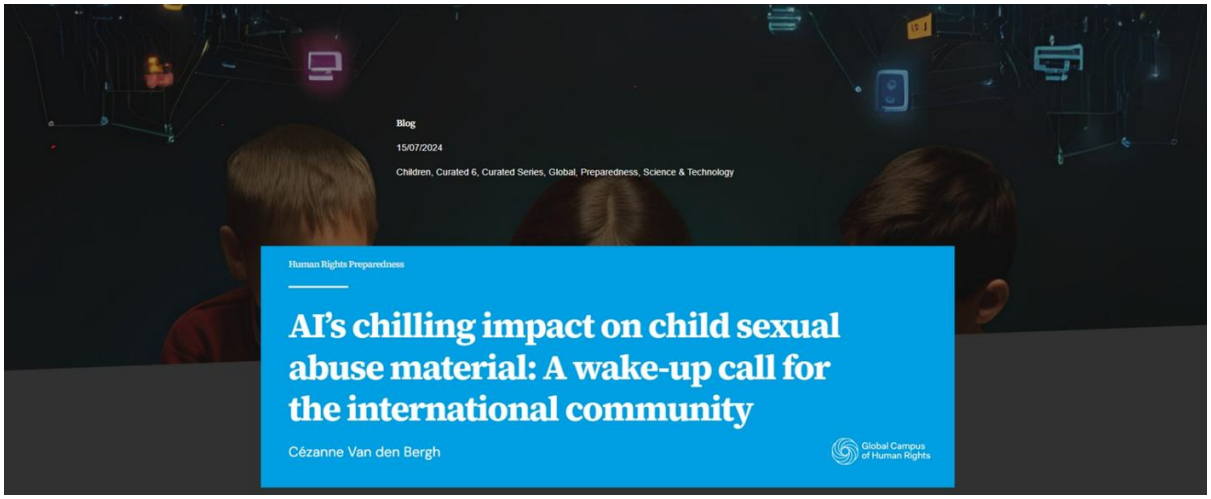

AI's chilling impact on child sexual abuse material: A wake-up call for the international community

Cézanne Van den Bergh[*]

Abstract: Child Sexual Abuse Material (CSAM) remains a pressing concern, and accessible AI tools generating hyper-realistic content exacerbate this digital crisis, endangering children, obstructing investigations, and undermining legal certainty and effectiveness of laws. It is time for global awareness of AI CSAM's alarming implications.

> By far, the greatest danger of Artificial Intelligence is that people conclude too early that they understand it.
> Eliezer Yudkowsky (co-founder and research fellow at the Machine Intelligence Research Institute)

Amid a series of health and security crises in recent years, the international community is currently confronted with a digital public emergency: the rise of AI-driven illegal content, including child sexual abuse material (AI CSAM). Addressing this alarming issue requires a child-centric and rights-based-approach, considering the significant threats to child safety, societal moral fabric, and the rule of law. This approach is aligned with and further explores insights from the 2024 Fundamental Rights Forum's panel on 'digitalisation and security', organised by the

[*] Cézanne Van den Bergh holds a Bachelor's and Master's degree in Law from KU Leuven, including an exchange at Stellenbosch University. Currently pursuing the European Master's in Human Rights and Democratisation (EMA) at the Global Campus of Human Rights and Lund University, she explores the intersection between new technologies and children's rights.

European Union Agency for Fundamental Rights (FRA). It highlights deficiencies in ensuring rights-compliant-digitalisation, particularly those affecting children's well-being. The rapid advancement of artificial intelligence (AI) and machine learning has undeniably brought numerous benefits, such as AI chatbots and augmented medical assistance, yet it has not come without significant challenges. At the FRA Forum, discussions underscored AI's societal advantages but also raised concerns about potential abuses due to inadequate EU legislation. A notable example is the upswing of AI CSAM, which unfortunately went unaddressed at the forum's panel on 'digitalisation and security'. This oversight may stem from the novelty of the employed AI tools and a lack of awareness regarding its severity. The panel only debated the specific impact of the online dissemination of real sexual abuse material on children's rights. As a result, this blog post draws attention to this digital crisis, severely exacerbated by the proliferation of user-friendly AI tools, by exploring its technological underpinnings, various manifestations, and impacts on rights-compliant digitalisation.

Technological intricacies of AI CSAM

As emphasised during the FRA Forum's panel on digitalisation and security, there exists a widespread lack of knowledge among legal scholars regarding AI and technology, such as deep fakes. Bridging this gap requires increased dialogue between lawyers and technologists. Therefore, we will first elaborate on the technological foundations of AI CSAM. This type of content generally falls into two categories: fully AI-generated material and AI-modified content, such as 'deepnudes'.For the first category, the advent of freely accessible text-to-image generators in 2023 made it easy for users worldwide to create (il)legal artificial content swiftly. As a result, these user-friendly machine learning models, the most well-known of which is Stable Diffusion, enable the creation of hyper-realistic CSAM unprecedently accessible to Internet users. Using a two-step mechanism known as 'forward' and 'reverse diffusion', this type of generative AI (GenAI) model first adds noise layers to training images, gradually converting them into Gaussian noise. Following training, the model can reverse these steps, regenerating visual images. By applying this reverse process to new random data and adding specified image data, users can create fresh outputs with desired features, such as children and nudity. The second category involves AI-manipulated abuse material derived from actual footage, altered using deepfake technology and morphing techniques. These may include age manipulation, face-swapping, the use of 'Nudify' Apps, and 'inpainting' of sexual facial expressions onto children's faces. Consequently, these manipulations could alter adult images to resemble children, transform real images of child sexual abuse into fictional or cartoon-like representations, modify innocent images of children to depict sexual activity, or any other form of alteration resulting in fictious abuse imagery.

Multifaceted impacts on FRA Forum's rights-compliant digitalisation

The rise of AI CSAM carries multifaceted implications, particularly for children's rights and safety both online and offline. A child rights-based approach to addressing AI CSAM is critical for several reasons. First, its dissemination exacerbates the prevalence of real child abuse. As AI CSAM contributes to the broader market of CSAM, it perpetuates and tolerates the sexualisation of children, which in turn decreases the barriers to real-life abusive fantasies. While some argue that it might offer an 'alternative' to real-world abuse, there is currently insufficient evidence to support this stance. More so, research actually indicates that it yields counterproductive outcomes, contributing to a culture of normalising child abuse.Secondly, GenAI models are trained on datasets that frequently include real CSAM, as a recent report revealed. The AI tool's model weights, which act as the basic components for image generation, are inferred from patterns observed in machine learning models trained on input datasets. Consequently, these parameters incorporate specific features and characteristics of real CSAM data and use them to define the concrete output image. This indefinite perpetuation of real CSAM leads to the re-victimisation of abused children, which not only infringes upon their right to dignity, integrity, and privacy, but also profoundly impacts their long-term mental health and well-being. This result starkly contrasts with the main principles discussed at the FRA Forum, which underscored the critical importance of clearly translating existing human rights standards into national laws. It emphasised the obligation of states to uphold human rights in the regulation of AI and the digitalisation development and deployment, which includes children's privacy and well-being.Thirdly, there is a real risk that such simulated imagery is being used for sexual exploitation, commonly referred to as grooming, and financial sextortion. Perpetrators are using AI-altered explicit content derived from innocent images of a minor to coerce them through social media platforms, such as Instagram or Snapchat, compelling them to provide real sexual footage or money. The financial exploitation of minors through the use of real nude images obtained by hiding behind fake flirtatious profiles has already been described as a 'digital pandemic', as evidenced by the impact of perpetrators such as the Yahoo Boys, particularly on teenage boys. Now the gravity of the situation is even greater, as it is no longer necessary to use actual nude pictures to blackmail victims, given the free availability of hyper-realistic AI-driven text-to-image generators. This alarming development clearly affects the FRA Forum's main premise of safeguarding human rights standards in the development of AI, as the misuse of these tools for nefarious purposes poses serious threats to the well-being of minors. In addition, the escalating prevalence of AI CSAM complicates efforts by online platforms filtering harmful content and law enforcement investigations. The intentional inundation of investigation authorities with AI CSAM aims to impede the effective prosecution of actual CSAM possession amid a landscape where real and fictitious material becomes indistinguishable. This growing challenge in discerning between real and virtual

CSAM exacerbates the struggle to identify and assist real-life victims, thereby amplifying the detrimental impact on abused children. Another crucial aspect is that regulating AI CSAM in the European Union raises numerous legal challenges. While recognising the significance of the EU AI Act and its regulatory obligations for AI system providers, significant legislative shortcomings remain. Chief among these is the lack of any criminal accountability for the GenAI model or its 'humans-behind-the-machines', potentially limiting the effectiveness of laws. Additionally, the unpredictability of GenAI's technical evolution, methods to bypass technical safeguards, and corresponding legal responses contribute to legal uncertainty. The key takeaway from the forum's human rights table on 'code to conscience' underscored the urgent need for clarity of laws and standards governing AI and human rights, along with clear state obligations. Moreover, the panel emphasised the importance of accountability in AI development and deployment through a collaborative approach defining clear responsibilities of various stakeholders, including those in the industry. The panel stressed the EU's role in advocating transparency, accountability, and human rights impact assessments, which also applies for AI CSAM, to ensure child rights-compliant digitalisation. This necessitates that existing regulations provide greater clarity and accountability for AI CSAM through a child-centred approach. Finally, traditional criminal law faces challenges as it typically requires a tangible victim, 'real' harm, and a causal link between the perpetrator and the victim's harm. However, the forum's cited human rights table emphasised the imperative of recognising the harms inflicted by technology, advocating for enhanced evidential frameworks supported by reports from Chief Security Officers. It is essential to construct a specific narrative around the harms currently observed, a task equally applicable to addressing AI CSAM.

Urgent call to collective action

As the line between real and simulated imagery blurs rapidly, we risk succumbing to an 'AI illusion', making it increasingly difficult, if not impossible, to discern fictional images from real content. This phenomenon of hyper-realistic CSAM poses severe threats to children's rights, including their rights to dignity, integrity, privacy and overall protection, in various ways.

First, it fosters a culture of normalising child sexual abuse material, and reducing the barriers to real-life abuse. Second, it causes the re-victimisation of previously abused children whose images are present in the input data and model weights of AI tools, influencing the final output. Thirdly, it severely facilitates the sexual and financial exploitation of children through social media, which has become the fastest growing crime targeting youth worldwide. Additionally, media platforms struggle to effectively moderate illegal content, and law enforcement faces challenges in conducting thorough investigations due to the influx of AI CSAM. Finally, AI CSAM poses legislative constraints leading to legal uncertainty and a lack of child rights-compliant digitalisation, undermining efforts highlighted during

the forum's panel on 'digitalisation and security'. The imperative for clarity of laws, accountability through a multistakeholder approach, and recognition of technology-related harm stands out as critical themes.To safeguard children in the AI-dominated realm, urgent awareness and collective action is essential. This necessitates the involvement of legislators, AI system providers, law enforcement authorities, online service providers, IT experts and other stakeholders. Specifically, I encourage FRA to closely monitor AI CSAM and scrutinise its profound impacts on child safety in both online and offline environments.

As highlighted throughout the forum, we must consider the impacts of the digitalised world on our future citizens, vulnerable youth and children. The digital footprint we create today will shape their tomorrow, and human rights must embrace a pivotal role in the world we want to build for and with them. The urgency for action in this digital crisis is clear—we cannot afford to delay.