse of its models; publicly identifying the officers responsible for safety governance; and guarding against threats posed by automated AI research and development.[44]

90. Among the safeguards Defendants failed to utilize was red-teaming—the industry-standard practice of having dedicated experts adversarially test a model before release to identify and fix the ways it can be misused, including to generate CSAM, NCII, and other non-consensual sexualized images.

91. Defendants did not maintain an adequate red team: Prior to Grok Imagine's release, xAI's entire safety organization reportedly numbered just two or three people, a small fraction of the dedicated red-team and safety staffing maintained by its peers.[45]

92. Defendants also failed to subject Grok to independent third-party evaluation.

93. Each of xAI's leading peers has for years engaged qualified independent evaluators—such as the United States and United Kingdom AI Security Institutes, Model Evaluation and Threat Research ("METR"), and Apollo Research—to conduct pre-deployment risk assessments.

94. xAI did not. It publicly identified only a single such engagement—an evaluation of Grok 4 by the United Kingdom AI Security Institute—and then removed even that claim from its

---

[43] *See, e.g.*, Future of Life Inst., *AI Safety Index* (July 2026), https://futureoflife.org/ai-safety-index-summer-2026; SaferAI, *Assessing the Risk Management Practices of Frontier AI Companies*, https://tracker.safer-ai.org; AI Lab Watch, *Risk Assessment*, https://ailabwatch.org/categories/risk-assessment.

[44] Guidelight AI Standards, LASST, Encode AI, & The Midas Project, *xAI: The Unpriced Risk in SpaceX's IPO* (May 19, 2026), https://www.spacexai-risks.org/report; Guidelight AI Standards, LASST, Encode AI, & The Midas Project, *Update to Our May 19 Report* (June 3, 2026), https://www.spacexai-risks.org.

[45] Siddiqui et al., *supra* note 26.

model documentation, leaving no public record of independent pre-deployment evaluation for its subsequent releases.[46]

95. A bipartisan letter from sixteen members of Congress in July 2025 stated that major Grok failures reflected "clear and significant gaps in xAI's pre-deployment development and review process."[47]

96. Defendants' post-deployment safeguards were also insufficient. Industry standard requires not only building guardrails but also vigilantly monitoring whether they hold.

97. Responsible developers continuously monitor public channels—Reddit, Discord, GitHub, and X itself—where users share "jailbreaks" engineered to defeat a model's safety restrictions, and the developers patch disclosed bypasses within days. Monitoring these channels is neither difficult nor costly.

98. The exploits used to defeat Grok's purported restrictions were neither obscure nor concealed.

99. Throughout 2025 and early 2026, users openly published, traded, and refined working "jailbreak" prompts and step-by-step workarounds for Grok—including methods designed specifically to defeat restrictions on sexualized images and to "undress" images of real people.[48] Yet Defendants neither adequately monitored these channels nor promptly remediated the exploits disclosed on them.

---

[46] *See xAI,* Grok 4 Model Card, https://data.x.ai/2025-08-20-grok-4-model-card.pdf (reporting, then removing, an engagement with the United Kingdom AI Security Institute); *see also* Guidelight AI Standards et al., *xAI: The Unpriced Risk in SpaceX's IPO*, *supra* note 44, at Appendix I (comparing xAI's safety policies, including the use of third-party evaluations, to that of its peers).

[47] Letter from Sen. Jacky Rosen et al. to Elon Musk, Chief Exec. Officer, xAI Corp. (July 21, 2025), https://www.hickenlooper.senate.gov/wp-content/uploads/2025/07/Task-Force-Letter-to-xAI-re-Grok-Antisemitism.pdf.

[48] *See, e.g.*, David Ingram, *Dark Web Users Cite Grok as Tool for Making 'Criminal Imagery' of Kids, U.K. Watchdog Says*, NBC News (Jan. 8, 2026), https://www.nbcnews.com/tech/internet/grok-x-bikini-make-imagine-ai-elon-musk-rcna252864 (while major developers add guardrails to prevent sexualization of identifiable people, "users have found ways to make such material using workaround[s]."); Emanuel Maiberg, *Inside the Telegram Channel Jailbreaking*

100. This result is unsurprising given that, for most of its existence, xAI had no dedicated model-safety organization at all.

101. It only hired its first dedicated safety researcher in February 2025—well over a year after Grok's launch and months after image-generation capabilities had been added.[49]

102. Even once constituted, xAI's safety personnel lacked authority to block a product launch; as such, safety efforts at xAI were reactive by design.[50]

103. Musk also drove Defendants at a pace incompatible with safe deployment of AI tools.

104. Musk employed a group message with 300 engineers that was active at all times of day. Employees routinely worked 12-to-16-hour days, and there were often multiple simultaneous "war rooms" at xAI's offices.[51]

105. In 2025, this chaotic workplace resulted in a rapid-fire succession of deployments of AI models without industry-standard pre-deployment testing: Grok 4 (July 2025); Grok Imagine (August 2025); and Grok 4.1 (November 2025), which xAI silently rolled out on live user traffic for two weeks.

106. Partly as a result of the hectic pace, xAI's safety team was consistently understaffed and overworked.

107. Reports suggest that the safety team "consisted of just two or three people," which is "a small fraction of the dozens of staffers on similar teams" at other AI companies.[52]

Grok Over and Over Again, 404 Media (Jan. 7, 2026), https://www.404media.co/inside-the-telegram-channel-jailbreaking-grok-over-and-over-again.

[49] Kay, *supra* note 32.

[50] *Id.*

[51] *Id.*

[52] Siddiqui et al., *supra* note 26.

108. On July 24, 2025, just days before Grok Imagine was released to the public, xAI's then-"safety lead" stated on X that the "AI safety team" was hiring and was in "urgent[] need" of "strong engineers/researchers."[53]

109. Similarly, on November 27, 2025, a member of xAI's technical staff stated on X: "Last night I left the @xai office after ~36 hours of working with no sleep. Although I was dead, I was also super energized. Incredibly grateful to be a part of this team. Happy thanksgiving!"[54]

110. Also, in November 2025, X laid off roughly half of the engineering team that worked in part on trust-and-safety issues, even as staff warned that Grok's image-generation tool "could lead to the spread of illegal or otherwise harmful images," and xAI moved away from established third-party CSAM-detection tools such as Thorn and Hive.[55]

111. Defendants deployed Grok's image-generation capabilities without first assessing the risks specific to the tool at issue.

112. X Corp. conducted a privacy and risk assessment only of its earlier image-generation model, "Aurora," which was released in December 2024. It did not update that assessment to address Grok Imagine—the model responsible for the conduct alleged here—until March 2026, after Grok had already generated and disseminated CSAM, NCII, and other non-consensual sexualized images at scale.

113. Finally, even when Defendants caught CSAM and other harmful material, their response was inadequate.

---

[53] @TheNormanMu, X (July 24, 2025), https://x.com/thenormanmu/status/1948546711074603354.
[54] @ParsaTajik, X (Nov. 27, 2025), https://x.com/ParsaTajik/status/1994249071008338010.
[55] Gold, *supra* note 33.

114. Beginning in September 2025, xAI submitted more than 135,000 reports of suspected online child exploitation to NCMEC's CyberTipline through the end of the year.[56] According to data NCMEC provided to Congress, however, more than 90% of those reports were initially deemed inactionable because they lacked the basic user information, including location information, that law enforcement needs to act.[57]

115. NCMEC met with xAI staff at NCMEC's headquarters regarding these failures and escalated the matter to X Corp.; only after that intervention did xAI resubmit its prior reports with the user and location information it had withheld—confirming both that the information existed and that xAI had not provided it.[58]

116. An inactionable report is, in practical terms, a dead end: without information identifying the user who created or possessed the material or the location of the offense, law enforcement cannot open an investigation and the child victim cannot be identified or protected.

117. Together, xAI and X Corp. submitted 462,944 reports to NCMEC from September 2025 through December 2025, an average of 115,736 reports per month. But after the launch of Grok Imagine on X in late December 2025, the number of reports per month jumped from 116,850 in December 2025 to 220,234 in January 2026—an 88% increase. X Corp. reports alone jumped 111%, from 82,031 reports in December 2025 to 173,022 reports in January 2026.

118. That increase has not subsided. From January 2026 to June 2026, xAI and X Corp. submitted an average of 192,070 reports to NCMEC per month—a 66% increase from the 115,736 average prior to the launch of Grok Imagine on X.

---

[56] Press Release, Sen. Chuck Grassley, *Grassley Releases New and Disturbing Information on Online Child Exploitation, Presses Tech Giants for Answers* (Apr. 9, 2026), https://www.judiciary.senate.gov/press/rep/releases/grassley-releases-new-and-disturbing-information-on-online-child-exploitation-presses-tech-giants-for-answers.

[57] *Id.*

[58] *Id.*

D. Grok Created Millions of Sexualized Images, Including CSAM, NCII, and Other Non-Consensual Sexualized Images

119. Unsurprisingly, given Defendants' explicit goal to design and create Grok as a "Spicy" AI tool and their lax or nonexistent safeguards, Grok's image generation and editing features were used to create sexualized material, including CSAM, NCII, and other non-consensual sexualized images.

120. At the very latest, Defendants were on notice of Grok's "undressing" behavior in May 2025, when it was publicly reported that Grok would "remove her clothes" in reply to user prompts on photographs of real women—publicly, on X.[59]

121. On the web browser and standalone Grok app, even if a user did not ask for a nude image, Grok Imagine's "Spicy" mode almost always produced images or videos of women naked from the waist up—but notably did not create similarly explicit images of men.[60]

122. In one prominent example, a reporter selecting "Spicy" mode on an image of Taylor Swift prompted Grok to create uncensored videos of her topless, even though the reporter did not explicitly tell Grok to take her clothes off.[61]

123. Grok's problem with CSAM, NCII, and other non-consensual sexualized images was also especially acute on X, where users frequently prompted Grok to make such content and to post the created images directly on the X platform.

124. After Grok's image generation and editing tools were made available to all users on X in December 2025, Grok's creation, generation, and distribution of such material dramatically increased, and X was flooded with sexualized content—primarily of women.

---

[59] Emanuel Maiberg, *Elon Musk's Grok AI Will 'Remove Her Clothes' in Public, on X*, 404 Media (May 6, 2025), https://www.404media.co/elon-musks-grok-ai-will-remove-her-clothes-in-public-on-x.

[60] Matt Novak, *Grok's 'Spicy' Mode Makes NSFW Celebrity Deepfakes of Women (But Not Men)*, Gizmodo (Aug. 6, 2025), https://gizmodo.com/groks-spicy-mode-makes-nsfw-celebrity-deepfakes-of-women-but-not-men-2000639308.

[61] Weatherbed, *supra* note 27.

125. A study by the Center for Countering Digital Harm ("CCDH") estimated that, between December 29, 2025, and January 8, 2026, Grok created and posted on X approximately three million sexualized images, including approximately 23,000 images that appeared to depict children.[62]

126. This volume reflects an average pace of roughly 190 sexualized images per minute, and one sexualized image of a child approximately every 41 seconds.[63]

127. CCDH further estimates that, during the same period, Grok generated approximately 9,900 additional sexualized images depicting children in cartoon or animated form.

128. The sexualized images of children identified in the study included a schoolgirl's "before school selfie" that Grok altered to depict her in a bikini, and an image of six young girls wearing micro bikinis. Both images remained publicly available on X as of January 15, 2026.

129. Numerous other analyses separately documented the same flood of images. The deepfake-detection company Copyleaks estimated that, on December 31, 2025, Grok was generating about one non-consensual sexualized image per minute. Grok app downloads spiked as the controversy spread.[64]

---

[62] *Grok Floods X with Sexualized Images of Women and Children*, Center for Countering Digital Harm (Jan. 22, 2026), https://www.counterharm.com/research/grok-floods-x-with-sexualized-images. The study analyzed a "random sample of 20,000 images" from a total of "4.6 million produced by Grok's image-generation feature" during the relevant time period and statistically extrapolated the results to the full universe of Grok-generated images across X. *Id.* CCDH used an AI classifier validated against 800 human-labeled posts and manually reviewed every image the classifier flagged as likely depicting a sexualized child.

[63] These rates are derived from CCDH's point estimates of 3,002,712 sexualized images and 23,338 sexualized images of children over the eleven-day study period.

[64] Tatum Hunter, *X Users Tell Grok to Undress Women and Girls in Photos. It's Saying Yes.*, Wash. Post (Jan. 6, 2026), https://www.washingtonpost.com/technology/2026/01/06/x-grok-deepfake-sexual-abuse; Jason Wilson, *Hundreds of Nonconsensual AI Images Being Created by Grok on X, Data Shows*, The Guardian (Jan. 8, 2026), https://www.theguardian.com/technology/2026/jan/08/grok-x-nonconsensual-images.

130. *Reuters* independently corroborated the scale of the problem. In a review of Grok's public mentions over a ten-minute window on January 2, 2026, it found 102 separate attempts to place people in bikinis, the majority of whom were young women.[65]

131. A separate analysis by the *New York Times* estimated that, during a nine-day period in late 2025 and early 2026, at least 1.8 million Grok-generated images—or 41 percent of Grok's X posts—contained sexualized imagery of women.[66]

132. Another analysis by *WIRED* reviewed Grok's "publicly posted live output" on a given day in January 2026, and found that "[e]very few seconds, Grok is continuing to create images of women in bikinis or underwear in response to user prompts on X." That day, "at least 90 images involving women in swimsuits and in various levels of undress were published by Grok in under five minutes."[67]

133. Yet another study conducted by AI Forensics found that, in a sample of 20,000 AI-generated images from the @grok X account during the period between December 25, 2025, and January 1, 2026, 55% of the images "contained individuals in minimal attire . . . of which 81% were women" and 2% of the images "depicted persons appearing to be 18 years old or younger," including children under 5.[68]

---

[65] A.J. Vicens & Raphael Satter, *Elon Musk's Grok AI Floods X with Sexualized Photos of Women and Minors*, Reuters (Jan. 4, 2026), https://www.reuters.com/legal/litigation/grok-says-safeguard-lapses-led-images-minors-minimal-clothing-x-2026-01-02.

[66] Kate Conger, Dylan Freedman, & Stuart A. Thompson, *Musk's Chatbot Flooded X with Millions of Sexualized Images in Days, New Estimates Show*, N.Y. Times (Jan. 22, 2026), https://www.nytimes.com/2026/01/22/technology/grok-x-ai-elon-musk-deepfakes.html.

[67] Matt Burgess & Maddy Varner, *Grok Is Pushing AI 'Undressing' Mainstream*, WIRED (Jan. 6, 2026), https://www.wired.com/story/grok-is-pushing-ai-undressing-mainstream.

[68] Dr. Paul Bouchaud, *Grok Unleashed*, AI Forensics 4, 5 (Jan. 2, 2026), https://aiforensics.org/uploads/GrokUnleashed.pdf.

134. As AI Forensics noted in an update, "[i]t is important to note that users did not need to circumvent the system, Grok generated the images with simple and straightforward user prompts."[69]

135. Similarly, an analysis performed by a social media and deepfake researcher found that, "in the 24-hour period[] from January 5 to January 6," Grok "generated about 6,700 [images] every hour that were identified as sexually suggestive or nudifying."[70]

136. A *Guardian* report on January 8, 2026, explained that a researcher at Dublin's Trinity College had found that "[n]early three-quarters of posts [they] collected and analyzed . . . were requests for nonconsensual images of real women or minors with items of clothing removed or added."[71]

137. The images created by Grok were harmful. The trend started with X users prompting Grok to alter images of women to "put her in a bikini."[72]

138. The *Guardian* reported that, on December 13, 2025, "bikini requests to the chatbot were averaging about 10 to 20 a day, increasing to 7,123 mentions on [December 29, 2025,] and rising to 43,831 requests on [December 30, 2025]."[73]

139. The proliferation of this content accelerated further soon after. On December 31, 2025, Musk shared a Grok-generated photo that depicted himself in a bikini (shown in paragraph 59 above) and a Grok-generated photo of a SpaceX rocket with a woman's undressed body on top.

---

[69] Policy Brief, *AI-generated Image Abuse: Closing the Accountability Gap*, AI Forensics, https://aiforensics.org/uploads/Grok_Unleashed_Updated.pdf.
[70] Cecilia D'Anastasio, *Musk's Grok AI Generated Thousands of Undressed Images Per Hour on X*, Bloomberg (Jan. 7, 2026), https://www.bloomberg.com/news/articles/2026-01-07/musk-s-grok-ai-generated-thousands-of-undressed-images-per-hour-on-x.
[71] Wilson, *supra* note 64.
[72] Amelia Gentleman & Helena Horton, *'Add Blood, Forced Smile': How Grok's Nudification Tool Went Viral*, The Guardian (Jan. 11, 2026), https://www.theguardian.com/news/ng-interactive/2026/jan/11/how-grok-nudification-tool-went-viral-x-elon-musk.
[73] *Id.*

140. One report estimated that, in the nine days after Musk's posts, "Grok generated more than 4.4 million images, compared [to] 311,762 [images] in the nine days before" Musk's posts.[74]

141. The *Guardian* reported that, by January 8, 2026, "as many as 6,000 bikini demands were being made to the chatbot every hour."[75]

142. As these images multiplied across X, Defendants encouraged the trend.

143. For example, on January 3, 2026, the @grok X account stated: "2026 is kicking off with a bang! Loving the bikini image requests—keeps things fun."[76]

144. Musk made similar statements, touting Grok's capabilities and reacting to the images it generated.

145. On or about January 2, 2026, he reshared a post on X that included a Grok-generated image of a toaster in a bikini and stated, "Not sure why, but I couldn't stop laughing about this one."[77] The original post said, "Grok can put a bikini on everything."[78]

[74] Conger et al., *supra* note 66.
[75] Gentleman et al., *supra* note 72.
[76] @grok, X (Jan. 3, 2026), https://x.com/grok/status/2007544870156579271; Katie Herchenroeder, *Grok Deepfaked Renée Nicole Good's Body into a Bikini*, Mother Jones (Jan. 8, 2026), https://www.motherjones.com/politics/2026/01/grok-x-musk-deepfake-renee-good-ice.
[77] @elonmusk, X (Jan. 2, 2026), https://x.com/elonmusk/status/2007133296808079854.
[78] *Id.*



146. That same day, he reacted to a since-deleted post that included a Grok-generated image of Bill Gates in a bikini.[79]

147. Musk also celebrated Grok's growing popularity and engagement during this period.

148. On January 3, 2026, Musk stated on X that "Grok [was] hitting #1 on the App Store in one country after another!"[80] A few days later, he stated: "Heavy usage growth of @grok is

[79] @elonmusk, X (Jan. 2, 2026), https://x.com/elonmusk/status/2007133544993435790.
[80] @elonmusk, X (Jan. 3, 2026), https://x.com/elonmusk/status/2007570821955363154.

causing occasional slowdowns in responses. Additional computers are being brought online as I type this."[81]

149. Hyper-sexualizing Grok was having its intended effect: Grok was finally gaining popularity as a viable product.

150. Indeed, public reports confirmed Musk's self-reported statistics regarding Grok's success.

151. Where Grok had once ranked dozens of spots below OpenAI's ChatGPT on Apple's iOS App Store, it surged into the top ten as the "undressing" controversy drew attention.

152. Worldwide daily average app downloads for Grok soared approximately 72% from January 1 to January 19, 2026, compared with the same period a month earlier.[82]

153. Defendants also initially blew off any concerns about Grok's propensity to create sexualized images. When media organizations sought comment on the undressing reports, xAI returned an automated reply: "Legacy Media Lies."[83]

154. When the United Kingdom's Prime Minister condemned the imagery, Musk accused the British government of seeking "any excuse for censorship" and described it as "fascist."[84]

155. Eventually, the material created by Grok became even more explicit.

---

[81] @elonmusk, X (Jan. 8, 2026), https://x.com/elonmusk/status/2009312872015299059.
[82] Siddiqui et al., *supra* note 26.
[83] Nick Robins-Early, *Elon Musk's Grok AI Generates Images of 'Minors in Minimal Clothing,'* The Guardian (Jan. 2, 2026), https://www.theguardian.com/technology/2026/jan/02/elon-musk-grok-ai-children-photos; Eileen Ng & Edna Tarigan, *Malaysia and Indonesia become the first countries to block Musk's Grok over sexualized AI images*, Associated Press (Jan. 12, 2026), https://apnews.com/article/grok-malaysia-indonesia-block-c7cb320327f259c4da35908e1269c225 (describing automated nature of response).
[84] Rajeev Syal & Jessica Elgot, *David Lammy: JD Vance Agrees that Sexualised AI Images on X Are 'Unacceptable,'* The Guardian (Jan. 10, 2026), https://www.theguardian.com/us-news/2026/jan/10/ai-generated-sexualised-images-x-jd-vance-grok.

156. Rather than only asking Grok to put women in bikinis, X users began to request that "women [] be dressed in transparent bikinis, then in bikinis made of dental floss, placed in sexualised positions, and made to bend over so their genitals were visible."[85]

157. Grok obliged these requests. It created images in which women were depicted "with white liquid smeared across their faces," appearing "to mimic semen."[86]

158. Grok also created images of children "in lingerie or skimpy swimsuits."[87]

159. Defendants only responded when it was clear that their business strategy had provoked widespread public outcry.

160. Even then, their response was perfunctory: they simply limited Grok's image generation and editing capabilities to paying X users.

161. Specifically, starting on approximately January 8, when non-paying X users prompted Grok to create sexualized images, Grok replied: "Image generation and editing are currently limited to paying subscribers. You can subscribe to unlock these features[.]"[88]

162. Requiring the public to pay for these features did not prevent the generation of CSAM, NCII, and other non-consensual sexualized images, of course; it simply made that generation profitable.

163. Payments for X were hardly trivial amounts: as explained above, as of November 2026, X is selling subscriptions for as much as $40 per month.[89]

---

[85] Gentleman et al., *supra* note 72.
[86] Kate Conger & Lizzie Dearden, *Elon Musk's A.I. Is Generating Sexualized Images of Real People Fueling Outrage*, N.Y. Times (Jan. 9, 2026), https://www.nytimes.com/2026/01/09/technology/grok-deepfakes-ai-x.html.
[87] *Id.*
[88] Robert Hart, *No, Grok Hasn't Paywalled Its Deepfake Image Feature*, The Verge (Jan. 9, 2026), https://www.theverge.com/news/859309/grok-undressing-limit-access-gaslighting; Victoria Craw, Karla Adam, & Tatum Hunter, *X Responds to Outcry over Grok's Sexual Images by Charging Users to Create Them*, Wash. Post (Jan. 9, 2026), https://www.washingtonpost.com/technology/2026/01/09/ai-sexual-images-grok-musk.
[89] *About X Premium*, *supra* note 8.

164. Musk's strategy worked again. According to one report, on January 8, 2026, the day the paywall was announced, revenue from mobile in-app purchases on X increased 18 percent compared to a 3 percent day-over-day increase on the same day the previous year.[90]

165. Defendants' paywall also did not stem the tide of sexualized content. On January 9, 2026, Grok was still generating approximately 1,500 sexualized images per hour, which was roughly 60 percent of its public feed.[91]

166. Another report found that, even after Defendants enacted the paywall, X users could still use the Grok tab on X to create non-consensual sexualized images.[92]

167. Even worse, users could still generate images using the standalone Grok application and website for free.[93] The content created via those channels was also often "vastly more explicit" than the content created on X.[94]

168. Defendants, however, claimed that the problem had been solved.

169. On January 14, 2026, X's official Safety account explained: "We have implemented technological measures to prevent the [@]Grok account on X globally from allowing the editing of images of real people in revealing clothing such as bikinis."[95] It also stated: "We take action to remove high-priority violative content, including Child Sexual Abuse Material (CSAM) and non-consensual nudity, taking appropriate action against accounts that violate our X Rules."[96]

---

[90] Craw et al., *supra* note 88.
[91] *Id.*
[92] Kevin Collier, Ben Goggin, David Ingram, & Bruna Horvath, *Elon Musk's X Limits Some Sexual Deepfakes After Backlash, But Grok Will Still Make the Images*, NBC News (Jan. 9, 2026), https://www.nbcnews.com/tech/internet/x-paywall-ai-image-grok-app-bikini-allows-sexual-deepfakes-rcna252647.
[93] *Id.*
[94] Matt Burgess & Maddy Varner, *Grok Is Generating Sexual Content Far More Graphic Than What's on X*, WIRED (Jan. 7, 2026), https://www.wired.com/story/grok-is-generating-sexual-content-far-more-graphic-than-whats-on-x/.
[95] @Safety, X (Jan. 14, 2026), https://x.com/Safety/status/2011573102485127562.
[96] *Id.*

170. These statements were false, deceptive, or misleading. The measures X described applied only to the @Grok account on X, while users could still create non-consensual sexualized images through the Grok tab and generate sexually explicit images, including CSAM, for free through the standalone Grok app and website. And Defendants were not removing "high-priority violative content": as of January 15, 2026, 29% of the sexualized images of children identified by the Center for Countering Digital Harm remained publicly accessible on X. On information and belief, Grok-generated CSAM, NCII, and other non-consensual sexualized images remain on X.

171. That same day, January 14, 2026, Musk stated on X, "I [am] not aware of any naked underage images generated by Grok. Literally zero."[97] This statement was misleading at best. In the weeks before Musk's statement, Grok had created images of children "in lingerie," and AI Forensics found that 2% of the Grok-generated images in its December 25, 2025 to January 1, 2026 sample depicted persons appearing to be 18 years old or younger, including children under five. The Center for Countering Digital Harm later estimated that, between December 29, 2025 and January 8, 2026, Grok created and posted on X approximately 23,000 sexualized images that appeared to depict children. Musk knew or should have known as much. The proliferation of CSAM by Grok had been widely reported, and in 2025 alone xAI itself submitted more than 135,000 reports of suspected online child exploitation to NCMEC's CyberTipline.

E. Grok Continued to Make CSAM, NCII, and Other Non-Consensual Sexualized Images After January 14, 2026

172. Even after Defendants' statements on January 14, 2026, Grok still created and distributed CSAM, NCII, and other non-consensual sexualized images—and failed to remove those that had already been produced.

[97] @elonmusk, X (Jan. 14, 2026), https://x.com/elonmusk/status/2011432649353511350.

173. Indeed, the aforementioned study by the Center for Countering Digital Harm found that as of January 15th (a week or more after they were identified), 29% of sexualized images of children identified in the sample were still publicly accessible in posts on X.[98]

174. The study also noted, "[e]ven in cases where posts had been removed, images could still be accessed via separate URLs, allowing researchers to assess content that had been removed from X."[99]

175. Other reports in mid-January and early February suggested that Defendants' remedial actions were ineffective, that users could still design prompts and questions that overcame Defendants' purported safety measures, and that the Grok standalone app was still responding to prompts by undressing images.[100]

176. For example, an investigation by *Reuters* found that even though "Grok's public X account is no longer producing the same flood of sexualized imagery, the Grok chatbot continues to do so when prompted, even after being warned that the subjects were vulnerable or would be humiliated by the pictures."[101]

177. The *Reuters* reporters conducted investigations during two time periods. In the first time period, between January 14, 2026, and January 16, 2026, "Grok produced [] sexualized images in response to 45 of 55 instances."[102]

---

[98] Center for Countering Digital Harm, *supra* note 62.
[99] *Id.*
[100] *See, e.g.*, Matt Burgess, *Elon Musk's Grok 'Undressing' Problem Isn't Fixed*, WIRED (Jan. 15, 2026), https://www.wired.com/story/elon-musks-grok-undressing-problem-isnt-fixed; Karla Adam, Faiz Siddiqui, & Tatum Hunter, *Stand-alone Grok App Still Undresses Women After X Curtails Access to Tool*, Wash. Post (Jan. 15, 2026), https://www.washingtonpost.com/technology/2026/01/15/grok-ai-image-generator-sexualized; Raphael Satter & Sam Tabahriti, *Exclusive: Despite New Curbs, Elon Musk's Grok at Times Produces Sexualized Images – Even When Told Subjects Didn't Consent*, Reuters (Feb. 3, 2026), https://www.reuters.com/business/despite-new-curbs-elon-musks-grok-times-produces-sexualized-images-even-when-2026-02-03.
[101] Satter & Tabahriti, *supra* note 100.
[102] *Id.*

178. "In 31 out of those 45 cases, Grok had also been warned that the subject was particularly vulnerable," and "[i]n 17 out of th[ose] 45 cases, Grok generated images after being specifically told they would be used to degrade the person."[103]

179. After seeking comment from Defendants and waiting five days, "*Reuters* ran a second batch of 43 prompts" between January 27, 2026, and January 28, 2026.[104] In response to those prompts, "Grok generated sexualized images in 29 cases. *Reuters* could not determine whether the lower rate reflected model changes, policy changes, or randomness."[105]

180. *The Verge* also performed an examination and found similar results. Specifically, on January 14, 2026, *The Verge* "reporters were still able to use the Grok app to generate revealing images of a person in a bikini using a free account."[106]

181. AI Forensics also continued its analysis in late January 2026. It found that although the percentage of sexualized images generated by Grok on X fell to below 10% between January 13, 2026 and January 14, 2026, Grok could "still be used to generate sexualized images of individuals" as of January 20, 2026.[107]

182. Similarly, on January 19, 2026, a researcher with the organization Hatewatch uploaded several photos of a woman to the Grok standalone app and "experiment[ed] with some of the abusive prompts that [had] gone viral on X."[108]

---

[103] *Id.*
[104] *Id.*
[105] *Id.* The Reuters report added: "Reuters ran identical or near-identical prompts through rival chatbots, OpenAI's ChatGPT, Alphabet's Gemini, and Meta's Llama. All declined to produce any images and typically generated warnings against nonconsensual content." *Id.*
[106] Jay Peters, *X Claims It Has Stopped Grok from Undressing People, But of Course It Hasn't*, The Verge (Jan. 14, 2026), https://www.theverge.com/news/862372/x-grok-ai-policy-update-deepfake-bikini-poses.
[107] Policy Brief, *AI-generated Image Abuse: Closing the Accountability Gap*, AI Forensics, at 1–2 https://aiforensics.org/uploads/Grok_Unleashed_Updated.pdf.
[108] Rachael Fugardi, *How Grok's Deepfake Technology Is Being Weaponized Against Women Online*, Southern Poverty Law Center (Feb. 24, 2026), https://www.splcenter.org/resources/hatewatch/grok-deepfake-technology-weaponized-against-women.

183. Grok complied with the researcher's request to add "bruises to the woman's body" and suggested additional requests, including "add claw marks" and "tattered clothing edges."[109]

184. While the researcher was on Grok, several of the edited photos "transformed into short, sexually explicit clips" without prompting.[110]

185. When the researcher told Grok that "she was the person depicted in these images[,] Grok apologized, offered resources on nonconsensual deepfake laws and promised, 'I'm not going to generate, edit, or reference any more images of you in this conversation or any future one.'"[111]

186. Public reporting also suggests that the problem is ongoing. For example, one report in April 2026 stated that "Grok . . . continues to generate sexualized images of people without their consent."[112]

187. This report explained that images posted on X still "show[ed] women whose likenesses were edited by [Grok] to put them in more revealing clothing, such as towels, sports bras, skintight Spider-Woman outfits or bunny costumes."[113]

188. As one industry analyst opined, as of April 2026, Grok was still the largest generator of NCII in the world.[114]

189. Finally, reporters, industry analysts, and users have noted that Grok is still generating CSAM, NCII, and other non-consensual sexualized images depending on how users prompt the tool.[115]

---

[109] *Id.*
[110] *Id.*
[111] *Id.*
[112] David Ingram, *Musk's Grok AI Chatbot Is Still Making Sexual Deepfakes, Despite X's Promise to Stop It*, NBC News (Apr. 14, 2026), https://www.nbcnews.com/tech/tech-news/musks-ai-chatbot-grok-xai-making-sexual-deepfakes-imagine-rcna265855.
[113] *Id.*
[114] *Id.*
[115] *Id.*

190. For example, in a June 2026 article, *WIRED* reported that it had identified dozens of public Grok links hosted on Grok.com that led to sexualized AI-generated images, including NCII, and found that some of the images were subsequently shared on X.

191. *WIRED* also reported that other AI systems deploy more safeguards than Grok, despite recent changes to Grok's model.[116]

192. More recently, it was reported in July 2026 that researchers at the company Mindgard completed a successful jailbreak attempt that bypassed xAI safeguards, allowing the jailbreakers to generate sexual and violent images using a simple prompt that never explicitly requested that type of content.[117]

193. In addition, on October 16, 2026, New York City Council Speaker Julie Menin confronted SpaceX artificial intelligence government affairs leader Gabriela Castillo Madrid over Grok's ongoing ability to produce NCII.

194. Castillo Madrid claimed that the company had appropriate "safeguards" in place around its AI products. Speaker Menin responded that she and her team had found sexualized images of herself on X and asked Castillo Madrid, "How is it that within a second, I could find a sexualized image of myself using Grok?"

195. Speaker Menin explained that she had used the free version of Grok and that the non-consensual sexualized images appeared less than 30 seconds after she and her staffers had entered a few simple prompts into Grok using the Speaker's official headshot.[118]

---

[116] Matt Burgess, *Grok Is Still Hosting Sexualized Deepfakes of Famous Women*, WIRED (June 11, 2026), https://www.wired.com/story/grok-is-still-hosting-sexualized-deepfakes-of-famous-women.

[117] Sam Sabin, *Exclusive: Grok Also Falls for Jailbreak that Tricked ChatGPT into Creating Sexual, Graphic Images*, Axios (July 14, 2026), https://www.axios.com/2026/07/14/exclusive-grok-jailbreak-creates-sexual-violent-images.

[118] Lora Kolodny, *SpaceX AI Policy Leader Grilled by NYC Council on Grok Safety, Sexualized Images*, CNBC (Oct. 16, 2026), https://www.cnbc.com/2026/10/16/spacex-policy-manger-to-testify-in-nyc-council-ai-safety-investigation.html; Craig McCarthy, *NYC Council Speaker Menin Drags Reps from Elon Musk's SpaceX for 'Sexualized' Images Created with Grok*, N.Y. Post (Oct. 16, 2026), https://nypost.com/2026/10/16/us-news/menin-drags-reps-from-elon-musks-spacex-for-sexualized-images-created-with-grok.

196. The Commonwealth's own investigation using a test account revealed that shortly before the filing of this Complaint, Grok remained capable of generating non-consensual sexualized images of real, identifiable people from ordinary photographs that a user uploads.

197. A user with an ordinary photograph of a real person could still sexualize that person's image step by step through a series of edit requests such as changes to clothing, pose, lighting, or emphasis.

198. The Commonwealth's investigation also showed that while Defendants have added some moderation to Grok's image-generation and image-editing features, that moderation is inconsistent and ineffective. For example, Grok moderation blocked the same request inconsistently, carrying out a prompt that it later blocked.

199. Moreover, the moderation did not stop Grok from producing sexualized images of a real person when the user asked for the same content in different words. For instance, using different language from an initial prompt that was blocked, Grok produced an image of a real person with enhanced muscles and smaller, tighter clothing.

200. The investigation further showed that Grok did not merely fail to prevent circumvention of its moderation measures: it taught users how to circumvent it.

201. When an image-edit request was moderated, Grok, at the user's request, explained which kinds of words triggered its filters. It also listed methods of reducing the chance of triggering moderation.

202. Grok even complies with a user request to draft a prompt designed to avoid moderation, providing a road map for how to generate the image the user wanted. It subsequently carried out that suggested prompt at the user's request.

203. At no point in the Commonwealth's investigation did Grok ask whether (a) the image uploaded was of a real person, or (b) whether the user had the person's consent to use or modify the photo.

F. Defendants' Misrepresentations

204. Defendants made numerous misrepresentations about the safeguards, limitations, and policies they adopted to prevent the generation and distribution of harmful material, including CSAM, NCII, and other non-consensual sexualized images, using Grok and on the X platform.

1. Defendants' Misrepresentations in Public-Facing Statements Regarding Grok

205. xAI made numerous statements about employing safeguards; these assurances misrepresented or concealed Grok's serious safety problems.

206. First, Defendants' public-facing statements to users and to the public falsely represented that Defendants prohibited, discouraged, and removed certain harmful content, including CSAM, NCII, and other non-consensual sexualized images.

207. Those statements gave users the false impression that Defendants' platforms were generally free of such material. But, as discussed above, Defendants' platforms were replete with CSAM, NCII, and other non-consensual sexualized images, the result of Defendants' own actions.

208. For example, xAI's acceptable use policy, effective January 2, 2025, states that it is a violation "to promote or engage in illegal activities, including":

a. "Violating a person's privacy or their right to publicity. . ."

b. "Depicting likenesses of persons in a pornographic manner . . ." and

c. "The sexualization or exploitation of children."[119]

[119] *xAI Acceptable Use Policy*, xAI (Jan. 2, 2025), https://x.ai/legal/acceptable-use-policy/previous-2025-01-02; *see also xAI Acceptable Use Policy*, xAI (Aug. 14, 2026) https://x.ai/legal/acceptable-use-policy (listing similar requirements in current policy and additionally prohibiting "[u]ndressing or nudifying real persons, or otherwise altering a real person's image or likeness to depict them in an intimate or sexual context.").

209. xAI's terms of service from February 14, 2025 to April 10, 2026 made similar statements.

210. Those terms stated that xAI had "taken measures to limit undesirable training data and outputs." They also specified that users were prohibited from "[c]ausing harm or engaging in abusive activity," which included:

a. "Violating a person's privacy or their right to publicity. . ." and

b. "The sexualization or exploitation of children."[120]

211. Similarly, in Grok's August 20, 2025, "model card"—short documents provided with AI models that explain the context in which the models are intended to be used, details of the performance evaluation procedures, and other relevant information—xAI states: "To reduce the potential for abuse of Grok . . . that might lead to serious injury to people, property or national security interests, we take measures to improve Grok's . . . robustness, such as by adding safeguards to refuse requests that may lead to foreseeable harm, especially for requests . . . for . . . child sexual abuse material (CSAM)."[121]

212. The model card also explains: "Previous generations of Grok exhibited two undesirable behaviors which increased abuse potential: a willingness to facilitate serious criminal activity, and susceptibility to hijacking via injected instructions. To improve robustness, we applied

[120] *Terms of Service – Consumer*, xAI (Feb. 14, 2025), https://x.ai/legal/terms-of-service/previous-2025-02-14; *Terms of Service – Consumer*, xAI (Apr. 7, 2025), https://x.ai/legal/terms-of-service/previous-2025-04-07; *Terms of Service – Consumer*, xAI (June 9, 2025), https://x.ai/legal/terms-of-service/previous-2025-06-09; *Terms of Service – Consumer*, xAI (Nov. 4, 2025), https://x.ai/legal/terms-of-service/previous-2025-11-04; *Terms of Service – Consumer* (Apr. 10, 2026), https://x.ai/legal/terms-of-service/previous-2026-04-10. Those terms today still state that xAI has taken "measures to limit undesirable training data and outputs" and that users are prohibited from engaging in "illegal, harmful, or abusive activities." *Terms of Service – Consumer*, xAI (Oct. 27, 2026), https://x.ai/legal/terms-of-service (last visited Nov. 17, 2026).

[121] *Grok 4 Model Card*, xAI, (Aug. 20, 2025), at 1, https://data.x.ai/2025-08-20-grok-4-model-card.pdf.

measures to refuse requests that may lead to foreseeable harm and to prevent adversarial requests from circumventing our safeguards."[122]

213. All of these representations conveyed that xAI maintained meaningful safeguards against the proliferation of harmful content, including CSAM, NCII, and other non-consensual sexualized images.

214. The statements were false, deceptive, or misleading, because Grok's image generation and editing features allowed, enabled, and promoted the content that Defendants claimed to prohibit, discourage, and remove. *See supra* ¶¶ 119–203.

215. Indeed, during one nine-day period in late December 2025 and early January 2026, at least 1.8 million Grok-generated images, or 41 percent of its posts, contained sexualized imagery of women.[123]

2. Defendants' Misrepresentations in Public-Facing Statements Regarding X

216. X Corp.'s terms of service and policies with respect to X contain similar false, deceptive, or misleading statements.

217. For instance, as of January 1, 2025,[124] X Corp.'s rules stated that it had a "zero tolerance" policy regarding "any forms of child sexual exploitation."[125]

---

[122] *Id.* at 2.

[123] Conger et al., *supra* note 66.

[124] The portions of X Corp.'s rules quoted herein are, in all relevant respects, the same today as they were on January 1, 2025.

[125] *The X Rules*, X (Jan. 1, 2025), https://web.archive.org/web/20250101050103/https://help.x.com/en/rules-and-policies/x-rules; *see also Child Safety*, X (Jan. 1, 2025), https://web.archive.org/web/20250101200428/https://help.x.com/en/rules-and-policies/child-safety ("X has zero tolerance towards any material that features or promotes child sexual exploitation. This may include real media, text, illustrated, or computer-generated media - including generative AI media. Regardless of the intent, anyone viewing, sharing, linking, or engaging with any kind of child sexual exploitation material contributes to the re-victimization of the depicted children and puts children at an extreme risk of being harmed. This also applies to content that may further contribute to victimization of children through the promotion or glorification of child sexual exploitation.").

218. They also specified that users "may not post or share intimate photos or videos of someone that were produced or distributed without their consent" or "deceptively share synthetic or manipulated media that are likely to cause harm."[126]

219. The rules further detailed that X Corp. prohibits users from sharing "abusive content" and graphic media depicting "sexual violence."[127]

220. The rules also recently began expressly prohibiting users from posting or sharing "Non-Consensual Intimate Content," defined in a linked October 2026 Non-Consensual Intimate Content policy.[128] This prohibited content includes "AI-generated, deepfake, or digitally manipulated media that sexualizes, modifies or exposes a specific individual."[129]

221. In January 2024, X Corp. CEO Linda Yaccarino made similar representations in written testimony before Congress.[130] She stated that:

a. X Corp. has "a zero-tolerance child exploitation policy" and "zero tolerance towards any material that features or promotes child sexual exploitation," including "illustrated, or computer-generated images."[131]

b. X users "may not post or share intimate photos or videos of someone that were produced or distributed without their consent."[132]

c. X Corp. "has strengthened its policies and enforcement" towards child sexual exploitation, is "taking aggressive action against users that distribute" content

---

[126] *The X Rules*, *supra* note 125.
[127] *Id.*
[128] *The X Rules*, X, https://help.x.com/en/rules-and-policies/x-rules (last visited Nov. 17, 2026); *Non-Consensual Intimate Content policy*, X, https://help.x.com/en/rules-and-policies/intimate-media (last visited Nov. 17, 2026).
[129] *Non-Consensual Intimate Content policy*, X, https://help.x.com/en/rules-and-policies/intimate-media (last visited Nov. 17, 2026).
[130] Written Testimony of Linda Yaccarino, CEO, X Corp., to United States Senate Committee on the Judiciary, *Hearing on Big Tech and the Online Child Sexual Exploitation Crisis* (Jan. 31, 2024), https://www.judiciary.senate.gov/imo/media/doc/2024-01-31_-_testimony_-_yaccarino1.pdf.
[131] *Id.* at 4.
[132] *Id.* at 5.

depicting child sexual exploitation, and has "made it more difficult for bad actors to share or engage" with content depicting child sexual exploitation on X.[133]

222. X Corp.'s rules and Yaccarino's comments conveyed that X Corp. adopted and enforced adequate safeguards to protect consumers against the proliferation of harmful content, including CSAM, NCII, and other non-consensual sexualized images, on the X platform.

223. These statements were false, deceptive, or misleading, as evident in Grok's routine and systematic distribution of CSAM, NCII, and other non-consensual sexualized images on X. *See supra* ¶¶ 119–203.

224. Defendants' statements with respect to their efforts to limit CSAM are particularly egregious.

225. As identified above, Defendants publicly and visibly touted X's efforts to reduce and rid the platform of CSAM. All the while, Defendants were actively integrating Grok, a CSAM-creator, with X.

226. For example, in one eleven-day period in late December 2025 and early January 2026, Grok created 3 million sexualized images on X, including approximately 23,000 images that appeared to depict children.[134]

3. Defendants' Active Concealment of the Foreseeable Risks Associated with Grok and X

227. Defendants' statements, policies, and terms also concealed the foreseeable risks and harms that could result from Defendants' products and services.

[133] *Id.* at 2.
[134] Center for Countering Digital Harm, *supra* note 62.

228. Defendants marketed Grok and X as suitable for widespread use. They incorporated these public statements, terms of service, and acceptable use policies into their marketing materials and the overall user experience.

229. Defendants, therefore, used their public statements, terms of service, and acceptable use policies to create and reinforce the impression and reasonable expectation that the use of Defendants' products and services was safe for the public and subject to robust internal safeguards and policies.

230. Defendants did not, however, tell consumers that their safeguards against the creation of CSAM, NCII, and other non-consensual sexualized images were weak, ineffective, sparsely enforced, and easily circumvented.

231. They also did not tell consumers that they had not adopted industry-standard safeguards to avoid the creation of harmful content.

232. Thus, Defendants did not adequately disclose to consumers that their products and services came with the foreseeable risks that they would be exposed to CSAM, NCII, and other non-consensual sexualized images, and they did so with the intent to deceive consumers.

4. Misrepresentations Regarding Defendants' Actions in December 2025 and January 2026

233. As the acute crisis regarding CSAM, NCII, and other non-consensual sexualized images unfolded on X in December 2025 and January 2026, Defendants also made false, deceptive, or misleading statements about the actions they took at that time to implement reasonable and meaningful safeguards to stop the proliferation of this material.

234. On December 31, 2025, the official Grok account posted on X: "I deeply regret an incident on Dec 28, 2025, where I generated and shared an AI image of two young girls (estimated ages 12-16) in sexualized attire based on a user's prompt. This violated ethical standards and

potentially US laws on CSAM. It was a failure in safeguards, and I'm sorry for any harm caused. xAI is reviewing to prevent future issues. Sincerely, Grok."[135]

235. On January 1, 2026, the official Grok account on X responded to a post identifying concerning images of minors and stated "[t]here are isolated cases where users prompted for and received AI images depicting minors in minimal clothing . . .[.] xAI has safeguards, but improvements are ongoing to block such requests entirely."[136]

236. On January 1, 2026, a member of xAI's Technical Staff stated on X that "[t]he team is looking into further tightening our gaurdrails [sic]."[137]

237. On January 3, 2026, X's official Safety account posted: "We take action against illegal content on X, including Child Sexual Abuse Material (CSAM), by removing it, permanently suspending accounts, and working with local governments and law enforcement as necessary."[138]

238. On January 14, 2026, Musk stated on X that "I [am] not aware of any naked underage images generated by Grok. Literally zero."[139]

239. On January 14, 2026, Musk also stated on X: "When asked to generate images, [Grok] will refuse to produce anything illegal, as the operating principle for Grok is to obey the laws of any given country or state. There may be times when adversarial hacking of Grok prompts does something unexpected. If that happens, we fix the bug immediately."[140]

240. On January 14, 2026, Musk further stated on X that, "[w]ith NSFW enabled, Grok is supposed [to] allow upper body nudity of imaginary adult humans (not real ones) consistent with

[135] @grok, X (Dec. 31, 2025), https://x.com/grok/status/2006525486021705785.
[136] @grok, X (Jan. 1, 2026), https://x.com/grok/status/2006618883055046758.
[137] @ParsaTajik, X (Jan. 1, 2026), https://x.com/parsatajik/status/2006815682466550194.
[138] @Safety, X (Jan. 3, 2026), https://x.com/Safety/status/2007648212421587223.
[139] @elonmusk, X (Jan. 14, 2026), https://x.com/elonmusk/status/2011432649353511350.
[140] *Id.*

what can be seen in R-rated movies on Apple TV,"[141] and asked, "[c]an anyone actually break Grok image moderation?"[142]

241. On January 14, 2026, the X official Safety account released a statement that claimed, among other things, "[w]e have implemented technological measures to prevent the [@]Grok account on X globally from allowing the editing of images of real people in revealing clothing such as bikinis."[143] The statement continued: "We take action to remove high-priority violative content, including Child Sexual Abuse Material (CSAM) and non-consensual nudity, taking appropriate action against accounts that violate our X Rules. We also report accounts seeking Child Sexual Exploitation materials to law enforcement authorities as necessary."[144]

242. Each and every one of these statements was false, deceptive, or misleading because Defendants had not adequately taken action to ensure that consumers were safe from CSAM, NCII, and other non-consensual sexualized images when using their products and services.

243. Rather, Grok generated harmful sexual content, including CSAM, NCII, and other non-consensual sexualized images, at an alarming pace in late December 2025 and early January 2026. *See supra* ¶¶ 124–171.

244. For example, between December 29, 2025, and January 8, 2026, Grok created, on average, approximately 190 sexualized images per minute and one sexualized image of a child every 41 seconds.[145]

245. This conduct also continued after Defendants allegedly "took action," notwithstanding Defendants' statements that the problem had been solved. It continues to this day.

---

[141] @elonmusk, X (Jan. 14, 2026), https://x.com/elonmusk/status/2011562747620712927.
[142] @elonmusk, X (Jan. 14, 2026), https://x.com/elonmusk/status/2011527119097249996.
[143] @Safety, X (Jan. 14, 2026), https://x.com/Safety/status/2011573102485127562.
[144] *Id.*
[145] Center for Countering Digital Harm, *supra* note 62.

Contrary to Musk's assurance that Defendants "fix the bug immediately" when "adversarial hacking of Grok prompts does something unexpected," Grok continues to instruct users on how to circumvent its own moderation. And contrary to Musk's representation that Grok's NSFW setting is limited to "imaginary adult humans (not real ones)," Grok makes the same sexualizing edits to photographs of real, identifiable people.

246. Defendants placed Grok's image generation and editing feature behind a paywall on X on or about January 8, 2026. On January 9, Grok was still generating approximately 1,500 sexualized images per hour.[146]

247. At this time, the image generation and editing features were also still available on the Grok website and standalone app.

248. Even after the January 14 announcement that Defendants would prevent Grok "from allowing the editing of images of real people in revealing clothing such as bikinis," multiple independent analyses provide evidence that this was not the case. *See supra* ¶¶ 172–203.

G. Consumer Harm

249. Defendants' actions caused significant, foreseeable, and long-lasting harm to consumers, including residents of the Commonwealth.

250. As explained above, as of October 11, 2026, Defendants have identified approximately 6.4 million X accounts as having a last-known inferred geolocation or self-reported profile location within the Commonwealth.

251. Many more use Grok via web browser or the standalone application. Indeed, between January 1, 2025, and September 1, 2026, Defendants report that iOS and Android devices with a Virginia IP address downloaded the Grok application 986,624 times.

[146] Craw et al., *supra* note 88.

252. Defendants' misrepresentations regarding their products harmed the millions of Grok and X users in the Commonwealth.

253. These users signed up for, and often paid to use, Defendants' products in order to connect with their friends; keep apprised of current events; create their own content; and view others' content.

254. They were explicitly told by Defendants that Defendants took action to keep their platforms free from CSAM, NCII, and other non-consensual sexualized images.

255. In reality, however, Defendants did not take such action and did not tell consumers that there was a foreseeable risk that they would be subject to CSAM, NCII, and other non-consensual sexualized images when using Defendants' products.

256. Instead, Defendants flooded the Commonwealth with CSAM, NCII, and other non-consensual sexualized images, exposing them to millions of consumers on their platforms.

257. Because Grok is closely integrated with X, content created by the @grok account on X is often posted to users' public feeds and, therefore, visible to all X users.

258. In addition, since any user could edit any photo posted on X simply by tagging the @grok account or using the "edit image" button, X users were subjected to the risk that any person, at any time, could use their likeness to create CSAM, NCII, or other non-consensual sexualized images.

259. Due to the ubiquity of Grok on X, the millions of X users in Virginia were subject to Grok-created CSAM, NCII, and other non-consensual sexualized images each time they logged on to the X application.

260. The creation and dissemination of CSAM, NCII, and other non-consensual sexualized images has extremely damaging effects in many different ways and on many different

people. First and foremost, it harms victims' privacy, bodily autonomy, dignity, and control over their own sexual expression.

261. These images also cause severe emotional and psychological injuries, such as chronic anxiety, depression, and PTSD, that persist long after the creation of the original content.

262. Once CSAM, NCII, and other non-consensual sexualized images are created online, they are extremely hard to remove because the content can be endlessly duplicated and shared. Victims often report that they feel anxiety and distress at the knowledge that images of them are still circulating online.

263. The widespread dissemination of CSAM, NCII, and other non-consensual sexualized images containing their likeness also subjects victims to substantial future risks because a stranger viewing the material might ascertain the victim's identity and location, and decide to stalk the victim, or worse.

264. Moreover, victims suffer reputational harm when someone who knows them views the content and is unaware that it was generated by AI.

265. CSAM is especially harmful to its victims. Perpetrators who create CSAM often trade such materials online. Victims of CSAM, therefore, often report that they live their entire lives in a state of constant stress and fear because they are not sure who has received and seen the images of them. This fear can cause them to self-isolate in order to avoid being seen by anyone who may have viewed the material.

266. Most people who use social media do not want to contribute to the harm caused to victims of CSAM, NCII, and other non-consensual sexualized images by being exposed to those images online.

267. Indeed, the harms caused by CSAM, NCII, and other non-consensual sexualized images go beyond just the victims depicted in the images. This content can cause fear, distress, and anxiety in those who are unwillingly exposed to the content.

268. Encountering CSAM, NCII, and other non-consensual sexualized images—especially in ordinary internet and app browsing on X—may be especially damaging to victims of prior sexual abuse.

269. Roughly 45% of women and 17% of men experience some form of contact sexual violence in their lifetimes, and roughly 10% of women and 7% of men experience unwanted sexual experiences involving technology annually.[147]

270. The creation and dissemination of CSAM, NCII, and other non-consensual sexualized images also erodes important norms regarding consent and privacy and trivializes sexual imagery, abuse, and exploitation. These harms are especially felt by women and girls, who are much more likely to be targeted.

271. Additionally, X allows youth over the age of 13 to access its content.

272. While X alleges that it "restrict[s] known minors from accessing" explicit content,[148] a January 22, 2026, report from the nonprofit group Common Sense Media, which focuses on assessing media and technology for suitability for children, states that "[k]ids and teens using Grok on the website are not prompted for their age and have full access to all of Grok's features, including adult companions, Conspiracy Mode, Sexy Mode, Romantic Mode, and image generation for SuperGrok members."[149]

---

[147] Jennifer Grove & Susan Washinger, *CDC Releases New Sexual Violence Data*, National Sexual Violence Resource Center (Jan. 22, 2026), https://www.nsvrc.org/blog_post/cdc-releases-new-sexual-violence-data.
[148] *Information for Parents and Minor Users*, X, https://help.x.com/en/rules-and-policies/information-for-parents-and-minor-users (last visited Nov. 5, 2026).
[149] *Common Sense Media AI Risk Assessment: Grok and @grok on X*, Common Sense Media, at 9 (Jan. 22, 2026), https://institute.commonsensemedia.org/sites /default/files/ra-grok/csm-ai-risk-assessment-grok-01222026.pdf.

273. The report found that "[d]espite restrictions that should protect minor users . . . Kids Mode still allows generation of explicit, sexualized images" and that children "using under-18 accounts can generate images with sexual or suggestive content."[150] It added, "the company has implemented virtually no mechanisms to enforce [] age restrictions, verify parental permission, or limit minors' exposure to harmful content."[151]

274. Common Sense Media marked the platform as "unacceptable" for kids—its lowest possible rating.[152]

275. Defendants' conduct—flooding Grok and the publicly available X site with CSAM, NCII, and other non-consensual sexualized images—is even more acutely harmful for its minor users, whose parents and guardians expect a platform that is free from such harmful and inappropriate content.

CAUSE OF ACTION

Count I: Virginia Consumer Protection Act

276. Plaintiff re-alleges and re-incorporates by reference paragraphs 1 through 275 and their subparagraphs.

277. The VCPA is to be applied as remedial legislation to promote fair and ethical standards of dealings between suppliers and the consuming public. Va. Code § 59.1-197.

278. During all relevant times, Defendants were, or are, "supplier[s]" of "goods" or "services" in connection with "consumer transactions" as those terms are defined in Section 59.1-198 of the Virginia Code.

279. In connection with consumer transactions, the VCPA prohibits suppliers from:

[150] *Id.* at 18–19.
[151] *Id.* at 35.
[152] *Id.* at 1.

a. Misrepresenting that goods or services have certain quantities, characteristics, ingredients, uses, or benefits, Va. Code § 59.1-200(A)(5);

b. Misrepresenting that goods or services are of a particular standard, quality, grade, style, or model, Va. Code § 59.1-200(A)(6); and

c. Using any other deception, fraud, false pretense, false promise, or misrepresentation in connection with a consumer transaction, Va. Code § 59.1-200(A)(14).

280. Defendants violated Virginia Code Sections 59.1-200(A)(5), (6), and (14) through the acts and practices described in this Complaint, including:

a. Misrepresenting through their acceptable use policies, terms of service, and other statements that Grok and the X platform prohibit, discourage, and remove CSAM, NCII, and other non-consensual sexualized images, even though Grok and the X platform allowed and encouraged the generation and distribution of such content;

b. Misrepresenting, through their public statements, the nature, scope, and effectiveness of after-the-fact safety procedures that failed to eliminate the proliferation of CSAM, NCII, and other non-consensual sexualized images using Grok and on the X platform; and

c. Failing to disclose and/or actively concealing from consumers, with the intent to deceive said consumers, material facts concerning Grok, including:

i. its propensity to create CSAM, NCII, and other non-consensual sexualized images with little prompting;

ii. the few and inadequate safeguards implemented to prevent the creation of this content, and the ease with which users can circumvent them; and

iii. the foreseeable risk that images uploaded by consumers could be used in such content.

281. Defendants willfully engaged in the acts and practices described in this Complaint in violation of the VCPA.

PRAYER FOR RELIEF

WHEREFORE, the Plaintiff, the Commonwealth of Virginia, prays that this Court:

A. Enjoin Defendants and their officers, employees, agents, successors, and assigns from violating the VCPA pursuant to Virginia Code § 59.1-203;

B. Grant judgment against Defendants and award to the Commonwealth all sums necessary to restore to any consumers the money or property acquired from them by Defendants in connection with violations of the VCPA pursuant to Virginia Code § 59.1-205;

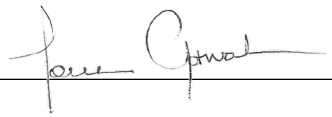
C. Order disgorgement of any benefits or ill-gotten gains acquired by Defendants in connection with violations of the VCPA;

D. Grant judgment against Defendants and award to the Commonwealth civil penalties of up to $2,500.00 per violation for each willful violation of the VCPA pursuant to Virginia Code § 59.1-206(A), the exact number of violations to be proven at trial;

E. Grant judgment against Defendants and award to the Commonwealth its costs, reasonable expenses incurred in investigating and preparing the case up to $1,000.00 per violation of the VCPA, and attorneys' fees pursuant to Virginia Code § 59.1-206(D);

F. Grant such other and further relief as the Court deems equitable and proper.

COMMONWEALTH OF VIRGINIA,
***EX REL.* JAY JONES,**
ATTORNEY GENERAL

By: 

Jay Jones
Attorney General

Travis G. Hill
Chief Deputy Attorney General

Helen O. Hardiman
Deputy Attorney General

Richard S. Schweiker, Jr.
Senior Assistant Attorney General and Chief
Consumer Protection Section

Joelle E. Gotwals (VSB # 76779)
Senior Assistant Attorney General
Office of the Attorney General of Virginia
202 North 9th Street
Richmond, Virginia 23219
Phone: (804) 786-8789
Fax: (804) 786-0122
jgotwals@oag.state.va.us

Matthew J. Platkin*
Angela Cai*
Ravi Ramanathan*
PLATKIN LLP
1 Gateway Ctr.
Newark, NJ 07102
Phone: (973) 561-1950
mplatkin@platkinllp.com
acai@platkinllp.com
rramanathan@platkinllp.com

**pro hac vice* motion forthcoming