

AI and CSAM: Current Standing and a Future Perspective

summary

Analiv Hasert, Andreia Costa Marchante, Terezia Kucarova
Erasmus University Rotterdam

Introduction

Child sexual abuse material (CSAM) constitutes a fast-evolving and dynamic threat, representing one of the most harmful forms of online exploitation targeting children. Recent advances in artificial intelligence (AI) have fundamentally changed this landscape: generative AI technology makes it possible for offenders to create hyper-realistic and easily scalable CSAM, without requiring direct access to a child. Both the quantity and the severity of AI CSAM have been growing rapidly; the convergence of increasingly accessible tools significantly lowers the technical barriers for the creation and distribution of such content. With legal frameworks and detection methods struggling to keep up with the pace of technological development, there is an urgent need for evidence-based knowledge on which to formulate effective preventive and responsive measures.

AI Creation Tools

The primary technology powering AI CSAM is the diffusion model, a type of generative AI that produces images based on textual input. Two broad categories of how AI CSAM is created encompass: AI-generated imagery and AI-manipulated content. AI-generated CSAM involves fictional imagery of minors, created largely via open-source diffusion models, which offer offenders fewer external safeguards than the closed-source alternatives. The emergence of fine-tuning techniques within this category allow models to be further specialized to produce targeted outputs.

AI-manipulated CSAM involves the use of an actual image of a person, further altering it through AI, directly victimizing the individual. Newly emerging techniques are rapidly expanding the threat, with synthetic motion video and audio deepfakes enabling the production of multimodal abuse material.

Motivation: Creators

Research shows that AI CSAM is created for a range of motivations, and that these are broader than simple sexual gratification. Some creators are driven by financial gain, using platforms and communities that support custom orders or model-sharing ecosystems. Others seek status, technical prestige, or belonging within creator communities. AI also enables coercive uses such as sextortion. A feature is that AI lowers the threshold for offending. Creation no longer requires direct access to a child, and even inexperienced users can produce highly realistic material with tools. This makes the pool of potential creators larger and shifts AI CSAM toward a low-cost, high-volume, and highly adaptable form of abuse. Younger users also appear in the picture, with evidence suggesting that some adolescents create this material out of curiosity, peer pressure, revenge, or experimentation, often without knowing of the long-term harm involved.

Motivation: Consumers

According to research, different motivators for CSAM consumers can comprise sexual gratification, addiction, paraphilic sexual interests, particularly pedophilia, and normalisation. Studies have demonstrated that people can get addicted to pornography, exhibiting similar basic mechanisms and underlying neural processes to substance addiction, and similarly, repeated CSAM users can exhibit compulsive behaviour and report their consumption as highly rewarding. Additionally, research also shows that pedophilia is a strong predictor of CSAM consumption, especially when hypersexuality is present.

Furthermore, normalisation may contribute to all these motivators as AI CSAM consumers may foster impunity and a lack of empathy for child victims through their engagement with AI CSAM, due to a sense of diminished or no real harm. However, research has also shown that viewing, possessing, or making CSAM strengthens deviant thoughts, fantasies and urges, and can increase the risk of escalation.

Psychological Impact and Harm: Victims

AI-driven sexual exploitation of children subjects its victims to severe psychological harm. Victims have described experiencing fear, shame, humiliation, and self-blame, alongside feelings of helplessness and loss of control over one's identity, indicating symptoms consistent with anxiety, depression, and PTSD. The nature of AI CSAM creates a unique challenge: removal efforts are complicated by infinite replication and rapid spread of the content online, enabling both perpetuating trauma and continued violation of the victim's dignity. Victims find it increasingly difficult to seek justice, as online platforms often delay removal requests and fragmented legal frameworks complicate access to justice. Research focused on victims of AI CSAM is limited, and existing interventions fail to adequately cater to these complexities.

Mitigation and prevention efforts

Multiple legislative and technological frameworks directed at mitigating CSAM have been created and/ or are being discussed across the world. In the European Union, the efforts are now centered on the Directive 2011/93/EU and its pending 2024 recast directive. This is complemented by the EU AI Act. In the United Kingdom legislation such as the Protection of Children Act, the Coroners and Justice Act, and the Crime and Policing Bill also address the rise of AI CSAM. In the United States, the PROTECT Act, and more recently the STOP CSAM Act and the ENFORCE Act also address the creation and possession of AI CSAM.

Moreover, even though these are secondary to legal enforcement, there are also technological and clinical interventions worth mentioning as some prevention strategies include industry-led digital watermarking, the curation of safer training datasets (e.g., LAION-5B), and clinical treatments for sexually deviant behaviours that offer some preventative potential.

Legal gaps and detection challenges

Existing child sexual abuse material laws have not fully caught up with AI-generated content. A central difficulty is that AI CSAM may not involve a real child in the image itself, complicating legal classification and culpability. Across jurisdictions, there is uncertainty around who is responsible, being the end-user, the model trainer, the platform, or the provider of the system that made the content possible. In the EU, the current legal framework was written before modern generative AI and therefore does not always speak clearly about synthetic material. This gap is especially visible in this period between older legislation and the pending rules. Detection is also becoming harder. Traditional hash-based tools are effective for known material but are less useful when every AI-generated image is unique. Watermarking and AI-detection systems help, but offenders can often remove, alter, or bypass them. At the same time, AI CSAM is becoming more photorealistic and more abundant.

A Future Perspective: Research Recommendations

After accumulating information in an evidence-based report on the current knowledge and legislation on the topic of AI-manipulated and -generated CSAM, we conclude by outlining research priorities to guide future policy and practice in this fast-evolving field. Research priorities include further investigation into systematic misuse of open-source AI models, the integrity of training datasets, and the implications of emerging technologies such as tool convergence or agentic AI.

Continuing to study the cognitive and moral frameworks of creators and consumers to map target groups for intervention is crucial, further research is needed on the social and developmental drivers of creation, including online communities and youth involvement, as well as on identifying consumer profiles and clarifying the link between (AI-) CSAM consumption and addiction, which remains underexplored beyond adult pornography. Additionally, research into the unique psychological impacts of AI CSAM victimization remains scarce, particularly regarding the gendered dimension of harm, secondary victimization, and the adequacy of existing victim support frameworks.