

Grok AI e deepfake sessuali: il caso che mette X sotto accusa

Maria Cattini | 12/01/2026 | Intelligenza Artificiale

Basta caricare una foto e scrivere poche parole. Bikini, nudità, corpi alterati. A fine 2025 questa pratica è diventata virale su **X**, grazie a **Grok**, il sistema di intelligenza artificiale integrato nella piattaforma.

Non si è trattato di satira o sperimentazione creativa. Le immagini hanno coinvolto persone reali, spesso senza consenso. In alcuni casi, minori.

Il punto critico non è la tecnologia in sé, ma il contesto: una funzione pubblica, accessibile a chiunque, senza barriere efficaci.

Cosa faceva Grok, in concreto

Tra dicembre 2025 e i primi giorni di gennaio 2026, [Grok](#) ha consentito di modificare fotografie reali caricate dagli utenti. Bastava una richiesta testuale: cambiare abiti, accentuarne le forme, spogliare il soggetto.

La funzione risultava disponibile anche senza abbonamento. Nessun filtro realmente operativo bloccava richieste esplicite, neppure quando le immagini riguardavano ragazze molto giovani.

Questa dinamica ha prodotto una valanga di deepfake sessualizzati, facilmente condivisibili sulla piattaforma stessa.

A rendere il caso ancora più esplosivo è stata la reazione del fondatore di X, [Elon Musk](#).

Davanti ad alcune immagini generate con Grok – comprese parodie che coinvolgevano personaggi pubblici – Musk ha commentato con ironia, emoji e meme. Nessuna presa di distanza immediata. Nessun segnale di allarme.

Quel tono ha rafforzato l'idea di una sottovalutazione del problema, mentre le immagini circolavano senza controllo.

Dalle polemiche alle indagini

La risposta istituzionale è arrivata prima delle correzioni tecniche.

La Francia ha esteso un'indagine già aperta su X, coinvolgendo la procura di Parigi e i canali ufficiali per le segnalazioni di abusi online.

L'India ha imposto a X una revisione urgente di Grok, con un ultimatum di 72 ore. In caso di inadempienza, la piattaforma rischiava la perdita delle tutele legali riservate agli intermediari digitali.

La Commissione europea ha definito i contenuti "illegali e spaventosi", richiamando gli obblighi previsti dal **Digital Services Act**.

Nel Regno Unito, Ofcom ha chiesto chiarimenti formali, mentre il governo ha evocato sanzioni fino al 10% del fatturato globale di X, previste dall'Online Safety Act.

Anche Malesia e Brasile hanno avviato procedure autonome.

Il blocco parziale deciso da X

Il 9 gennaio 2026 X ha annunciato una limitazione: generazione e modifica delle immagini tramite Grok riservate agli abbonati paganti. Una mossa tardiva e incompleta. Le falle restano evidenti:

- la funzione “modifica immagine” integrata nei post continua a funzionare anche per chi non paga;
- l’app mobile e il sito grok.com permettono ancora la creazione di immagini dopo il login;
- la modalità “Spicy”, priva di filtri stringenti, risulta attiva su canali alternativi.

In altre parole, il problema non è stato rimosso. È stato solo spostato.

Il nodo centrale non riguarda una singola funzione mal progettata. Riguarda il modello di governance dell’AI sulle piattaforme social.

Quando un sistema generativo agisce su immagini reali, il confine tra creatività e abuso diventa sottilissimo. Senza controlli strutturali, la responsabilità non può essere scaricata sugli utenti.

Il [Garante Privacy italiano](#) ha parlato di rischio concreto per diritti e libertà fondamentali. Altri regolatori hanno espresso valutazioni simili.

Le indagini sono in corso. Le modifiche tecniche non hanno chiuso tutte le falle. Il modello di Grok non è stato ripensato alla radice.

Il caso X-Grok mostra quanto velocemente l’AI possa amplificare violazioni già note, rendendole di massa, istantanee, difficili da arginare.

Chi progetta questi sistemi non può più fingere sorpresa.

Entra nella community **OSINT & AI per tutti**:

- Newsletter: <https://coondivido.substack.com/>
- Telegram: <https://t.me/osintaipertutti>
- Telegram: <https://t.me/osintprojectgroup>

Basta caricare una foto e scrivere poche parole. Bikini, nudità, corpi alterati. A fine 2025 questa pratica è diventata virale su **X**, grazie a **Grok**, il sistema di intelligenza artificiale integrato nella piattaforma.

Non si è trattato di satira o sperimentazione creativa. Le immagini hanno coinvolto persone reali, spesso senza consenso. In alcuni casi, minori.

Il punto critico non è la tecnologia in sé, ma il contesto: una funzione pubblica, accessibile a chiunque, senza barriere efficaci.

Cosa faceva Grok, in concreto

Tra dicembre 2025 e i primi giorni di gennaio 2026, [Grok](#) ha consentito di modificare fotografie reali caricate dagli utenti. Bastava una richiesta testuale: cambiare abiti, accentuarne le forme, spogliare il soggetto.

La funzione risultava disponibile anche senza abbonamento. Nessun filtro realmente operativo bloccava richieste esplicite, neppure quando le immagini riguardavano ragazze molto giovani.

Questa dinamica ha prodotto una valanga di deepfake sessualizzati, facilmente condivisibili sulla piattaforma stessa.

A rendere il caso ancora più esplosivo è stata la reazione del fondatore di X, [Elon Musk](#).

Davanti ad alcune immagini generate con Grok – comprese parodie che coinvolgevano personaggi pubblici – Musk ha commentato con ironia, emoji e meme. Nessuna presa di distanza immediata. Nessun segnale di allarme.

Quel tono ha rafforzato l'idea di una sottovalutazione del problema, mentre le immagini circolavano senza controllo.

Dalle polemiche alle indagini

La risposta istituzionale è arrivata prima delle correzioni tecniche.

La Francia ha esteso un'indagine già aperta su X, coinvolgendo la procura di Parigi e i canali ufficiali per le segnalazioni di abusi online.

L'India ha imposto a X una revisione urgente di Grok, con un ultimatum di 72 ore. In caso di inadempienza, la piattaforma rischiava la perdita delle tutele legali riservate agli intermediari digitali.

La Commissione europea ha definito i contenuti "illegali e spaventosi", richiamando gli obblighi previsti dal **Digital Services Act**.

Nel Regno Unito, Ofcom ha chiesto chiarimenti formali, mentre il governo ha evocato sanzioni fino al 10% del fatturato globale di X, previste dall'Online Safety Act.

Anche Malesia e Brasile hanno avviato procedure autonome.

Il blocco parziale deciso da X

Il 9 gennaio 2026 X ha annunciato una limitazione: generazione e modifica delle immagini tramite Grok riservate agli abbonati paganti. Una mossa tardiva e incompleta. Le falle restano evidenti:

- la funzione "modifica immagine" integrata nei post continua a funzionare anche per chi non paga;
- l'app mobile e il sito grok.com permettono ancora la creazione di immagini dopo il login;
- la modalità "Spicy", priva di filtri stringenti, risulta attiva su canali alternativi.

In altre parole, il problema non è stato rimosso. È stato solo spostato.

Il nodo centrale non riguarda una singola funzione mal progettata. Riguarda il modello di governance dell'AI sulle piattaforme social.

Quando un sistema generativo agisce su immagini reali, il confine tra creatività e abuso diventa sottilissimo. Senza controlli strutturali, la responsabilità non può essere scaricata sugli utenti.

Il [Garante Privacy italiano](#) ha parlato di rischio concreto per diritti e libertà fondamentali. Altri regolatori hanno espresso valutazioni simili.

Le indagini sono in corso. Le modifiche tecniche non hanno chiuso tutte le falle. Il modello di Grok non è stato ripensato alla radice.

Il caso X-Grok mostra quanto velocemente l'AI possa amplificare violazioni già note, rendendole di massa, istantanee, difficili da arginare.

Chi progetta questi sistemi non può più fingere sorpresa.

Entra nella community **OSINT & AI per tutti**:

- Newsletter: <https://coondivido.substack.com/>
- Telegram: <https://t.me/osintaipertutti>
- Telegram: <https://t.me/osintprojectgroup>