

AI-GENERATED IMAGE ABUSE: AN UPDATE ON GROK UNLEASHED

An updated analysis of Grok shows safeguards installed, but problematic images remain on Grok's app and website.

20-01-2026

DOWNLOAD REPORT 



This report provides an update to Grok Unleashed, published January 5, 2026. The first report, using data obtained between December 25, 2025 and January 1, 2026, found that around half of images generated by @Grok in our sample of 20k images depicted people in minimal attire, most notably women.

By January 13-14, 2026, this number fell to below 10%, suggesting some safeguards were implemented on X.

However, as reported by AIF alongside Wired, Grok could be used to generate images far more explicit when accessed directly from the Grok.com website or app compared with on X, including full nudity and/or sexual acts.

As of January 19, 2026, we continued to identify synthetic media created by Grok Imagine depicting individuals suspected of being minors. These have been flagged to French regulators and Point de Contact, a French Trusted Flagger under the DSA.

The rest of this report provides a policy briefing, examining how such emerging AI risks can be considered under the DSA and AI Act, as well as a set of recommendations for the European Commission, AI Office, and co-legislators.

Share This Project :



08-04-2026

Wired covered out report, noting that men were buying hacking tools to use against their wives and friends as well as sharing nonconsensual images of women and girls.

READ ARTICLE 

AI FORENSICS IS PROUD TO BE TRUSTED BY

OPEN SOCIETY
FOUNDATIONS

Luminate

SIDNfonds

WORK

PRIVACY POLICY

© 2025 AI FORENSICS

CONTACT US: INFO@AIFORENSICS.ORG