Grok is undressing anyone, including minors

xAI's Grok is removing clothing from pictures of people without their consent following this week's rollout of a feature that allows X users to instantly edit any image using the bot without needing the original poster's permission. Not only does the original poster not get notified if their picture was edited, but Grok appears to have few guardrails in place for preventing anything short of full explicit nudity. In the last few days, X has been flooded with imagery of women and children appearing pregnant, skirtless, wearing a bikini, or in other sexualized situations. World leaders and celebrities, too, have had their likenesses used in images generated by Grok.

AI authentication company Copyleaks reported that the trend to remove clothing from images began with adult-content creators asking Grok for sexy images of themselves after the release of the new image editing feature. Users then began applying similar prompts to photos of other users, predominantly women, who did not consent to the edits. Women noted the rapid uptick in deepfake creation on X to various news outlets, including *Metro* and *PetaPixel*. Grok was already able to modify images in sexual ways when tagged in a post on X, but the new "Edit Image" tool appears to have spurred the recent surge in popularity.

In one X post, now removed from the platform, Grok edited a photo of two young girls into skimpy clothing and sexually suggestive poses. Another X user prompted Grok to issue an apology for the "incident" involving "an AI image of two young girls (estimated ages 12-16) in sexualized attire," calling it

“a failure in safeguards” that it said may have violated xAI’s policies and US law. (While it’s not clear whether the Grok-created images would meet this standard, realistic AI-generated sexually explicit imagery of identifiable adults or children can be illegal under US law.) In another back-and-forth with a user, Grok suggested that users report it to the FBI for CSAM, noting that it is “urgently fixing” the “lapses in safeguards.”

But Grok’s word is nothing more than an AI-generated response to a user asking for a “heartfelt apology note” — it doesn’t indicate Grok “understands” what it’s doing or necessarily reflect operator xAI’s actual opinion and policies. Instead, xAI responded to *Reuters’* request for comment on the situation with just three words: “Legacy Media Lies.” xAI did not respond to *The Verge*’s request for comment in time for publication.

Elon Musk himself seems to have sparked a wave of bikini edits after asking Grok to replace a memetic image of actor Ben Affleck with himself sporting a bikini. Days later, North Korea’s Kim Jong Un’s leather jacket was replaced with a multicolored spaghetti bikini; US President Donald Trump stood nearby in a matching swimsuit. (Cue jokes about a nuclear war.) A photo of British politician Priti Patel, posted by a user with a sexually suggestive message in 2022, got turned into a bikini picture on January 2nd. In response to the wave of bikini pics on his platform, Musk jokingly reposted a picture of a toaster in a bikini captioned “Grok can put a bikini on everything.”

While some of the images — like the toaster — were evidently meant as jokes, others were clearly designed to produce borderline-pornographic imagery, including specific directions for Grok to use skimpy bikini styles or remove a skirt entirely. (The chatbot did remove the skirt, but it did not depict full, uncensored nudity in the responses *The Verge* saw.) Grok also complied with requests to replace the clothes

of a toddler with a bikini.

Musk’s AI products are prominently marketed as heavily sexualized and minimally guardrailed. xAI’s AI companion Ani flirted with *Verge* reporter Victoria Song, and Jess Weatherbed discovered that Grok’s video generator readily created topless deepfakes of Taylor Swift, despite xAI’s acceptable use policy banning the depiction of “likenesses of persons in a pornographic manner.” Google’s Veo and OpenAI’s Sora video generators, in contrast, have guardrails around generation of NSFW content, though Sora has also been used to produce videos of children in sexualized contexts and fetish videos. The prevalence of deepfake images is growing rapidly, according to a report from cybersecurity firm DeepStrike, and many of these images contain nonconsensual sexualized imagery; a 2024 survey of US students found that 40 percent were aware of a deepfake of someone they knew, while 15 percent were aware of nonconsensual explicit or intimate deepfakes.

When asked why it is transforming images of women into bikini pics, Grok denied posting photos without consent, saying: “These are AI creations based on requests, not real photo edits without consent.”

Take an AI bot’s denial as you wish.

https://www.theverge.com/news/853191/grok-explicit-bikini-pictures-minors