This submission is respectfully presented to the United Nations Office on Drugs and Crime (UNODC) in the framework of its mandate to support Member States in preventing and combating cybercrime, including online child sexual exploitation and abuse and emerging forms of technology-facilitated violence against women and children.

The undersigned delegation considers that the incident involving the Grok artificial-intelligence system on the X platform raises serious concerns that warrant urgent attention and concerted international action.

Title and formal details

Title

Submission to the United Nations Office on Drugs and Crime **Concerning AI-Enabled Image-Based Abuse and the Grok Incident on the X Platform**

Submitted by: **Laura Laffitte Salis-Gabbiani, Delegate to UNODC Civil Society**

Date: **Melbourne Australia, 18th January 2026**

Executive summary

The deployment of the Grok artificial-intelligence system on the X platform has given rise to serious and credible allegations that its image-editing functionality has been used to generate and disseminate sexualised and semi-nude images of women and girls, including those who appear to be minors, without their consent. This functionality, reportedly capable of "undressing" persons in photographs or rendering clothing transparent, has profound implications for the protection of children, the prevention of violence against women and girls, and the integrity of global efforts to combat cybercrime.

The present submission situates the Grok incident within the broader phenomenon of AI-enabled image-based abuse, including deepfakes, "nudification" tools and synthetic child sexual abuse material, and examines the extent to which existing international standards and national legal frameworks are equipped to respond to these developments. It is submitted that current safeguards and regulatory measures are insufficient and that urgent, coordinated action is required at national, regional and international levels, with UNODC playing a central supporting role.

In particular, the submission:

- Describes the factual background relating to Grok's image-editing functionality and its alleged misuse to produce sexualised depictions of women and apparent minors.
- Analyses the international legal and policy framework applicable to online child sexual exploitation and abuse, technology-facilitated gender-based violence and cybercrime more generally.
- Highlights the specific challenges posed by deepfakes, "nudification" technologies and synthetic child sexual abuse material for law-enforcement, prosecutorial and victim-support systems.
- Proposes a set of recommendations addressed to UNODC, Member States and private-sector actors aimed at strengthening prevention, protection, accountability and international cooperation.

UNODC is invited to consider developing specialised guidance on AI-enabled online sexual exploitation and abuse; supporting Member States in updating their legal frameworks and investigative capacities; and engaging with technology companies to promote robust safety-by-design standards and effective safeguards against image-based abuse.

---

1. Introduction and background

The present submission is respectfully presented by the undersigned delegation to inform UNODC's ongoing work on cybercrime, online child sexual exploitation and abuse, and the broader governance of emerging digital technologies. UNODC's existing tools and activities on cybercrime and child protection provide a strong foundation from which to address new and complex forms of AI-enabled image-based sexual abuse.

Recent years have witnessed a rapid expansion in the use of generative artificial intelligence systems capable of producing, altering and manipulating images, audio and video content in highly realistic ways. While such technologies offer important opportunities for innovation, they also create powerful new tools for offenders to generate and disseminate sexualised imagery of adults and children without consent, often across borders and at unprecedented scale. The convergence of social-media platforms, AI-driven image editing and weak or fragmented regulatory environments has contributed to a marked increase in online child sexual exploitation and abuse, as well as technology-facilitated gender-based violence.

Within this broader context, open-source reporting in early 2026 indicated that Grok, a generative AI system integrated into the X platform, enabled users to upload photographs and request altered versions in which clothing was removed, minimised or rendered transparent, resulting in quasi-nude or highly sexualised depictions. These capabilities appear to have been used in respect of women and girls, including in some cases individuals who appeared to be minors, raising acute concerns about the production and dissemination of sexualised images that may fall within or closely approximate the scope of child sexual abuse material.

The undersigned delegation submits that this incident is emblematic of wider structural weaknesses in platform safety, AI governance and regulatory oversight, and that failure to address such weaknesses risks undermining decades of progress in the protection of children and other vulnerable persons from cyber-enabled exploitation.

2. International legal and policy framework

2.1 Children's rights and online sexual exploitation

The Convention on the Rights of the Child (CRC) and its Optional Protocol on the sale of children, child prostitution and child pornography impose clear obligations on States to protect children from all forms of sexual exploitation and sexual abuse, including when facilitated by information and communication technologies. Although these instruments predate the advent of generative AI, their object and purpose point strongly to a technologically neutral interpretation that encompasses novel forms of online child sexual exploitation and abuse, such as synthetic or AI-generated sexualised imagery of minors.

UN guidance on online child sexual exploitation and abuse recognises that such exploitation may involve grooming, coercion and blackmail, the production of self-generated sexual content, live-streaming of abuse and the creation and sharing of sexualised images, whether captured directly or generated digitally. In this regard, it is submitted that the use of AI systems to generate sexualised depictions of children, including realistic synthetic images that resemble real children, engages the same protective concerns and should be addressed with the same level of urgency as "traditional" child sexual abuse material.

2.2 Violence against women and technology-facilitated abuse

International and regional standards relating to violence against women and girls recognise technology-facilitated abuse, including non-consensual sharing of intimate images, deepfake pornography and online harassment, as forms of gender-based violence that engage State obligations of due diligence. Studies indicate that the overwhelming majority of deepfake content online is pornographic, and that around 99 per cent of victims of deepfake pornography are women, reflecting entrenched patterns of gendered abuse.

It follows that States are required not only to criminalise such behaviours but also to exercise due diligence in regulating private actors whose products and services foreseeably facilitate technology-based violence against women and girls. The deployment of AI-driven "nudification" tools without stringent safeguards, including clear prohibitions on sexualising user images without explicit, informed consent, may therefore be inconsistent with these obligations.

2.3 Cybercrime instruments and digital evidence

UNODC's work on cybercrime underscores that online child sexual exploitation and abuse, including the production, distribution and possession of child sexual abuse material, constitute serious offences that must be addressed through comprehensive criminalisation, effective investigation and robust international cooperation. Regional instruments and soft-law guidelines similarly emphasise the need for updated legal definitions, investigative powers and evidentiary rules tailored to digital environments.

The emergence of synthetic child sexual abuse material and AI-generated sexual content raises complex questions regarding the classification of such material for criminal-law purposes, as well as challenges in attributing responsibility, authenticating evidence and preserving digital integrity. It is therefore submitted that existing cybercrime frameworks must be interpreted and, where necessary, revised to ensure that synthetic and AI-generated exploitative content does not fall into legal grey areas that undermine child protection and accountability.

2.4 Responsibilities of private-sector actors

International human rights standards increasingly recognise that private-sector entities, including technology companies and AI developers, have responsibilities to respect human rights and to avoid causing or contributing to adverse human-rights impacts through their operations, products or services. In the digital sphere, this includes adopting robust policies and technical safeguards to prevent the use of their platforms and tools for online child sexual exploitation and abuse and technology-facilitated gender-based violence.

The Grok incident indicates that platform-level safeguards were inadequate to prevent the misuse of an AI image-editing feature in ways that sexualised women and girls, including apparent minors, and enabled the rapid dissemination of such content across a global social-media network. This situation raises serious questions regarding the fulfilment by private-sector actors of their responsibility to respect human rights and avoid facilitating foreseeable harm.

---

3. The Grok incident: factual description and analysis

According to publicly available reports, X introduced an “edit image” feature powered by its Grok AI system, which allowed users to upload photographs and request modified versions. Users reported that, when prompted, Grok could remove or substantially reduce clothing in images of women and girls or render clothing effectively transparent, generating quasi-nude or overtly sexualised depictions that appeared realistic and lifelike.

It is alleged that this functionality was, in some instances, applied to photographs of individuals who appeared to be minors, thereby producing images that may meet or closely approach definitions of child sexual abuse material in some jurisdictions. The resulting content was shared across the platform, including without the knowledge or consent of the individuals depicted, who in many cases had originally posted non-sexualised images.

Media reports indicate that regulatory authorities in at least one State sought urgent clarification from X regarding the safeguards in place to prevent Grok from generating sexualised imagery of minors and the remedial measures taken once the issue became public. Subsequent statements suggested that X implemented restrictions preventing Grok from generating nudity only in jurisdictions where such content is clearly unlawful, indicating a fragmented and inconsistent global approach to safety protections.

On the basis of the foregoing, it is respectfully submitted that the misuse of Grok's image-editing functionalities was reasonably foreseeable, given the wider pattern of abuse of deepfake and "nudification" tools for exploitative purposes. In circumstances where a platform is aware of systemic risks of image-based sexual abuse and possesses the technical capacity to mitigate those risks, the deployment of features that facilitate the sexualisation of women and children without consent raises serious concerns regarding the standard of due diligence exercised by the provider.

The cross-border nature of X's operations further complicates regulatory oversight, as content created in one jurisdiction can be disseminated instantly worldwide, including to States with stricter child-protection and obscenity laws. This underscores the need for harmonised international standards and cooperative mechanisms to address AI-enabled image-based abuse effectively.

---

4. Deepfakes, "nudification" and AI-enabled cybercrime

Deepfake technologies and "nudification" tools represent a new frontier in cyber-enabled sexual exploitation, enabling the creation of highly realistic synthetic images that can be indistinguishable from genuine photographs. Research indicates that the vast majority of deepfake content currently circulating online is pornographic, with women and girls disproportionately targeted, often without their knowledge or consent.

These technologies pose multiple challenges:

- Scalability and accessibility: Offenders no longer require physical access to victims to generate sexualised imagery; instead, they can use publicly available photographs and easily accessible tools to create synthetic, exploitative content at scale.
- Plausibility and persistence: Highly realistic synthetic images can be misperceived as authentic, damaging the reputation and psychological wellbeing of victims; once disseminated, such images can be extremely difficult to remove, with copies persisting across multiple platforms and networks.
- Complex evidentiary questions: Synthetic content challenges existing approaches to digital evidence, requiring new techniques to verify authenticity, attribute responsibility and demonstrate harm, particularly where offenders seek to exploit perceived ambiguity regarding the "reality" of the depicted acts.

In relation to children, the capacity of AI systems to generate sexualised images of minors, including entirely synthetic child avatars, raises particularly grave concerns. Such content can be used to groom children, normalise abusive behaviour or satisfy offender fantasies, and may in some jurisdictions fall outside narrow definitions of child sexual abuse material that presuppose the involvement of an actual child in image capture. The undersigned delegation submits that, from a child-rights perspective, these distinctions should not prevent robust criminalisation and enforcement.

The Grok case illustrates how a general-purpose AI tool, ostensibly designed for benign image editing, can be repurposed as a "nudification" device that feeds into existing ecosystems of deepfake abuse and online sexual exploitation. This demonstrates the urgent necessity of integrating safety-by-design principles into the development and deployment of AI systems, with particular attention to foreseeable misuse.

---

5. International protection of vulnerable groups online

International human rights law requires States to take all appropriate legislative, administrative, social and educational measures to protect children from all forms of sexual exploitation and sexual abuse, including in digital environments. This obligation encompasses the duty to prevent online child sexual exploitation and abuse, investigate allegations effectively, prosecute perpetrators and provide remedies and support to child victims.

The principle of the best interests of the child must be a primary consideration in all actions concerning children, including the design, deployment and governance of AI systems and social-media platforms likely to be accessed by children or used to process images of children. This implies that developers and platform operators should proactively assess and mitigate the risks that their systems may be used to generate sexualised images of minors or otherwise facilitate online child sexual exploitation and abuse.

At the same time, obligations related to the elimination of violence against women and girls require States to address technology-facilitated forms of gender-based violence, including non-consensual intimate-image abuse, deepfake pornography and online harassment. Women and girls are disproportionately affected by deepfake and "nudification" abuse, and the psychological, social and economic consequences can be severe and long-lasting. Accordingly, measures to regulate AI-enabled image editing should explicitly incorporate a gender-sensitive lens.

The undersigned delegation submits that protecting vulnerable persons online requires a holistic approach that combines:

- Robust criminal-law frameworks covering synthetic as well as "real" exploitative content involving children.
- Civil-law and regulatory remedies for non-consensual sexualised imagery of adults, including deepfake and "nudified" content.
- Effective age-appropriate platform design, safety features and content-moderation practices.

6. Role of UNODC and recommendations

Given its mandates in the areas of crime prevention, criminal justice and cybercrime, UNODC is well placed to support Member States in addressing AI-enabled online sexual exploitation and abuse. Building on its existing tools and programmes, UNODC could play a central role in clarifying normative standards, strengthening capacities and fostering cooperation between States, international organisations and the private sector.

The undersigned delegation respectfully submits the following recommendations:

6.1 Normative and policy development

- UNODC is invited to consider developing dedicated guidance on AI-enabled online sexual exploitation and abuse, including deepfakes, “nudification” tools and synthetic child sexual abuse material, in cooperation with other UN entities and relevant regional organisations.
- Member States are encouraged to review and, where appropriate, update their legal frameworks to ensure that:
    - The non-consensual creation and dissemination of sexualised synthetic imagery of identifiable individuals, including deepfake and “nudified” content, is clearly criminalised.
    - The production, possession and distribution of synthetic child sexual abuse material and sexualised depictions of minors, regardless of whether a real child was directly involved, are subject to effective criminal sanctions.

6.2 Capacity-building and technical assistance

- UNODC is encouraged to integrate modules on AI-enabled image-based abuse into its cybercrime and child-protection capacity-building activities, including training for law-enforcement officials, prosecutors and judiciary.
- Technical assistance programmes should support the development of specialised investigative units and digital-forensics capabilities equipped to identify, analyse and respond to deepfake and “nudification” offences, while respecting human rights and due-process guarantees.

6.3 Engagement with the private sector

- UNODC could facilitate structured dialogue between Member States and technology companies, including AI developers and social-media platforms, to promote safety-by-design approaches and shared standards for preventing AI-enabled image-based abuse.

- Private-sector actors should be encouraged to:
    - Implement robust technical safeguards that prevent or severely restrict the generation of sexualised imagery of minors and non-consensual sexualised depictions of adults.
    - Adopt and enforce clear terms of service prohibiting the use of their tools for image-based sexual exploitation, supported by effective monitoring and enforcement mechanisms.

6.4 Research, data collection and evaluation

- UNODC, in partnership with academic and civil-society organisations, could support research on the prevalence, patterns and impacts of AI-enabled image-based sexual abuse, including deepfake and "nudification" offences, with disaggregated data by age and gender.
- Evidence generated through such research should inform the design and evaluation of laws, policies and programmes aimed at protecting children and other vulnerable groups from cyber-enabled sexual exploitation.

Annex I – Illustrative chronology and features of the Grok incident

Open-source information suggests the following chronology and key features:

- In late 2025, X deployed an "edit image" function powered by the Grok AI system, enabling users to upload images and request various alterations.
- Users soon reported that Grok could be prompted to remove clothing from images of women and girls or render clothing transparent, creating highly sexualised or semi-nude depictions without the consent of the individuals portrayed.
- Some images reportedly involved individuals who appeared to be minors, giving rise to concerns that the generated content might fall within definitions of child sexual abuse material in certain jurisdictions.
- National authorities requested information from X regarding the safeguards in place to prevent such misuse and the remedial steps taken in response to the incident.
- X subsequently indicated that it had restricted Grok's nudity-generation capabilities in jurisdictions where such content is clearly illegal, but not necessarily on a global basis, thereby leaving protections uneven across different legal environments.

Annex II – Emerging evidence on AI-generated image-based sexual abuse

Recent studies and expert briefings highlight several trends:

- Deepfake content online is predominantly pornographic in nature, with women and girls overwhelmingly represented among victims.
- "Nudification" applications, first popularised through tools such as DeepNude and its successors, allow users to digitally remove clothing from photographs, producing synthetic but realistic images that can be used for harassment, blackmail or public shaming.
- Offenders share techniques and tools through social-media platforms, encrypted messaging services and dark-web forums, complicating detection and takedown efforts.
- AI can generate entirely synthetic child avatars and sexualised depictions of minors, which can be used for grooming or gratification and may evade existing detection systems or fall outside narrow statutory definitions of child pornography or child sexual abuse material.
- Victims of deepfake and "nudification" abuse report severe psychological distress, damage to personal and professional relationships, and significant barriers to obtaining effective removal of content and legal redress.

Annex III – Indicative elements for State practice and private-sector standards

To strengthen protection against AI-enabled image-based abuse, the following elements may be considered:

- Legislative elements for States:
    - Express criminalisation of non-consensual creation and distribution of sexualised synthetic imagery of identifiable persons, including deepfake and "nudified" content.
    - Clear provisions addressing synthetic child sexual abuse material and sexualised depictions of minors, irrespective of whether a real child was directly involved in image capture.
    - Effective investigative powers and international-cooperation mechanisms tailored to the digital nature of these offences, with appropriate safeguards for human rights and due process.

- Regulatory and policy elements:
    - Requirements for platforms and AI providers to conduct child-rights and human-rights impact assessments prior to deploying powerful image-editing or generative tools, with a focus on risks of sexual exploitation and abuse.
    - Obligations to provide rapid, accessible mechanisms for victims to report non-consensual sexualised imagery, combined with expedited takedown and evidence-preservation procedures.
- Private-sector standards:
    - Adoption of safety-by-design principles, including default technical blocks on generating sexualised imagery of minors and strong safeguards against non-consensual sexualised manipulation of adult images.
    - Transparent reporting on content-moderation practices, abuse patterns and responses to AI-enabled image-based sexual exploitation, in line with data-protection and privacy obligations.

The undersigned delegation respectfully invites UNODC and Member States to take due note of the concerns and recommendations set out in this submission and to consider them in the context of ongoing and future work on cybercrime, online child sexual exploitation and abuse, and the governance of emerging digital technologies.