# Private Images, Public Emergency: Explainer on Sexualised Deepfakes

While using technologies to carry out gender-based violence (GBV) is no longer a novel social harm, the emergence of Generative AI (Gen-AI) tools have certainly increased the ease with which perpetrators can now target women online. In the past, the crime of publishing, sharing or distributing non-consensual intimate imagery (NCII)* of a person online would require perpetrators to have some way of accessing pre-existing intimate images. However, Gen-AI tools remove even this barrier, making it nearly seamless for perpetrators to 'undress' women online with a simple click of their mouse. Such tools are now easily available, cost little to nothing and the process takes minimal time.


In December 2025, users on Elon Musk-owned microblogging site X, started tagging GrokAI, making requests to 'nudify' images of women and children – including depicting them in transparent clothing and bikinis.

These targeted not only regular users but also celebrities and key politicians, like the Deputy Prime Minister of Sweden and the former Vice President of the United States of America.

Research found that Grok produced an estimated 3 million sexualised images over a period of eleven days, resulting in global outrage and renewed regulatory attention on the platform worldwide.

Harms arising from such content go beyond psychological or personal repercussions. Victims — encompassing both urban women as well as women belonging to more rural, marginalised backgrounds — can suffer professional setbacks and direct physical violence, where cultural contexts catalyse honour-based crimes. Given the relative novelty of this social harm, many technological dimensions, and varied actors involved in the spread of such content, regulation around the world is still playing catch-up.

*'Non-consensual intimate imagery' is a term encompassing broad range of content including images and videos of an explicit nature, that are created and/or distributed without the subject's consent. In this explainer, we use 'sexualised deepfakes', the term widely used in media, policy, and public discourse when discussing AI-generated intimate content.

# Main Challenges

## THE VIRALITY PROBLEM

Targeted applications such as ClothOff, which have soared in popularity in the past few years, have amplified the ease of creating, sharing and duplicating such content online. These services are often advertised through social media platforms and available on app-stores. Given the speed with which such content goes viral, the reputational and psychological damage to victims is already done before legal remedies can even intervene. While taking the images and videos down can limit the continuing harm to a certain extent, the perceived '**immortality' of content on the internet and the delay in takedowns** is a key challenge when tackling NCII.

> The adage of everything on the internet is forever and cannot be deleted is rooted in the idea that content, once uploaded online, can be copied, shared and reproduced endlessly, even if the original piece has been deleted.

The virality problem is also rooted in online platforms' preexisting issues with misinformation, hate speech, and their business models incentivising increased traffic to their platforms as opposed to protecting individuals whose rights are being infringed on.

## IDENTIFIABLE VICTIMS, INVISIBLE PERPETRATORS

End-to-end encrypted communication apps promise and deliver private messaging services. While a win for privacy and relief from surveillance, bad actors exploit the encryption and anonymity afforded to them through these platforms, rendering it nearly impossible to identify a perpetrator in the fight against NCII. The dissemination of sexual deepfakes is therefore simplified by the architecture of these platforms, where the ease of creating large-scale closed groups on platforms like Telegram, creates a thriving environment for the diffusion of NCII. Perpetrators for NCII also hide behind VPNs* and proxy servers, making it difficult to track the location and request accurate information on their identities to hold them culpable.

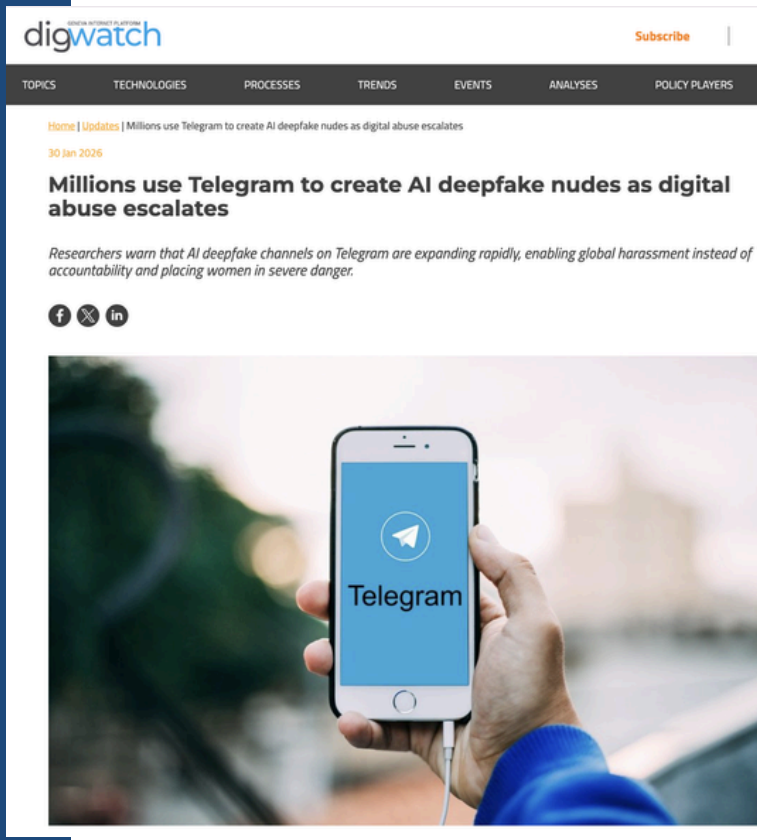
A remedy victims often rely on is requesting search engines to 'de-index' their content hosted on offending websites, which means such content no longer shows up on search results. 'De-indexing' can potentially reduce traffic to websites hosting such sexual deepfakes, but this is an imperfect solution.

This does not preclude perpetrators from reuploading such content, and puts the burden, once again, on victims to repeatedly reach out to search engines to remedy this situation.

The reality of victims being more recognisable than perpetrators is furthered by the recent lawsuit against xAI. Deepfake images created using Grok led to a lawsuit, but SpaceX, the parent company, has now asked the claimants to be publicly identified, or has asked them to drop the lawsuit. Such a demand for the victims to identify themselves can be perceived as an intimidation tactic, which speaks to the persistence of social stigma that surrounds litigation and enforcement efforts for NCII.

The moderation of illegal content on popular communication app, Telegram, depends on individual users flagging such content. Combined with the lack of general oversight from an internal board, this framework signifies how dissemination can fly under the radar. Unless victims file complaints, or a whistleblower flags a specific channel or group within such platforms, NCII on the platform can be shared without any intervention.

Telegram states that their moderators monitor public parts of the platform and remove content breaching the NCII prohibition, but the private and anonymous nature of the application hinders identification of perpetrators - as opposed to the victims, whose identities are being misused and abused actively. The platform has previously been accused of abetting the distribution of such content due to poor handling of NCII material and has 'apologised' to national authorities.




*A 'VPN' is a virtual privacy network, which creates a private, encrypted connection between a device and the internet. It hides a user's personal IP address and scrambles browsing data, which allows people to access the internet and not be tracked easily.

StopNCII.org
Stop Non-Consensual Intimate Image Abuse

StopNCII.org indicates resources on how a victim can reach out to platforms to have their images taken off the platform.

## CHALLENGES OF CROSS-BORDER ENFORCEMENT

In general, cross-border cybercrime enforcement is hindered by conflicting national laws, competing jurisdictions, and slow evidence sharing processes. Traditional law enforcement mechanisms are ill equipped to handle the rapidly developing avenues for technology facilitated criminal activity. The borderless nature of NCII abuse means that authorities, in order to track down perpetrators, rely on cooperation amongst national authorities, domestic laws, and communication platforms like Telegram and content delivery networks like Cloudflare. Tackling sexual deepfake creation and distribution can span across multiple countries, tracking, prosecuting and extraditing them can take months or years.

## BALANCING FREE SPEECH RIGHTS

Creative applications of deepfakes can be innocuous, legitimate journalistic and satirical use-cases, which are usually protected by constitutional frameworks of free speech. However, the potential for harm caused by deepfake technologies in lieu of nudification tools is particularly grievous and tests the limits of protections granted by freedom of speech. While constitutional frameworks globally value freedom of speech and expression, such rights are not absolute. Most countries curb free speech protections on grounds of reasonable restrictions, such as sovereign interests, public order, decency and defamation, among others. Accordingly, prohibitions on NCII should be centred on the illegality of the erasure of victims' agency. This approach does not negate free speech, but balances it with the fundamental rights of the victim.

# Best Practices

## 1 REGULATION AT AN APP-STORE LEVEL:

Globally, efforts to regulate specific applications that use Gen-AI tools to create NCII are beginning to include targeted bans. For instance, EU lawmakers are now explicitly banning 'nudification' apps, requiring developers and app stores to remove them from their services by December 2026. They have also included criminal prosecution for end-users creating NCII with such tools, effectively creating a dual liability model - a structural departure from previous host platform-centric regulation. Additionally, developers building GenAI imaging tools need to audit their products to evaluate if the service can be used for NCII generation.

## 2 CONTENT TAKEDOWN AND MODERATION AT SOCIAL MEDIA PLATFORM LEVELS:

For governments, one of the most commonly adopted regulatory tools has been to order online platforms to rapidly take down such content. However, globally, the timelines provided for platforms to respond to such content differ. The UK, for example, has a 48 hour window for platforms. The US's recent Take It Down Act also requires social media platforms to remove NCII content 48 hours within a victim's request. In contrast, India's recently introduced rules, require intermediaries to take down NCII content within 2 hours, a notable reduction from the previous **24-hour timeline.**

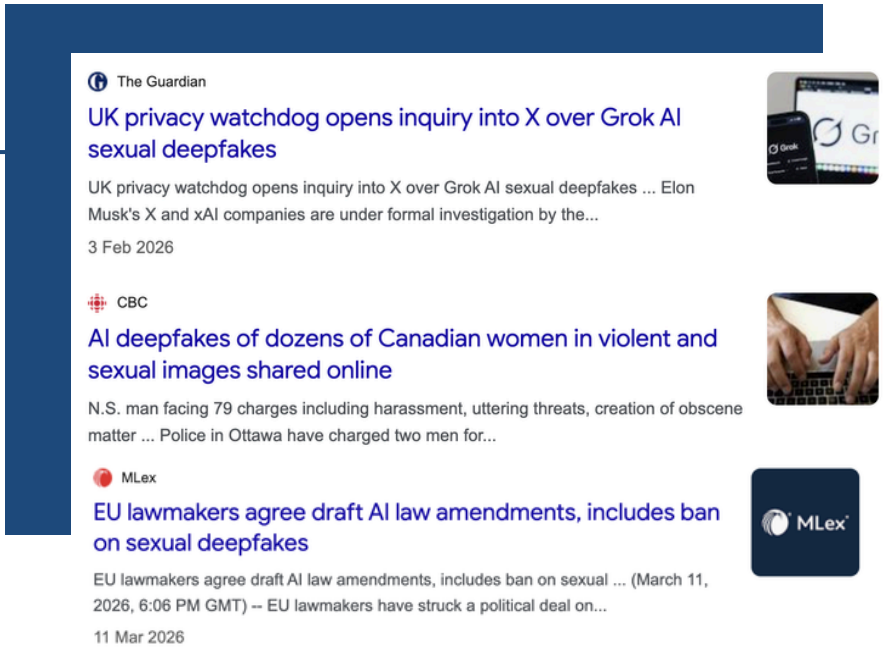
The efficacy of extremely short timelines for content takedown must be questioned. While a shorter response time makes sense from a protection perspective once a victim has reported a post, this could lead to increased reliance on AI detection tools. These tools have limited accuracy and could misidentify legitimate speech as sexualised deepfakes.

Technical tools like hash-matching are also being used to ensure NCII content is automatically taken down if it is re-uploaded. StopNCII.org, for example, generates a digital fingerprint or a **'hash' of such content,** enabling participating companies to locate and identify similar content that must be removed if it matches the hash.

> Ofcom, the British regulator, have indicated hashmatching as a key in limiting circulation and speedier takedown of NCII.

## 3 CRIMINALISING BOTH CREATION AND PRODUCTION OF DEEPFAKES:

Criminalising the creation of NCII is just as important as its distribution, since the wide ambit of harm begins with the creation of such content itself. Whether or not this content is actually distributed further, for victims, the threat of distribution looms from the moment of creation, and can become a tool for perpetrators to further coerce or control the victims' behaviour. Countries across the world have begun adopting laws addressing the various levels of sexual deepfakes. South Korea criminalises both creation and distribution levels, while Australia's 2024 amendment bans sharing sexual deepfakes and explicit material. The legal evolution is perhaps most explicitly visible in the UK, where the earlier offence of distribution and dissemination of NCII has now been expanded to include the creation of such content as well, following critique from scholars and legislators.



# Future Approaches to Adressing Sexual Deepfakes

1

## CROSS BORDER ENFORCEMENT CHALLENGES:

Evolving best practice should focus on expanding the extraterritorial application of laws in the absence of globally recognised regulations, like international treaties, agreements or UN guidelines. International efforts must include streamlining of coordination mechanisms between police, regulators, applications, app stores, courts, both nationally and internationally to target the distribution of NCII.

2

## USING CONSENT AS A BENCHMARK FOR REGULATION:

Regulators must take stock of the fact that the fundamental crime underlying the creation, dissemination and distribution of sexualised deepfakes is that the victim did not consent to their likeness being used in this manner, or being objectified online. Which means, it is important for laws to explicitly criminalise the non-consensual nature of such actions.

In the absence of the same, in countries like India, regulators have often used archaic provisions on obscenity and decency to criminalise not just non-consensual intimate imagery, but also any content that the government finds 'immoral'. This is problematic because historically, the online sphere and technological tools have been used by marginalised creators — including queer and transfolks — to diversify the meaning and depiction of acceptable sexualities and bodies, often through expressions of art, imagery and videos, all of which is consensual and legitimate expression of the fundamental right to speech. Without the crime of sexualised deepfakes being explicitly framed as being non-consensual, such content can also be potentially criminalised by more repressive and authoritarian regimes, thereby diminishing the democratic diversity of the digital public sphere.

*This explainer has been authored by the team at Centre for Communication Governance at National Law University Delhi (CCG-NLUD).*

*Written by Rishiti Choudaha and Torsha Sarkar. Design by Shivani Mago. Reviewed by Jhalak Kakkar and Shashank Mohan.*

*With support from* Ford Foundation FRIEDRICH NAUMANN FOUNDATION For Freedom.


**Centre for Communication Governance**
National Law University Delhi
Sector 14, Dwarka
New Delhi – 110078
011- 28031265

ccgdelhi.org | privacylibrary.ccgnlud.org |
ccg@nludelhi.ac.in | x.com/CCGNLUD