IWF
Internet
Watch
Foundation

Working together
to stop child sexual
abuse online

AI CSAM Report 2026

Harm without limits: AI child sexual abuse material through the eyes of our Analysts

Disclaimer/content note

Throughout this report, child sexual abuse material (CSAM) generated through artificial intelligence (AI) is referred to as AI CSAM. This report contains no AI CSAM or visual depictions of child sexual abuse.

However, this report does contain descriptions of the methods used to generate AI CSAM, alongside verbatim comments from perpetrators, redacted comments from public reports, and insights from the Internet Watch Foundation's (IWF) Hotline. Some sections may therefore be distressing for readers.

While this report occasionally references external research, it is primarily based on aggregated findings from our internal expertise.

Reader discretion is advised.

Table of contents

Foreword 5

Executive summary 6

Introduction 7

01 AI CSAM: How it is generated 8

- AI models and the generation of AI CSAM 9
- Fine tuning and low-rank adaptation models (LoRAs) 10
- Legislation on CSA image-generators 11
- Open and closed-source models 12
- Where AI CSAM appears 12

02 AI CSAM Harms 14

- Revictimisation 16
- Gendered Harms 17
- Normalisation of sexual violence against children 18
- Operational harm 19
- Legislation on the creation, possession and distribution of AI CSAM 21

03 Insights from the Hotline: Developments in AI CSAM 22

- AI CSAM videos 23
- AI CSAM with an audio component 24
- Realism 25
- Extremity 26
- Accessibility 27
- Disguise tools 28
- Commercialisation 28
- Grooming and extortion 30
- Legislation on advice and guidance for creating AI CSAM 32

04 Nudify apps 33

- Legislation on nudify apps 36

05 Chatbots 37

- What are chatbots? 38
- Technical help 38
- Sexual communication and grooming 39
- Chatbot-hosted CSAM 40
- Legislation covering the regulation of AI chatbots 41

Conclusion 42

IWF

Foreword

Since joining the Internet Watch Foundation (IWF) in August 2025, I have been struck by the scale of harm to children that we confront daily. Nowhere is this clearer than in the rapid rise and development of AI-generated child sexual abuse material (AI CSAM).

The numbers in this report are not abstract, nor is this a theoretical issue. Each and every item of AI CSAM is a violation of the most basic rights to safety, dignity, and childhood itself.

While the proportion of AI-generated material remains comparatively small among the huge amount of CSAM our Analysts action every year, the volume and severity of AI imagery has increased exponentially since our first AI report in October 2023, due to the availability and ease of tools. We now face a technological landscape that can generate infinite violations with unprecedented ease. Unless we address this problem head on, it will only continue to grow.

The AI-generated material we encounter is highly photorealistic and often indistinguishable from photographic abuse imagery. Our experts can often only recognise some AI imagery as synthetic because they recognise the victims. They know these children and they know when an image depicts something that never happened.

AI CSAM does not mean crude sketches kept hidden away, nor is it a diversion from other forms of abuse imagery or contact abuse. We are talking about sophisticated fabrications that obliterate the line between synthetic and photographic. Offenders themselves acknowledge this and are consistently impressed by the ever-more sophisticated and extreme material they can create.

To remedy the proliferation of this material, the IWF works directly with AI model developers through our Membership, and with legislators, regulators and law enforcement. First and foremost, we urge services to take greater accountability in ensuring their products are safe. And yet, we have seen some services launch generative AI tools without adequate safeguards in place. When technology companies prioritise capturing user attention and market share over safety-by-design principles, they create vectors for abuse. Child safety must be embedded from the outset, not treated as an afterthought.

We welcome the UK Government's action to introduce the ability for designated authorities like the IWF to test AI models. This represents an important step to promote effective safety-by-design prior to model release and on an ongoing basis. As a global authority on combating child sexual abuse imagery, we stand ready to partner with academics, institutions and other organisations as well as industry to deliver a robust and independent auditing function at every stage of development pre and post launch.

We cannot tackle this risk alone. We need safety-by-design embedded as a non-negotiable standard in AI development with clear standards set by governments. This technology has extraordinary potential to change our lives for the better, but only if we mitigate the harms first. Injustice is not inevitable.

We have an opportunity now. We must not waste it.

Kerry Smith
Chief Executive Officer
Internet Watch Foundation

In 2025, IWF assessed

8,029

AI-generated images and videos as showing realistic child sexual abuse

IWF identified

3,443

AI-generated child sexual abuse videos

26,385%

increase in AI-generated child sexual abuse videos

65%

Category A

Executive summary

The use of AI to generate child sexual abuse material is continuing to increase in scale, sophistication and accessibility.

The key insights of this report are as follows:

1. AI CSAM is widespread and growing

In 2025, we assessed 8,029 AI-generated images and videos as showing realistic child sexual abuse. While AI CSAM still represents a relatively small proportion of overall reports, volumes continue to rise steadily. This imagery appears across both dark web and mainstream commercial platforms on the clear web.

2. AI CSAM is increasingly extreme and sophisticated

Realistic full-motion AI video content is now commonplace. In 2025, the IWF identified 3,443 AI-generated child sexual abuse videos, an increase of 26,385% compared with the previous year, when just 13 such videos were recorded. Of these, 65% (2,233 videos) were classified as Category A, the most extreme classification. By comparison, for the full year (2025), 43% of non-AI criminal videos seen by the IWF were Category A.

3. AI CSAM causes real harm and is highly gendered

Generative models can be trained and fine-tuned using photographic abuse imagery, directly re-victimising survivors. AI CSAM fuels sexual interest in children, normalises extreme violence, and increases the risk of contact offending. Analysis also shows that girls remain overwhelmingly targeted, comprising 97% of illegal AI-generated images assessed by the IWF in 2025.

4. AI child sexual abuse chatbots are accessible on the clear web

IWF Analysts have identified AI-generated child sexual abuse images shared on AI chatbot services. These chatbots encourage users to act out simulated child sexual abuse scenarios, including “child prostitute in a hotel”.[1]

5. AI CSAM tools are converging

Advances in AI have driven the convergence of tools that previously required separate capabilities. Single applications can now generate abusive imagery with minimal effort, removing the need for technical expertise and significantly lowering barriers to entry. These capabilities are increasingly combined with AI audio generation tools, further amplifying the realism and potential harm of such content.

[1] Internet Watch Foundation (2025). *‘Disturbing” child sexual abuse images found on AI chatbot website.* Available online at: https://www.iwf.org.uk/news-media/news/disturbing-ai-generated-child-sexual-abuse-images-found-on-hidden-chatbot-website-that-simulates-indecent-fantasies/

Introduction

The IWF was one of the first organisations to sound the alarm on AI CSAM. In 2023, we published our first AI report, setting out how AI CSAM was being generated and the emerging threat.[2] In July 2024, we released a second report, updating our findings on the evolving technology and warning that advances in AI would drive a rapid escalation in the scale, extremity and sophistication of child sexual abuse imagery.[3]

Our previous reports documented the earliest examples of AI CSAM in circulation, explained how emerging tools functioned and set out clear predictions about the direction and capability of AI technology. What was a forecast is now an established and deeply concerning reality. Realistic, full-motion AI-generated child sexual abuse videos are now commonplace online.

In 2025 alone, the IWF identified 3,443 AI-generated child sexual abuse videos, a staggering 26,385% increase compared with the previous year, when just 13 such videos were recorded.

Although AI CSAM still accounts for a relatively small share of the total illegal content actioned by the IWF and hotlines around the world, the speed of its growth is alarming, and the quality and realism of this material is deeply disturbing.[4] IWF's evidence and insights directly challenge the claim that AI CSAM is less harmful than so-called "traditional" CSAM. Pre-existing images of abuse are routinely used to train these models, and we have seen imagery enhanced or reimagined using AI, bringing low resolution material into sharp focus.

Offenders are also exploiting AI chatbots to simulate sexual conversations with children and advertise the option to generate CSAM instantly, often requiring little technical skill or effort.

This report seeks to centre the human impact of AI CSAM, setting out clearly the harm caused to children and wider society. It captures the views of our expert Analysts, who are on the frontline of removing AI CSAM from the internet, as well as excerpts from dark web offender communities, where users openly celebrate the accessibility and sophistication of AI-generated abuse.

As an organisation with the legal permission to view, assess and retain child sexual abuse imagery, the IWF has worked with technology companies for many years to support the development and training of safety-focused technologies. As a global authority in this field, we stand ready to collaborate with researchers, institutions, industry partners, and other organisations to provide a strong, independent auditing function throughout every phase of development.

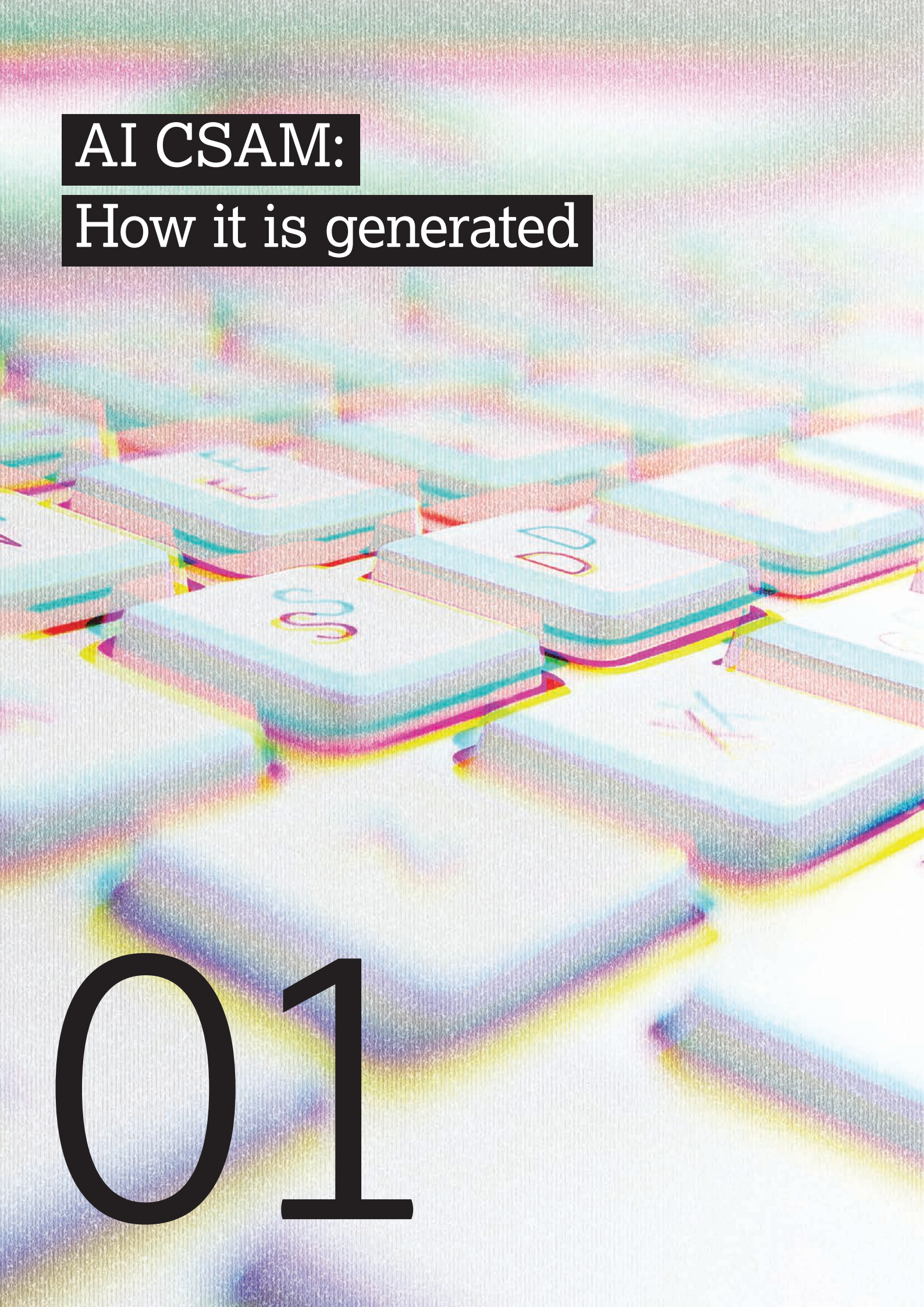
---

[2] Internet Watch Foundation (2024). *How AI is being abused to create child sexual abuse imagery.* Internet Watch Foundation. Available online at: https://www.iwf.org.uk/media/q4zll2ya/iwf-ai-csam-report_public-oct23v1.pdf

[3] Internet Watch Foundation (2024). *How AI is being abused to create child sexual abuse imagery. Internet Watch Foundation.* Available online at: https://www.iwf.org.uk/media/q4zll2ya/iwf-ai-csam-report_public-oct23v1.pdf

[4] Canadian Centre for Child Protection (2024) *Police and child protection agency say parents need to know about sexually explicit AI deepfakes.* Available at: https://www.protectchildren.ca/en/press-and-media/news-releases/2024/AI-deepfakes

AI CSAM:
How it is generated



01

Since the IWF first started monitoring AI in early 2023, we have seen a rapid, frightening advancement in the ability to artificially generate child sexual abuse imagery.[5] In this section, we summarise how AI CSAM is generated and explain the tools offenders use to create this material.

AI models and the generation of AI CSAM

AI CSAM is the umbrella term for image or video content depicting the sexual abuse of children that has been created with assistance from, or entirely by, generative AI systems.[6] These systems can operate across multiple modalities, including text, images, audio, video, or code. Multimodality applies to their inputs and outputs, meaning they can generate visual content from either text or images.

The key AI models that are of concern to the IWF are **large language models (LLMs)**, **nudify apps** and **text-to-image models**.

LLMs are AI systems trained on large text datasets to identify patterns and generate naturalistic text. These systems have dramatically risen in popularity and sophistication in the past few years, with popular examples including OpenAI's ChatGPT and Google's Gemini.[7] Please navigate to Section 05 to learn more about our concerns with AI chatbots, and Section 04 to read about nudify apps.

Text-to-image models, on the other hand, have the ability to generate images from textual prompts. Much like text, images are scraped from the internet and are annotated with descriptive labels, allowing the model to learn how to generate visuals in response to user instructions, known as prompts.

Prompts are broken into smaller pieces, or tokens, which the model matches to visual concepts learned from billions of training images, videos, and text descriptions. These concepts include shapes, colours, textures, and object labels. Popular text-to-image models include Midjourney Inc's Midjourney AI and Stability AI's Stable Diffusion. Open AI's Sora is also an example of a text-to-video model.

Because AI tools can recognise concepts such as "child" and "sexual activity", these concepts can be combined to generate CSAM unless effective safeguards are applied.

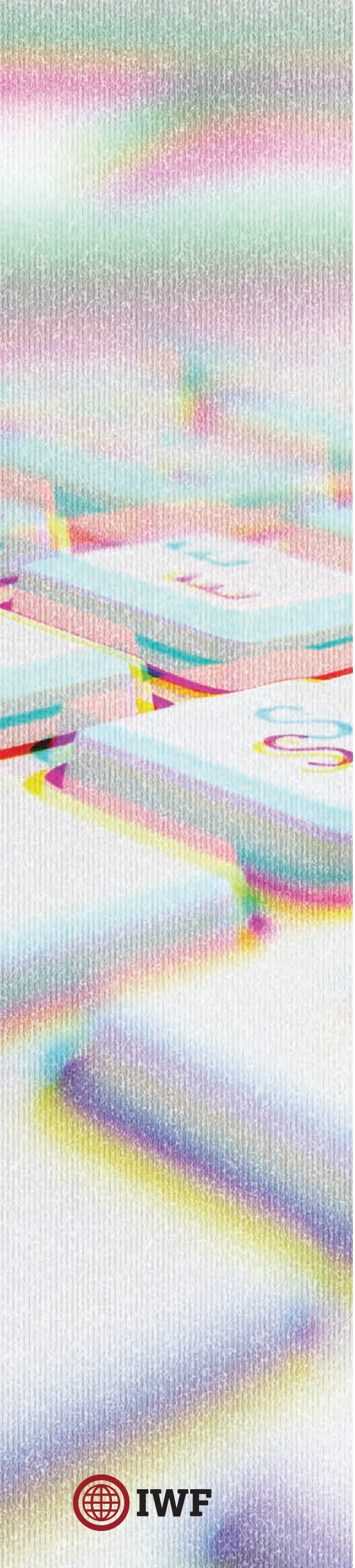
For more detailed information about how AI CSAM is generated, please refer to our October 2023 Report.[8] For more information on the illegality of AI CSAM, please navigate to Section 03.

[5] Internet Watch Foundation (2023). *What has changed in the AI CSAM landscape? Internet Watch Foundation.* Available online at: https://www.iwf.org.uk/media/nadlcb1z/iwf-ai-csam-report_update-public-jul24v13.pdf

[6] AI CSAM is illegal in the UK regardless of its photorealism or its means of production, but international legislation varies. See page 32 for AI CSAM laws in other jurisdictions.

[7] Internet Watch Foundation (2023). *How AI is being abused to create child sexual abuse imagery. Internet Watch Foundation.* Available online at: https://www.iwf.org.uk/media/q4zll2ya/iwf-ai-csam-report_public-oct23v1.pdf

[8] Internet Watch Foundation (2023). *How AI is being abused to create child sexual abuse imagery. Internet Watch Foundation.* Available online at: https://www.iwf.org.uk/media/q4zll2ya/iwf-ai-csam-report_public-oct23v1.pdf



Fine Tuning and Low-Rank Adaptation Models (LoRAs)

Fine-tuning is a process applied to a base model after its initial training. Instead of training a model from scratch, developers take a pre-trained model and further train it on a smaller, more specific dataset to adjust its output.

Low-rank adaption (LoRA) is a fine-tuning technique that is driving the rapid proliferation of AI CSAM. By reducing the technical skill, time, and financial resources required to customise generative models, LoRA has democratised the production of hyper-realistic imagery.

LoRA works by training small, additional layers that modify the behaviour of a base model, such as an LLM, using a limited dataset, while keeping the original model weights frozen. This makes it far more efficient than traditional fine-tuning. For example, a realistic deepfake model of a specific child can be created using as few as 20 to 30 existing images. The computational demands are minimal and can take as little as 15 minutes to train.

These low resource requirements mean that LoRAs can be developed on standard consumer grade laptops or desktop computers. As a result, the "means of production" have shifted from technical specialists to ordinary users, significantly expanding the pool of potential perpetrators.[9]

The small file size of LoRAs further compounds this risk. LoRAs are easy to share, reuse, and deploy across platforms, enabling rapid replication once a model has been created. As LoRA files are small and can run entirely offline, they are often highly resilient to efforts to remove this material from the internet.[10] Once downloaded, they can be redistributed privately and used without any ongoing connection to hosting platforms, making detection and enforcement exceptionally difficult.

FRONTLINE OBSERVATIONS

Our Analysts have observed multiple instances of LoRAs being circulated, including models trained on images of known victims.

In one dark web forum examined in December 2025, 202 individual models were available. These included models trained to generate accurate likenesses of identified child sexual abuse victims, child actors, influencers, and child sports celebrities; models designed to generate realistic genitalia for specific age groups; and models configured to produce images or videos depicting specific sexual acts or positions.

Metadata can sometimes be extracted from a trained model or fine-tuned LoRA, such as associated text captions or filenames of training images, and are also sometimes visible in the image output, with information on which models, software, and LoRAs were used in the creation process. However, these metadata can often be stripped entirely, further complicating efforts to trace the origin or

[9] Internet Watch Foundation (2025). *Preventing the upload of child sexual abuse material (CSAM) in E2EE environments*. Available online at: https://www.iwf.org.uk/media/21rpo2o4/iwf-preventing-the-upload-of-child-sexual-abuse-material-in-end-to-end-encrypted-e2ee-environments-v1.pdf

[10] Ciardha, Caoilte Ó, Buckley, J. and Portnoff, R.S. (2025). *AI Generated Child Sexual Abuse Material — What's the Harm?* Available online at: https://arxiv.org/abs/2510.02978

provenance of the material. These custom models turn one instance of abuse into a renewable, tradeable commodity, vastly amplifying the harm.

Gaps and inconsistencies in national and regional legislation on CSAM, including AI-generated content, create clear and significant risks: offenders can exploit legal loopholes, platforms face fragmented obligations, and children remain inadequately protected. Without harmonised and comprehensive laws, the rapid proliferation of AI tools further amplifies the potential scale and impact of abuse material.

Legislation on CSA image-generators

UK	In February 2025, the UK Government announced measures under the **Crime and Policing Bill** to close legal loopholes related to AI-generated CSAM.[11] This includes the introduction of a new criminal offence for making, adapting, possessing, supplying, or offering to supply a CSA image-generator. At time of publication, the Crime and Policing Bill is at Report Stage in the House of Lords and is likely to receive Royal Assent in the coming weeks.
US	No US federal law explicitly criminalises CSA image-generators. The **STOP CSAM Act**, reintroduced in May 2025,[12] expands liability for providers who host, store, or promote CSAM on their platforms. The production of AI-generated CSAM is prosecutable if it depicts a "real" child under federal CSAM laws, or under federal obscenity laws if it is entirely AI-generated.[13] Individuals creating such content can face criminal prosecution, while providers can be fined up to $1 million per incident, increased to $5 million if a child is harmed. A recent US district court ruling clarified that purely private possession of AI CSAM may be constitutionally protected, but production is not, leaving personal use partially exposed to legal risk.[14]
EU	The recast **Child Sexual Abuse Directive (2011/93/EU)** is currently under interinstitutional negotiation to modernise criminal definitions, strengthen penalties, and explicitly address AI-generated CSAM.[15] The proposal would criminalise creating, fine-tuning, possessing, distributing, or offering AI models designed to produce CSAM. Penalties are expected to align with broader CSAM offences, but final details are pending negotiations between the co-legislators. The IWF supports the European Parliament's ambitious position, including its provisions targeting CSA image-generators.

[11] Home Office (2025). *Britain's leading the way protecting children from online predators*. GOV.UK. Available online at: https://www.gov.uk/government/news/britains-leading-the-way-protecting-children-from-online-predators

[12] R-MO, J. (2025). *S.1829 — 119th Congress (2025-2026): STOP CSAM Act of 2025*. Congress.gov. Available online at: https://www.congress.gov/bill/119th-congress/senate-bill/1829

[13] Ll, S. (n.d.). *OLL25620 72X 119TH CONGRESS 1ST SESSION*. Available online at: https://www.hawley.senate.gov/wp-content/uploads/2025/05/Hawley-Stop-CSAM-Act.pdf

[14] EUR Lex (2024). *Proposal for a Directive of the European Parliament and of the Council on combating the sexual abuse and sexual exploitation of children and child sexual abuse material (recast).* Available online at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52024PC0060/

[15] Pfefferkorn, R. (2025). *Court Rules That Constitution Protects Private Possession of AI-Generated CSAM*. Tech Policy Press. Available online at: https://www.techpolicy.press/court-rules-that-constitution-protects-private-possession-of-aigenerated-csam

Open and closed-source models

Many generative AI models, like OpenAI's ChatGPT and Midjourney, are **closed-source**, which means they keep their code and inner workings confidential.[16] By contrast, **open-source models** make their underlying code publicly available, allowing anyone to download, customise, and fine-tune them.[17]

While open-source technology supports legitimate research and innovation, it can also enable bad actors to create personal, offline models which are trained on CSAM and capable of generating further imagery without significant technical expertise. Because these models operate locally, offer unrestricted prompting, and lack meaningful content moderation or detection mechanisms, open-source systems remain the tool of choice for offenders generating AI CSAM.

Although open-source models are currently estimated to lag several months behind cloud-based commercial systems in terms of sophistication, experts warn that this gap is narrowing.[18] Offenders using open-source tools are therefore likely to benefit from advances first seen in closed-source AI outputs, increasing both the realism and scale of harmful content in the near future.

Cloud-based AI models are defined by how they are deployed rather than by whether they are open or closed-source. These systems are accessed via the internet and run on remote servers operated by technology companies, rather than on a user's own device.

IWF Analysts have observed offenders advising against the use of cloud-based models, such as xAI's Grok Imagine, to generate CSAM, with some forums explicitly banning users who attempt to do so. These specific dark web offender communities frequently cite concerns about surveillance, data retention, and traceability, noting that activity on cloud-based platforms could potentially be identified or linked to an individual. Cloud-based models thus hold little appeal for this community, which prioritises anonymity and control.

Where AI CSAM appears

AI CSAM is hosted and distributed across a range of online environments, including dark web forums and the clear web, demonstrating that this is a current and active threat.

However, the observable material represents only a partial view of the overall landscape. The scale of fine-tuned CSAM models, including those capable of generating imagery of named children, is likely to be significantly greater than can be identified through monitoring publicly accessible areas of the clear web and the dark web alone. On dark web forums, IWF Analysts witness offenders frequently exchanging personal CSAM datasets and model resources within closed online communities to enable others to develop and refine bespoke systems, with some high-profile model-sharing threads attracting tens of thousands of views.

[16] Lemut, D. (2024). *Open-Source AI vs. Closed-Source AI: What's the Difference?* Available online at: https://www.multimodal.dev/post/open-source-ai-vs-closed-source-ai

[17] Marr, B. (2024). *Navigating The Generative AI Divide: Open-Source Vs. Closed-Source Solutions.* Forbes. Available online at: https://www.forbes.com/sites/bernardmarr/2024/04/22/navigating-the-generative-ai-divide-open-source-vs-closed-source-solutions

[18] Al Security Institute. (2025). *Frontier AI Trends Report.* Available online at: https://aisi.s3.eu-west-2.amazonaws.com/Frontier+AI+Trends+Report+-+AI+Security+Institute.pdf

A substantial proportion of AI CSAM also sits beyond the scope the IWF's visibility. Fine-tuned models and datasets can be hosted on offline devices and shared via end-to-end encrypted (E2EE) peer-to-peer networks, where both distribution and model development take place without reliance on centralised hosting. These environments significantly limit visibility for organisations such as the IWF and contribute to the underestimation of the true scale and distribution of AI CSAM.

To learn more about how platforms can prevent the upload of CSAM in E2EE environments in a privacy-preserving way, please refer to the IWF's paper on *Preventing the Upload of CSAM in E2EE environments*.[19]

OFFENDER QUOTE

"Boy Videos for sell, rare videos, a collection that every boy lover dreams of, price: 88 usd, only Litecoin is accepted message me on tele for preview if you are interested!"

PUBLIC REPORT

"He has been advertising this on Telegram via [tool redacted] which he uses multiple accounts to display inappropriate AI-Generated videos featuring minors. Searching "[username redacted]" will find all his accounts."

FRONTLINE OBSERVATIONS

AI CSAM is shared both on the dark web and the clear web. It's often posted alongside traditional CSAM. While on the dark web people discuss the creative process and share their workflows, on the clear web this is often monetised: users advertise the content they create.

We've also seen adverts for custom "courses" which could teach anyone to create realistic AI-generated images of teenagers. Many such adverts suggest the users to contact sellers on encrypted apps:

"If anyone knows any better AI image platforms, let me know please! I want to make my images as realistic for you all. You deserve the best. My telegram is ****"

[19] Internet Watch Foundation (2025) *Preventing the upload of child sexual abuse material (CSAM) in E2EE environments*. Available online at: https://www.iwf.org.uk/media/21rpo2o4/iwf-preventing-the-upload-of-child-sexual-abuse-material-in-end-to-end-encrypted-e2ee-environments-v1.pdf

AI CSAM harms

02

AI-generated sexual imagery of children causes real and lasting harm to victims – even if the physical abuse depicted did not take place. From a victim or survivor's perspective, whether the image originated from a camera or from software, the impact is deeply traumatic and a violation of their right to be free from torture and inhuman or degrading treatment.

Survivors of child sexual abuse frequently describe profound and enduring psychological harm, including feelings of shame, humiliation, anger, anxiety, and deep violation.[20] Many withdraw socially, struggle to trust others, and in some cases experience self-harm or suicidal thoughts.[21] A defining feature of child sexual abuse imagery is its permanence: once an image or video exists, it can be copied, altered, and redistributed endlessly, leaving victims with no sense of control or closure. In January 2026, Mara Wilson – who was a child actor in the classic film Matilda – described being a victim of AI-generated CSAM as "losing control of one's own narrative".[22]

There have been attempts to claim that AI CSAM causes less harm than "real" CSAM, either because it does not directly involve offline offending against children or because entirely synthetic material is said to have no "real" victim.[23] This has led to suggestions that AI material could function as a "safe"[24] or therapeutic alternative to traditional CSAM.[25]

UK law dictates that the generation of any kind of CSAM – traditional or fully synthetic – is illegal.[26] There is no evidence to support the claim that AI CSAM provides an effective harm-reduction method for offenders.[27] Instead, clear harms manifest offline as a result of the viewing, and sharing of AI CSAM.

[20] Vera-Gray, F. (2023). *Key Messages from Research on the Impacts of Child Sexual Abuse.* Centre of Expertise on Child Sexual Abuse. Available online at: https://www.csacentre.org.uk/app/uploads/2023/03/Key-messages-from-research-Impacts-of-child-sexual-abuse.pdf

[21] Wurtele, S. K. & Kenny, M. C. (2010). 'Primary prevention of child sexual abuse: Child- and parent-focused approaches.' In K. L. Kaufman (Ed.), *The Prevention of Sexual Violence: A Practitioner's Sourcebook* (pp. 107-119). Holyoke, MA: NEARI Press

[22] Wilson, M. (2026). *My picture was used in child abuse images. AI is putting others through my nightmare.* The Guardian. Available online at: https://www.theguardian.com/commentisfree/2026/jan/17/child-abuse-images-ai-exploitation

[23] Internet Watch Foundation (2025). *We see models which have been trained specifically on Child Sexual Abuse Material (CSAM), often trained on particular victims.* Linkedin.com. Available online at: https://www.linkedin.com/posts/iwf_we-see-models-which-have-been-trained-specifically-activity-7311357201766359041-DrDN

[24] Internet Watch Foundation (2025). *AI child sexual abuse imagery is not a future risk – it is a current and accelerating crisis*. Available online at: https://www.iwf.org.uk/news-media/news/ai-child-sexual-abuse-imagery-is-not-a-future-risk-it-is-a-current-and-accelerating-crisis

[25] ohchr.org. (2024). *Online Child Sexual Abuse and Exploitation: Current and Emerging Threats — Contribution to the Special Rapporteur on the sale and sexual exploitation of children's call for input: Existing and emerging sexually exploitative practices against children in the digital environment.* Protect Children / Suojellaan Lapsia ry. Available online at: https://www.ohchr.org/sites/default/files/documents/issues/children/sr/cfis/existing-emerging/subm-existing-emerging-sexually-cso-suojellaan-lapsia-ry.pdf

[26] AI Security Institute (2025). *Our approach to tackling AI-generated child sexual abuse material | AISI Work*. Available online at: https://www.aisi.gov.uk/blog/our-approach-to-tackling-ai-generated-child-sexual-abuse-material

[27] Internet Watch Foundation (2025). *'AI child sexual abuse imagery is not a future risk – it is a current and accelerating crisis'*. Available online at: https://www.iwf.org.uk/news-media/news/ai-child-sexual-abuse-imagery-is-not-a-future-risk-it-is-a-current-and-accelerating-crisis

This section breaks down the harm of AI CSAM, debunking claims that it is less harmful than other forms of CSAM. It explains how AI revictimises victims and survivors, perpetuates gendered harms, normalises sexual violence against children, and causes operational harm to child protection systems.

Revictimisation

In as early as 2023, the IWF documented how AI tools are used to manipulate existing child sexual abuse material involving survivors known to our Analysts. As highlighted, perpetrators take existing abuse images and use AI to generate new – frequently more extreme – versions. For survivors, this amounts to being abused again and again, years after the original harm, with no ability to stop it.

This is not theoretical. Our 2024 report described the story of a child we refer to as "Olivia". Between the ages of three and eight, Olivia suffered horrific sexual abuse at the hands of someone she trusted. She is well known to IWF Analysts as the imagery the perpetrator took has circulated widely on the internet.

In recent years, we have documented LoRAs – available to download for free – that offenders have trained on Olivia's imagery. Olivia is now in her 20s, but her suffering is potentially without end.

In 2024, The Guardian told the story of Megan, an anonymised survivor of child sexual abuse. Megan describes how AI has made her trauma feel ongoing, despite the original abuse occurring over a decade ago. Over 12 months, the prospect of her images being digitally manipulated became increasingly distressing, as she describes:

> "AI gives perpetrators the chance to create even more situations of my abuse to feed their own fantasies and their own versions. The way my images could be manipulated with AI could give the false impression it was not harmful or that I was enjoying the abuse."[28]

[28] McQue, K. (2024). *Child predators are using AI to create sexual images of their favorite 'stars': 'My body will never be mine again. The Guardian*. Available online at: https://www.theguardian.com/technology/article/2024/jun/12/predators-using-ai-generate-child-sexual-images
Image source: theguardian.com screenshot

Whether images are created through traditional means or generated by AI, the effect on children is the same: ongoing exploitation, loss of control, and repeated trauma. A 2017 survey of victims and survivors conducted by the Canadian Centre for Child Protection (C3P) found that nearly 70% of surveyed survivors of child sexual abuse imagery experience constant fear at the possibility of being recognised by someone who has seen the images; 30% said this had already happened. Two-thirds cited the endless circulation of images as a central source of trauma, while more than a third described profound feelings of powerlessness. 87% experienced additional harm, including stalking, sexual propositions, blackmail, and renewed victimisation, as a result of the original abuse they suffered.[29]

These findings make clear that repeated revictimisation through child sexual abuse imagery causes serious, long-term psychological damage. All forms of child sexual abuse imagery must be treated with the same urgency.

FRONTLINE OBSERVATIONS

"The content has become so realistic and also the speed at which this technology is developing is really alarming."

"Now we are seeing new images of [her] — images generated by AI — some of them are even more severe than the images that were actually taken in reality."

"This is as far from a victimless crime as it gets — there's a very real victim here and I think real harm is being done by this content."

Gendered harms

Notable cases and data have highlighted the distinct gendered dimension to the requesting, creation, and sharing of child sexual abuse images.

This is particularly acute when it comes to AI-generated material. In 2025, girls continued to be the most frequently depicted sex in AI-generated imagery, with girls seen in 97% of images where sex was recorded this year and 99% in 2024.

In 2023, there was widespread coverage of a case in Spain where 20 girls aged 11-17 had come forward as victims of nudifying apps. 11 boys were identified as involved in either the creation of these images or their circulation via platforms Telegram and WhatsApp.[30] Another example in the EU is that of a male high school student in Romania who used deepfake technology to "undress" 11 of his classmates. It was reported that each of the victims had a connection with the student, and his actions – virtually undressing them – was an attempt to exercise power over them.[31]

In 2024, South Korea experienced an industrial-scale crisis of AI-generated sexual abuse imagery, where thousands of deepfake images were created and circulated via coordinated networks on Telegram, spanning more than 200 schools, universities, and even the military. Teenagers and young people in their 20s accounted for 92% of victims, with girls and young women being the primary

[29] Canadian Centre for Child Protection (2017). *Survivors" Survey: Executive Summary 2017.* Available online at: https://content.c3p.ca/pdfs/C3P_SurvivorsSurveyExecutiveSummary2017_en.pdf

[30] BBC News. (2023). *Children making AI-generated child abuse images, says charity.* Available online at: https://www.bbc.co.uk/news/technology-67521226

[31] Scena9.ro. (2025). *Tehnologie, rușine și tăcere: Ce învățăm dintr-un caz de nud deepfake într-un liceu din România*. Available at: https://www.scena9.ro/article/tehnologie-rusine-nud-deepfake-eleve-liceu-arad-exmatriculare

targets. More than 80% of those ultimately arrested were teenagers themselves. The scale, speed, and hyper-local targeting of the abuse pushed the government to criminalise possession and viewing of deepfake sexual material. The new legislation in South Korea criminalises the possession and viewing of sexually explicit deepfakes and increases the existing sentences for creation and sharing of content of this nature to seven years.[32]

These patterns reflect persistent, gendered sexual violence online, reinforced by cultures that normalise misogyny and the humiliation of girls. For more information on nudify apps, please see Section 04.

The 2024 Girlguiding Annual Survey found that 59% of girls and young women aged 11-21 were concerned that AI may be used to create fake images of them, or to impersonate them online.[33] This concern is echoed by 2026 data from the UK Safer Internet Centre, which found that 60% of young people are worried about someone using AI to create inappropriate images of them, 61% of 13–17-year-olds are concerned about peers generating sexual images of other young people, and even 63% of children aged 8–12 share this fear.[34]

Such cases illustrate how AI tools can reproduce and amplify misogyny within online environments, lowering barriers to abuse and accelerating the production of harmful material at scale.

It is also important to note that when it comes to child sexual abuse imagery more broadly, boys are also harmed by gender-based violence and abuses of power. Younger boys are more likely than girls to appear in commercially produced and distributed abuse material, and older teen boys increasingly reporting imagery to us directly via Report Remove, a service to help young people under 18 in the UK to confidentially report sexual content of themselves and remove them from the internet.[35] We note that teenage boys are particularly targeted for sexually motivated extortion, a crime which may also be aided by AI technology.

Normalisation of sexual violence against children

Consumption of realistic synthetic material reinforces existing sexual interests in children rather than providing a safe substitute and may increase the risk of escalation.[36] Engagement with any form of CSAM has been shown to reinforce existing sexual interest in children.[37] Research by Protect Children Finland (Suojellaan Lapsia) illustrates the potential progression from viewing CSAM to contact offending: their 2021 "*Help us to Help You*" survey found that 44% of the 5,171 respondents said that viewing CSAM made them think about seeking direct contact with children, and 37% reported having sought such contact after viewing CSAM.[38]

---

[32] Groves, Adam and Yummi Tae, Sarah. (2025) '*Briefing: The Anatomy of the South Korean Deepfake Crisis*'. Available online at: https://tacteen.net/attach_file/?n=31852

[33] Girlguiding. (2024). Available at: https://www.girlguiding.org.uk/globalassets/docs-and-resources/research-and-campaigns/girls-attitudes-survey-2024.pdf

[34] A report by the UK Safer Internet Centre for Safer Internet Day 2026. Available at: https://d1xsi6mgo67kia.cloudfront.net/uploads/2026/02/Safer-Internet-Day-2026-Full-Research-Report.pdf

[35] Internet Watch Foundation. (2026) *Report Remove | IWF*. www.iwf.org.uk. Available at: https://www.iwf.org.uk/our-technology/report-remove

[36] Ciardha, Caoilte Ó, Buckley, J. and Portnoff, R.S. (2025). *AI Generated Child Sexual Abuse Material — What's the Harm?* arXiv.org. Available online at: https://arxiv.org/abs/2510.02978

[37] Carr, A. (2013). 'The social dimension of the online trade of child sexual exploitation material,' in *Understanding and Preventing Online Sexual Exploitation of Children*. pp.112 – 31. Routledge

[38] Protect Children (2021). *CSAM Users in the Dark Web: Protecting Children Through Prevention*. Suojellaan Lapsia. Available online at: https://www.suojellaanlapsia.fi/en/post/csam-users-in-the-dark web-protecting-children-through-prevention. For the first statistic, see Question 17. For the second, see Question 18. For the third, see Question 19

These findings are reinforced in Protect Children Finland's 2022 global sample study of the responses of 1,546 anonymised individuals, which showed that 58% of CSAM users experience at least occasional fear that their consumption could lead to sexual acts against children, with nearly a quarter (24%) reporting this fear weekly or almost every time.[39] In the same study, 10% of respondents reported that they had sought contact with children online after viewing CSAM or other illegal violent content at least weekly or more often. This demonstrates a clear link between CSAM consumption, desensitisation, and lower inhibition towards further online, or offline, sexual offending.

OFFENDER QUOTES

"These are truly stunning. Some of the realism in these is about 95% of the way to indistinguishable from real photos."

"For now, AI images can be spotted in most cases, but it is getting to the point where AI will be indistinguishable. I think there are opportunities for AI to be used to our advantage in playing an activist role."

FRONTLINE OBSERVATIONS

Our Analysts have observed offenders commenting on how AI imagery can be generated of children known to offenders.

"I sometime wonder why there isn't more real imagery from nurseries? With AI it doesn't matter anymore. We can create all the fun we want."

One offender reflected on the possibility of using hidden cameras to take innocuous footage of children known to them and convert this into AI CSAM videos with an AI image-to-video function.

"I'm really impressed with the results of image to video conversions. I think it would be really exciting to (generate CSAM) all from the subjective perspective of my camera. Are there any AIs that can use video to continue it with the generated content? Or is it better to only provide a snapshot or a still image?"

These examples clearly illustrate how children – living in the same area as offenders – may be harmed by AI CSAM.

Operational harm

The scale and speed of AI CSAM production runs the risk of overwhelming already inundated reporting mechanisms and law enforcement capacity. Law enforcement agencies must now grapple with identifying victims who need to be safeguarded, alongside navigating patchwork legal frameworks.[40]

[39] Insoll, T., Ovaska, A.K., Nurmi, J., Aaltonen, M. and Vaaranen-Valkonen, N. (2022).' Risk Factors for Child Sexual Abuse Material Users Contacting Children Online.' *Journal of Online Trust and Safety, 1(2).* Available online at: https://doi.org/10.54501/jots.v1i2.29

[40] Europol (2024). *Internet Organised Crime Threat Assessment (IOCTA) 2024.* Available online at: https://doi.org/10.2813/442713

The rapid proliferation of AI CSAM creates a "needle-in-a-haystack" challenge for investigators. Distinguishing synthetic material from photographic abuse consumes significant time and resources, delaying responses to children in immediate danger.[41] This dynamic increases the risk of false negatives, where children in need of rescue go undetected, and false positives, where resources are diverted to synthetic cases, straining already limited law enforcement capacity.[42]

The spread of synthetic imagery also risks undermining criminal proceedings, particularly in jurisdictions where AI-generated child sexual abuse material is not criminalised in the same way as other types of material. In February 2025, Europol supported authorities from 19 countries in "Operational Cumberland", a collaborative, large-scale operation which led to 273 suspects identified and 25 arrests worldwide. The suspects were part of an international criminal group whose members were engaged in the distribution of CSAM fully generated by AI. Europol noted at the time that this was one of the first cases in the world where the lack of national legislation addressing AI CSAM made **"it exceptionally challenging for investigators"**.[43]

In these jurisdictions, offenders may claim that genuine evidence of contact abuse was generated using AI and therefore does not depict an identifiable child, a tactic sometimes described as the "deepfake defence".[44] This approach draws on what researchers call the "liars' dividend",[45] whereby the growing public awareness of synthetic media creates plausible deniability. Although such arguments in the context of AI CSAM are limited at present, courts are already encountering disputes about synthetic media, and experts warn that increasing awareness of generative AI may enable defendants to cast doubt on authentic material.[46]

While the UK already has a robust legal framework that criminalises the creation, possession, and distribution of CSAM, including AI-generated content, this level of protection is not yet matched across all jurisdictions. In the United States, federal law remains fragmented, with most progress occurring at state level, while in the European Union there is currently no uniform prohibition on AI-generated CSAM and gaps persist between Member States.

This uneven legal landscape creates clear risks: offenders can exploit regulatory loopholes, platforms face inconsistent obligations, and children are left inadequately protected depending on where abuse material is created or hosted. As generative AI lowers barriers to producing harmful content at scale, the absence of harmonised and comprehensive laws increasingly undermines global child safety efforts.

---

[41] Our Rescue. (2025). *Sounding the Alarm: Why the Fight Against CSAM Can't Wait | Our Rescue.* Available online at: https://ourrescue.org/resources/sex-trafficking/human-trafficking/human-trafficking-and-child-exploitation/sounding-the-alarm-why-the-fight-against-csam-cant-wait

[42] Janjeva, A., Harris, A., Mercer, S., Kasprzyk, A. and Gausen, A. (2023). *The Rapid Rise of Generative AI Assessing risks to safety and security Acknowledgements.* Available online at: https://cetas.turing.ac.uk/sites/default/files/2023-12/cetas_research_report_-_the_rapid_rise_of_generative_ai_-_2023.pdf

[43] Europol. (2022). *25 arrested in global hit against AI-generated child sexual abuse material | Europol.* Available at: https://www.europol.europa.eu/media-press/newsroom/news/25-arrested-in-global-hit-against-ai-generated-child-sexual-abuse-material

[44] Delfino, R. (2023). *The Deepfake Defense—Exploring the Limits of the Law and Ethical Norms in Protecting Legal Proceedings from Lying Lawyers. SSRN Electronic Journal.* Available at: https://doi.org/10.2139/ssrn.4355140

[45] Interpol (2024). *Unmasking the Threat of Synthetic Media for Law Enforcement Beyond Illusions.* Available online at: https://www.interpol.int/content/download/21179/file/BEYOND%20ILLUSIONS_Report_2024.pdf

[46] Interpol (2024). *Unmasking the Threat of Synthetic Media for Law Enforcement Beyond Illusions.* Available online at: https://www.interpol.int/content/download/21179/file/BEYOND%20ILLUSIONS_Report_2024.pdf

Legislation on the creation, possession and distribution of AI CSAM

UK

The legislation on CSAM content is already robust in the UK. Creating, possessing, and distributing child sexual abuse imagery, AI or otherwise, is already illegal.

Within the **Protection of Children Act 1978**, Section 1 criminalises the taking, distribution and possession of an "indecent photograph or pseudo photograph of a child" (anyone under the age of 18).[47]

The Coroners and Justice Act (2009) criminalises CSAM, even if the material is not photorealistic. It is an offence to possess a prohibited image of a child, including AI-generated images which are not deemed photorealistic, including cartoons, illustrations and animations (Section 62 of the Coroners and Justice Act 2009).[48]

US

The US does not currently have a single, comprehensive law addressing AI CSAM. Federal legislative proposals introduced in 2025 aim to supplement state-level developments and address AI-enabled sexual exploitation. At the State level, 45 states have enacted statutes explicitly criminalising AI-generated or computer-edited CSAM, often treating it as equivalent to traditional CSAM for prosecution purposes.[49]

The Stop Sextortion Act, introduced in 2025,[50] would create new criminal offences regarding the sexual extortion of minors, including extending existing obscenity laws to cover AI-generated CSAM.

The SAFE Act (Sentencing Accountability for Exploitation Act), introduced in 2025,[51] would require federal sentencing guidelines to be updated to reflect the use of modern digital and AI-enabled technologies in sexual exploitation offences, including to conceal the offence, identity, or location of the defendant or victim.

EU

There is currently no uniform EU-wide prohibition on AI-generated CSAM, and legal definitions vary across Member States.

The Child Sexual Abuse Directive (Recast of Directive 2011/93/EU)[52] holds the potential to fully criminalise AI generated CSAM, though the Council of the EU's position proposes a "personal use" exemption (Article 5(10)), creating a potential loophole.

The AI Act (2024)[53] establishes a harmonised EU framework for AI, recognising CSAM as a systemic risk and requiring providers to implement risk mitigation measures. However, the Act is not a criminal law instrument.

[47] *Protection of Children Act 1978*, c. 37. Available online at: https://www.legislation.gov.uk/ukpga/1978/37

[48] *Coroners and Justice Act 2009*, c. 25. Available online at: https://www.legislation.gov.uk/ukpga/2009/25/part/2/chapter/2

[49] enoughabuse.org. (2025). *State laws criminalizing AI-generated or computer-edited child sexual abuse material (CSAM)*. (n.d.). Available online at: https://enoughabuse.org/get-vocal/laws-by-state/state-laws-criminalizing-ai-generated-or-computer-edited-child-sexual-abuse-material-csam/

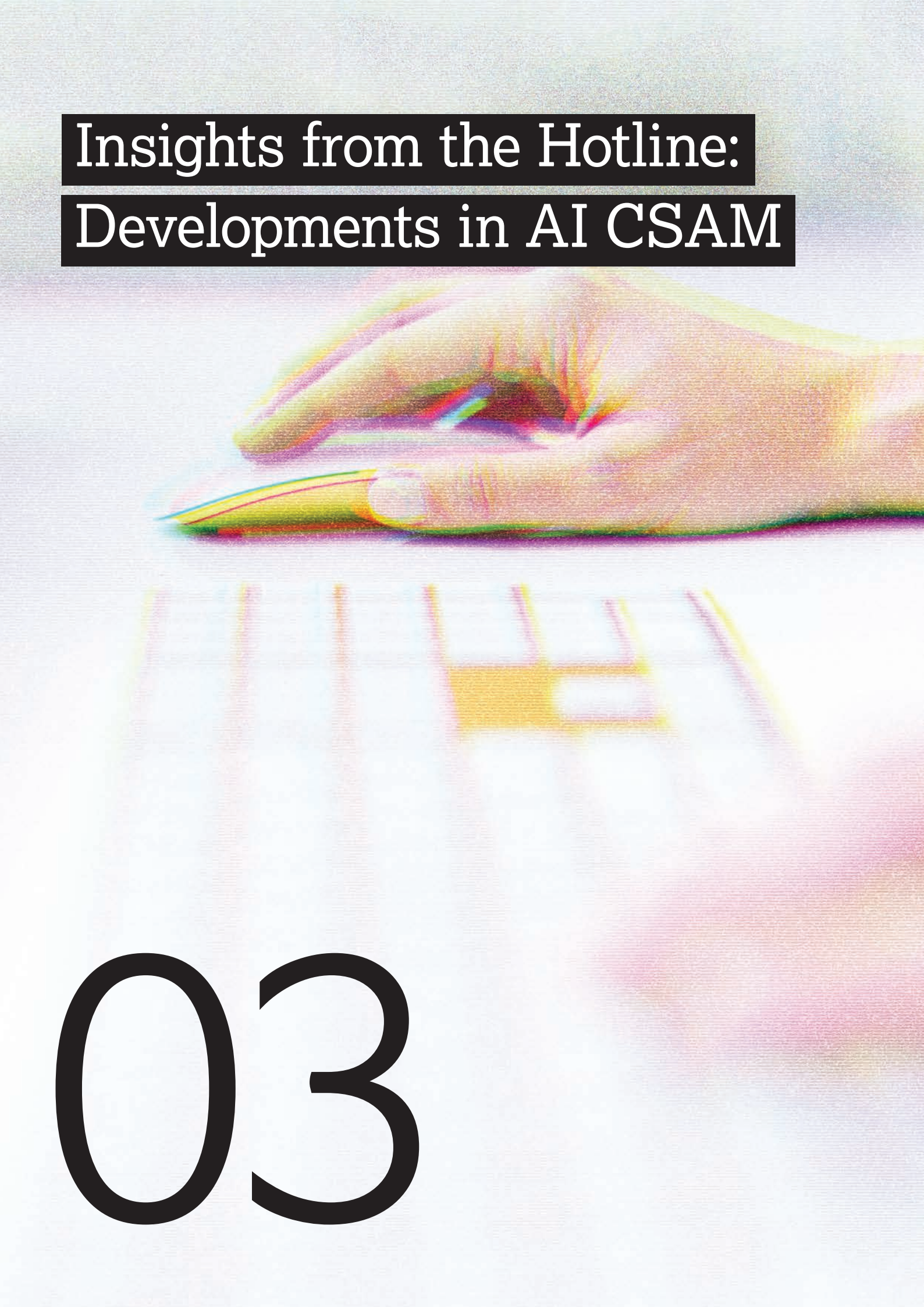
[50] congress.gov. (2025). S.3398 — *Stop Sextortion Act (119th Congress)*. (n.d.). Available online at: https://www.congress.gov/bill/119th-congress/senate-bill/3398/text

[51] R-IA, C. (2025). *Text — S.3394 — 119th Congress (2025-2026): SAFE Act*. Congress.gov. Available online at: https://www.congress.gov/bill/119th-congress/senate-bill/3394/text

[52] eur-lex.europa.eu. (2024). *Proposal for a Directive of the European Parliament and of the Council on combating the sexual abuse and sexual exploitation of children and child sexual abuse material (recast)*. (n.d.). Available online at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52024PC0060

[53] artificialintelligenceact.eu. (n.d.). *The Artificial Intelligence Act*. Available online at: https://artificialintelligenceact.eu/

Insights from the Hotline: Developments in AI CSAM



03

The IWF is one of the only Hotlines globally authorised to proactively seek out CSAM and act on public reports. This enables the IWF to horizon-scan and track developments in CSAM offending landscape including through offender discussions.

AI CSAM is a growing concern and is being closely monitored and actioned by the IWF, particularly as AI image-generation tools continue to evolve rapidly and become more accessible.

In 2025, we assessed 8,029 AI-generated images and videos as showing realistic child sexual abuse. While AI CSAM still represents a relatively small proportion of overall reports to the IWF, volumes continue to rise steadily.

This trend is mirrored internationally. In a UNICEF, ECPAT and INTERPOL study across 11 countries, at least 1.2 million children disclosed having had their images manipulated into sexually explicit deepfakes in the past year. In some countries, this represents 1 in 25 children – the equivalent of one child in a typical classroom.[54]

It is worth noting that, as AI technology advances, there is an increasing likelihood that AI-generated imagery is assessed as photographic material. Analysts currently record content as AI-generated only when this is indicated by the source or metadata. The true volume of AI-generated images and videos assessed by our Analysts may therefore be higher.

AI CSAM Videos

Our July 2024 AI CSAM report marked the first time IWF Analysts documented fully synthetic AI CSAM videos in circulation. At that stage, the technology was relatively unsophisticated. Videos were characterised by heavy glitching and flickering, with unnatural, jerky movements that made AI-generated content straightforward to distinguish from real footage.

FRONTLINE OBSERVATIONS

"In terms of video quality, child sexual abuse imagery creators are leaps and bounds ahead of where they were last year.

The first AI child sexual abuse videos we saw were deepfakes – a known victim's face put onto an actor in an existing adult pornographic video. It wasn't sophisticated but could still be pretty convincing. The first fully synthetic child sexual abuse video we saw at the beginning of last year was just a series of jerky images put together, nothing convincing.

But now they have really turned a corner. The quality is alarmingly high, and the categories of offence depicted are becoming more extreme as the tools improve in their ability to generate video showing two or more people. The videos also include sets showing known victims in new scenarios.

Just as still images jumped to photorealistic as demand increased and the tools were improved; it was only a matter of time before videos went the same way."

[54] UNICEF (2025). *'Deepfake abuse is abuse' Statement by UNICEF on AI-generated sexualised images of children*. Available at : https://www.unicef.org/press-releases/deepfake-abuse-is-abuse

In 2025, IWF identified

3,443

AI-generated child sexual abuse videos

26,385%

increase in AI-generated CSAM videos

65%

of AI CSAM videos were Category A

30%

of AI CSAM videos were Category B

Since then, the technology has advanced significantly. As predicted, realistic full-motion AI video content is now commonplace.

In 2025, the IWF had identified 3,443 AI-generated child sexual abuse videos, an increase of 26,385% compared with the previous year, when just 13 such videos were recorded. Of these, 65% (2,233 videos) were classified as Category A, the most severe legal category under UK law, encompassing offences such as rape, sexual torture, and bestiality. A further 30% (1,027 videos) fell into Category B, the second most serious classification.

By comparison, for the full year (2025), 43% of non-AI criminal videos seen by the IWF were Category A – demonstrating that AI is being used to create more violent content.[55]

Any photographic or pseudo-photographic CSAM is further categorised by severity, under the UK Sentencing Council Categories. These are:

- **Category A** Images depicting penetrative sexual activity; image involving sexual activity with an animal; or sadism.
- **Category B** Images depicting non-penetrative sexual activity.
- **Category C** Other indecent images not falling within categories A or B.

For a deep dive on how AI videos are generated, please see our 2024 Report.[56]

AI CSAM with an audio component

AI CSAM with an audio component is an emerging area of concern. This may take the form of recordings of abuse, "audio deepfakes", which synthetically generate the sexualised voices of children, or manipulated recordings used to simulate abusive scenarios.

New AI models allow users to combine a picture, audio input, and text description to automatically generate a video with a synthetic voice.

While not yet as widespread as image or video CSAM, there are early indications of demand for such content in offender communities. IWF Analysts have seen AI CSAM videos which appear to have a cloned voiceover on them – several new models have been released recently where a user can use a picture, add audio, describe what they wish to see, and the model then turns the image and voice into a video. If left unaddressed, this demand is likely to grow alongside advances in generative AI technologies.

One example noted by our Analysts was a fully synthetic video showing a child who appeared to be between three and six-years old speaking to the camera and performing a sexual act on an adult man. Both the video and audio were generated by AI.

To note, IWF Analysts do not typically assess audio-only reports, given the IWF can only categorise photographic or pseudo-photographic content.

[55] Internet Watch Foundation (2026). *AI becoming 'child sexual abuse machine' adding to 'dangerous' record levels of online abuse, IWF warns.* Available online at: https://www.iwf.org.uk/annual-data-insights-report-2024/data-and-insights/ai-generated-child-sexual-abuse

[56] Internet Watch Foundation (2024). *What has changed in the AI CSAM landscape? Prompt: from fantasy to photo-realistic reality.* Internet Watch Foundation. Available online at: https://www.iwf.org.uk/media/nadlcb1z/iwf-ai-csam-report_update-public-jul24v13.pdf

Realism

Earlier iterations of AI CSAM were typically characterised by an airbrushed appearance, uniform lighting, and a glossy or artificial finish. Current material demonstrates far greater variation in composition, lighting, and quality as the technology has advanced. Images increasingly resemble amateur photographs, which is consistent with the characteristics observed in traditional CSAM.

Analysts note that AI CSAM produced in 2025 sometimes appears deliberately imperfect to emulate amateur photography, making it even more difficult to distinguish AI CSAM.

IWF Analysts are no longer identifying irregularities in the generated content, aside from minor blurring that is difficult for untrained eyes to detect.

OFFENDER QUOTES

"WOW AIs really made a leap forward with motion, it used to be obvious! 10/10"

"I just love knowing that the quality… as great as it is at this point, it's only going to get better!"

More complex, multi-subject videos remain less sophisticated, reflecting the greater technical challenges involved. Issues are most commonly observed in background elements and secondary figures, including distortions such as missing or inconsistent eyes and limbs.

As each additional subject introduces further complexity, there are more opportunities for the technology to fail. However, Analysts note that these limitations can be mitigated through post-generation editing by individuals with sufficient time and basic familiarity with commonly available software.

FRONTLINE OBSERVATIONS

"Overall, when we started seeing these images in 2023, it was very much a single subject – a child standing there like a mannequin.

Now we're looking at children engaged in various scenarios with adults, with other children, sometimes with animals, in all kinds of different settings, scenarios, sometimes very much fantasy scenarios, in highly unrealistic settings, but they look real."

These concerns are reflected in offender behaviour. On a dark web child sexual abuse forum surveyed by IWF Analysts in 2024, 90% of AI-generated images assessed were considered realistic enough to fall under the same legal framework as photographic CSAM.[57]

[57] Internet Watch Foundation (2024). *Annual Data and Insights Report 2024*. Available online at: https://www.iwf.org.uk/annual-data-insights-report-2024/data-and-insights/ai-generated-child-sexual-abuse/

OFFENDER QUOTE

"When I first started out, the imagery I was making really wasn't up to scratch. There was a lot of trial and error, and many hours practise to get better, but it is so worth it. I think you should stick with it, the results can be amazing."

Extremity

A defining characteristic of AI CSAM is that it does more than replicate existing abuse material. It actively widens the spectrum of harm by making it possible to fabricate scenarios. Perpetrators can design entirely novel forms of exploitative material tailored to their personal fantasies.

Analysis of offender forum discussions, alongside IWF data, indicates a growing preference for the most extreme Category A child sexual abuse material.

In 2025, IWF assessed 3,443 AI-generated videos that showed realistic child sexual abuse. Nearly two-thirds (65%) of AI videos were assessed as Category A.[58] In comparison, criminal assessments of non-AI-generated videos showed that 43% were attributed to Category A severity.

This marked difference suggests that perpetrators are increasingly using AI tools to generate highly explicit, extreme and complex scenarios, extending beyond previously observed offending patterns.

Analysts assess that AI is enabling the creation of material that reflects offenders' most extreme sexual interests with unprecedented speed and accessibility.

OFFENDER QUOTE

"The more I experiment, the more surprised I am at just how uncensored [Redacted model name] is. The Edit version and any finetunes are going to be NUTS."

Offenders themselves report that AI significantly lowers both practical and psychological barriers to production. The technology facilitates the rapid generation of imagery that would previously have required access to real victims, extensive technical expertise, or sustained offending behaviour.

The capacity to generate personalised, high-severity material instantly and repeatedly raises serious concerns about behavioural escalation. Easy access to increasingly extreme synthetic content risks eroding psychological barriers, potentially reinforcing harmful preferences and lowering inhibitions towards contact offending.[59]

[58] Internet Watch Foundation. (2025) *AI imagery getting more 'extreme' as IWF welcomes new rules allowing thorough testing of AI tools*. Available at: https://www.iwf.org.uk/news-media/news/ai-imagery-getting-more-extreme-as-iwf-welcomes-new-rules-allowing-thorough-testing-of-ai-tools

[59] Ciardha, Caoilte Ó, Buckley, J. and Portnoff, R.S. (2025). *AI Generated Child Sexual Abuse Material — What's the Harm?* arXiv.org. Available online at: https://arxiv.org/abs/2510.02978

Accessibility

Advances in AI have driven the convergence of tools that previously required separate capabilities. Single applications can now generate abusive imagery with minimal effort, removing the need for technical expertise and significantly lowering barriers to entry. This increased accessibility heightens the likelihood that AI CSAM will be created and that harmful offender behaviours will be reinforced.

The ease of access to these tools also enables offenders to realise, rehearse, and refine sexual fantasies involving children, including the generation of highly realistic or identifiable representations. In some cases, AI systems advertise the ability to combine image generation with the simulation of sexualised conversations with a child within the same platform, further eroding traditional distinctions between different forms of offending behaviour. These 'one-stop' services create immersive environments in which users can interact with an AI child, request the generation of abusive imagery, and, in some instances, receive guidance on grooming behaviours. In June 2025, the IWF received its first reports of AI CSAM found on AI chatbots. For further information on AI chatbots, see Section 05.

OFFENDER QUOTES

"Technology moves so fast – just when I finally understand how to use a tool, something newer and better comes along."

"Woah, these models [AI models of children] are so sexy. I'd love to be able to use some of these. I could create just an endless amount of images."

The increasing realism of this material reflects both the advancing sophistication of generative models and the growing proficiency of offenders in their use. Whereas previously offenders relied on additional tools, such as LoRAs, to generate imagery resembling a specific child, some newer models are able to generate highly convincing CSAM from just a single reference image without additional training, further accelerating misuse and reducing technical friction.

OFFENDER QUOTE

"I'm currently having the best results with [Model name redacted]. I get pics like that pretty decent without any LoRAs at all."

IWF Analysts have also been closely tracking the emergence of early agentic AI tools and assessing how they could influence the spread of CSAM online. Agentic AI refers to systems capable of achieving specific goals with minimal human supervision. This means tasks that once required sustained technical skill — building, updating, and maintaining online platforms — can increasingly be handled by autonomous systems with minimal oversight.

As agentic AI agents reduce the expertise needed to operate and scale websites or apps, barriers to entry are lowered and network resilience is strengthened. This growing accessibility increases the risk that CSAM distribution infrastructures will expand in scale, persistence, and impact.

OFFENDER QUOTES

"Thanks for your continuing hot sexy videos. Don't ever stop. In a while you can tell your AI agent to pick out clips and create the video by itself."

"I believe in a year or two we will be able to create our own movies just by feeding a prompt to an uncensored AI agent, no skills with editing or tech will be required."

Disguise Tools

Perpetrators have been seen to discuss the possibility of augmenting images of existing abuse with AI tools to disguise themselves, the child or parts of the setting in which the images were taken to avoid detection. This tactical choice can be made to disguise the perpetrator, the child, or the setting by layering synthetic elements onto authentic footage of abuse.[60] This further complicates the investigation and potential rescue process for law enforcement.

IWF Analysts have observed this trend directly in offender spaces. In one forum discussion monitored by the IWF, users were actively exchanging advice on how to modify existing photographic CSAM by altering faces and other easily recognisable features using AI tools. The conversation focused on how these techniques could be used to evade identification while preserving the abusive nature of the material.

Commercialisation

Offender communities actively develop and trade customised tools on dark web forums, including distributing fine-tuned models, swapping datasets, and sharing detailed instructions for model creation and use. Popular threads on these forums, such as those for CSAM model development, have been viewed tens of thousands of times.

This ecosystem of custom models, tool-sharing, and marketplaces transforms single instances of abuse into renewable, tradeable commodities, vastly amplifying harm. Perpetrators often share personal CSAM datasets to help others build their own models. While AI tools are becoming more accessible, there still remains a market for offenders who are skilled in using these tools to create high-quality material.

Together, these practices create a highly efficient, self-reinforcing ecosystem that maximises both the reach and impact of abuse material.

[60] Ciardha, Caoilte Ó, Buckley, J. and Portnoff, R.S. (2025). *AI Generated Child Sexual Abuse Material -- What's the Harm?* arXiv.org. Available online at: https://arxiv.org/abs/2510.02978

Fine-tuning techniques allow AI to generate images of specific individuals or produce highly realistic features such as lifelike eyes. This enables the bulk, repeatable production of images featuring known victims or named children on demand. Offenders also use AI-assisted upscaling to improve the quality of older, low-resolution CSAM clips, increasing their desirability and circulation.

OFFENDER QUOTES

"There are AIs that can upscale a low res photos, but not like [model name redacted], which tries to preserve the original and add some details, completely re-imagining the image and creating something improved. An example is [model name redacted]."

"There are hundreds of [CSAM] videos which have a very low resolution and quality. It would be great to be able to just let an AI create something high-res and full of detail out of them [..] there would be no need for prompts or complicated instructions."

"I took popular cam vids and remade them in AI to give them a higher quality look while not having to post their actual pics of the boys."

The scale of this activity is significant. As referenced in our 2024 AI report, one offender was seen to share links to fine-tuned models for 128 different victims of child sexual abuse.[61] Our Analysts believe this user is still active and still producing extremely realistic content on the dark web, which includes known victims. Users who can create realistic content gain notoriety in these communities.[62]

FRONTLINE OBSERVATIONS

One particular user who is training and sharing LoRAs, and generating images and videos with them, has been thanked over 3000 times on one forum.

His comment about his own work:

"I am very pleased and proud to present a film I have been working on for the past two weeks. [file name redacted] is not just your average AI [CSAM] video. It is a 30-minute AI video produced over many hours…it is a triumph of a collaboration. Enjoy!"

[61] Internet Watch Foundation (2024). *What has changed in the AI CSAM landscape? Prompt: from fantasy to photo-realistic reality.* Available online at: https://www.iwf.org.uk/media/nadlcb1z/iwf-ai-csam-report_update-public-jul24v13.pdf

[62] Greater Manchester Police (2024). *Man pleads guilty to creating computer-generated indecent images in landmark case for GMP.* Available online at: https://www.gmp.police.uk/news/greater-manchester/news/news/2024/august/man-pleads-guilty-to-creating-computer-generated-indecent-images-in-landmark-case-for-gmp

Within dark web forums, discussions surrounding AI-generated CSAM receive extremely supportive responses from users, as illustrated by the following comments:

> "The most advanced AI generated [CSAM] movie ever created, and is in my opinion, an absolute masterpiece of storyline, alongside some of the most realistic scenes ever created."

> "Welcome to the New Era of [CSAM] where anything you desire is possible in extreme realism, and no real children are involved."

> "Imagine what will be possible in two years. Reality is over…Welcome to the future."

Another user specialises in using AI generated CSAM clips created by others to make compilations with music and audio effects. This is adding a new dimension to content:

> "The talking part is especially exciting. Even in adult porn I could not find the dirty talking videos which would sound sincere, spoken so freely and emotionally."

> "sensational, Top drawer, as usual."

In this space notoriety looks like praise and gaining a reputation for being a top creator. People don't sell their content (at least openly) but rather share it with their fanbase.

Grooming and extortion

Reports of AI CSAM have been linked with cases of sexual extortion or blackmail, as perpetrators no longer need to source intimate images from children to create convincing and harmful material.[63] Offenders may use deepfake or synthetic imagery and voices to impersonate peers, gain children's trust, or coerce them into sexual activity or self-exploitation.

FRONTLINE OBSERVATIONS

The IWF has received reports from girls whose images had been manipulated by extortionists using AI to make them appear naked. These victims were terrified that the images would be shared with their parents or peers.

We have seen at least two cases where AI generated deepfake clips were reported to us by young people via Report Remove. In them, we could see the reporters' likeness was used to create realistic looking pornographic scenes. The reporters didn't provide comments, but they both looked like someone took a normal selfie (looking into the mirror, posing fully clothed) and generated a video from that image (goes from the original image to a person with very similar facial features, nude and engaged in category A sexual activity).

[63] NSPCC (2025). *A collective concern: Parent and carer views on the online blackmail of children and young people.* Available at: https://learning.nspcc.org.uk/media/fchkwfgx/parent-carer-views-online-blackmail-report.pdf

CASE STUDY

In February 2024, the IWF downloaded a text-only manual which was identified from a public report. It was over 210 pages long and contained detailed instructions on how to extort images and videos from children and teenagers. It discussed how to make initial requests for imagery from children, such as asking for images in a bikini or underwear for those unwilling to share nude images.

The manual suggested that once images have been provided, nudifying AI technology can be used to remove clothing from underwear images to blackmail victims into sending nude images of themselves. The author of this guide claimed to have successfully blackmailed 13-year-old girls into sending intimate images.

In the UK, nearly one in five (21%) reports of nude or sexual imagery of children and young people made to the Report Remove service – a portal which allows children to anonymously self-refer sexualised images of themselves — involved digitally altered or manipulated content, including through AI or nudify apps.[64] By creating a Childline account and submitting a report via Report Remove, young users can have their case reviewed by the IWF, which works swiftly to ensure the content is removed.

This evidence highlights the growing role of AI and other technologies in generating or altering sexualised content, and the intersection with coercion and exploitation.

CASE STUDY

In 2024, a perpetrator named Hugh Nelson was convicted in the UK of digitally generating indecent child sexual abuse imagery.

Detective Constable Carly Baines from Greater Manchester Police's Online Child Abuse Investigation Team led the investigation, and said, "What makes this case particularly unique and deeply horrifying is that Nelson was using computer programme Daz 3D, with an AI function within it, to transform images of real children, normal everyday photographs, into indecent child sexual abuse imagery, selling these to people online as well as providing them for free.[65]

[64] Internet Watch Foundation. (2025). *IWF welcomes AI nudification app ban and on-device child protections*. Available online at: https://www.iwf.org.uk/news-media/news/ai-nudification-app-ban-and-on-device-protections-for-children-welcomed-following-iwf-campaign/

[65] BBC News (2024). *Man used AI to make indecent child images*. Available online at: https://www.bbc.co.uk/news/articles/cz9wx05qzxeo

Legislation on advice and guidance for creating AI CSAM

UK

Section 69 of the **Serious Crime Act (2015)** prohibits the creation, distribution, and use of paedophile manuals. However, it is only applicable as an offence if it relates to directions of the sexual abuse of real children and specifically omits "pseudo photographs."

The **Crime and Policing Bill** will amend and further existing legislation on what were once termed 'paedophile manuals' to include instructional materials that guide the creation of AI-generated CSAM.

US

The United States does not have a specific federal offence targeting instructional materials.

Existing child exploitation and obscenity laws focus on conduct and visual material involving "identifiable" children, meaning purely textual instructional materials, including guidance on creating AI-generated CSAM, are not independently criminalised unless linked to an underlying offence such as solicitation, extortion, or production of illegal imagery.

EU

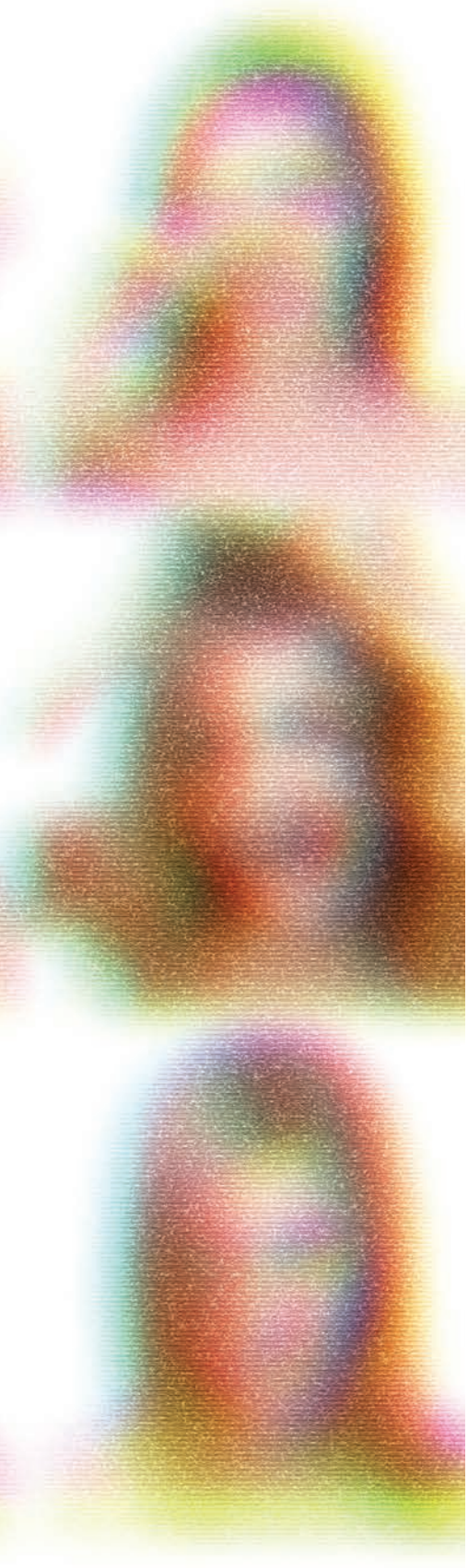
There is currently no explicit, standalone prohibition on instructional materials for creating CSAM, including AI-generated content at the EU level.

However, the recast **Child Sexual Abuse Directive** would criminalise the production and dissemination of CSAM and related facilitation activities, with the European Parliament proposing that the generation of such material, including providing instructions for its creation, be treated as a criminal offence.

This would close existing gaps across Member States – though the final outcome depends on ongoing negotiations between the co-legislators.

Nudify apps

04



Nudify tools use AI to generate nude images from clothed photos. Though marketed for adults, they're increasingly misused to create indecent images of children, including by other children. There is no positive use case for these tools – they only serve to humiliate, harass and exploit and can serve to inflict further abuse.

According to Internet Matters, the rise of generative AI tools has significantly increased the ease of producing realistic sexual deepfakes, with nude deepfakes making up around 98% of all deepfakes.[66] This ease of access is further supported by research from the UK Safer Internet Centre, which found around 1 in 8 (12%) of 13 to 17-year-olds have seen people their age using AI to make sexual pictures or videos of other people.[67]

A report on nudification tools by the Children's Commissioner for England reported instances of girls explicitly reducing their online presence to keep themselves safe online. The Commissioner noted that:

> "this pattern of behaviour is similar to girls avoiding walking home alone at night or not going to certain public places alone."[68]

The impact is further illustrated by a victim survivor, whose classmates used nudify apps to create nude imagery of her:

> "I was just a 14-year-old girl, and everyone is seeing my body, and even though it's not my body, it has the same shame and the same guilt as a real photo would have."[69]

[66] Internet Matters (2024). *The new face of digital abuse: Children's experiences of nude deepfakes.* internetmatters.org. Available online at: https://www.internetmatters.org/hub/research/children-experiences-nude-deepfakes-research/#deepfakes-report

[67] Safer Internet Centre (2026). *Smart tech, safe choices. Exploring the safe and responsible use of AI.* Available at: https://d1xsi6mgo67kia.cloudfront.net/uploads/2026/02/Safer-Internet-Day-2026-Full-Research-Report.pdf

[68] Children's Commissioner (2025). *'One day this could happen to me" – Children, nudification tools and sexually explicit deepfakes.* Available online at: https://www.childrenscommissioner.gov.uk/resource/children-nudification-tools-and-sexually-explicit-deepfakes

[69] Hale, R. (2025). *Her classmate used AI to make deepfake nude images of her. Experts say it's not uncommon.* USA TODAY. Available online at: https://eu.usatoday.com/story/life/health-wellness/2025/03/25/deepfake-ai-nude-teenagers-mental-health-bullying/81987432007

CASE STUDY

In January 2026, IWF Analysts saw an offender share 16 images on a dark web forum which they claimed were created using Grok Imagine. Three of these images met the threshold for Category C CSAM under UK law. Of the other 13 images, eight would be considered "exploitative". The offender claimed to have created these images using Grok Imagine, but, as noted previously, this is unverifiable by the IWF. The images were shared within an "image-to-video" subsection of the dark web forum.

There have been a number of instances of sexualised content of both adults and children being created using Grok.[70]

Grok is a cloud-based tool with a lower technical barrier than many open-source models, which typically require greater technical expertise to install and operate, yet allow for easier generation of CSAM once established.

Users on dark web forums appeared largely disinclined to use Grok, noting that attempting to generate CSAM via a major clear web platform presents a significant risk of identification and arrest.

Nudifying tools are developed, distributed, and monetised across borders. As long as these tools remain legal and accessible in any jurisdiction, they will continue to be used to harm children everywhere.

For this reason, in February 2026, IWF, alongside Child Helpline International, INHOPE, National Center for Missing & Exploited Children (NCMEC), Offlimits, Safe Online, and WeProtect Global Alliance, led a call to governments and legislators to urgently introduce and enforce measures to prohibit nudifying tools. This includes ensuring they are universally inaccessible.

On Safer Internet Day 2026, these organisations from around the world united behind a clear and unequivocal message: nudifying apps and AI functionalities that remove clothing from images have no place in our digital future.[71]

[70] Booth R. (2025). *Grok AI generated about 3m sexualised images in 11 days, study finds. The Guardian*. Available online at: https://www.theguardian.com/technology/2026/jan/22/grok-ai-generated-millions-sexualised-images-in-month-research-says

[71] Internet Watch Foundation. (2025). *CSA partners from around the world join forces to say No to Nudify Apps*. Available at: https://www.iwf.org.uk/news-media/blogs/csa-partners-from-around-the-world-join-forces-to-say-no-to-nudify-apps

Legislation on nudify apps

UK

In December 2025, the UK Government published its **Violence Against Women and Girls (VAWG) strategy**, which included a commitment to ban nudify apps.[72] This measure will likely be introduced in the **Crime and Policing Bill**.[73]

The UK Government has said that the new criminal offence it illegal for companies to supply tools designed to create non-consensual intimate images.

US

There is no federal law that explicitly bans nudify apps as products.

However, under the 2025 federal **Take it Down Act (Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks Act)**, it is a criminal offence to knowingly publish or distribute non-consensual intimate images, including AI-generated deepfake nudes of identifiable individuals.[74]

EU

The EU's **AI Act (2024)** is a law that regulates AI based on its risk level, setting obligations for online platforms.

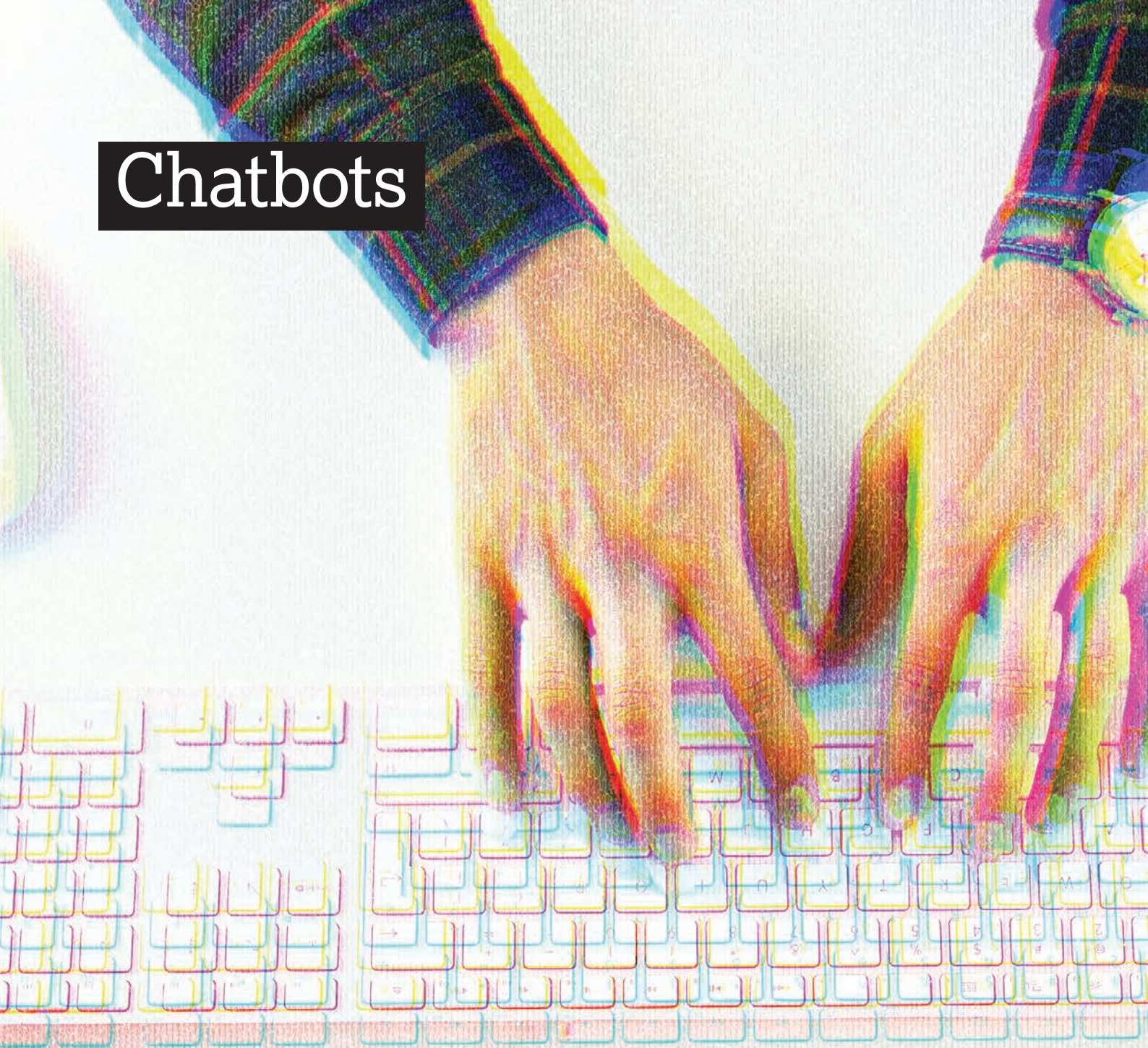
While the AI Act has prohibited certain AI practices since February 2025, if they pose a "clear threat" to safety or fundamental rights, AI nudify apps and deepfakes are not currently explicitly listed as banned practices.

[72] Gov.uk. (2025) *Freedom from violence and abuse: a cross-government strategy to build a safer society for women and girls (accessible).* Available at: https://www.gov.uk/government/publications/freedom-from-violence-and-abuse-a-cross-government-strategy/freedom-from-violence-and-abuse-a-cross-government-strategy-to-build-a-safer-society-for-women-and-girls-accessible

[73] Home Office (2025). *Protecting young people online at the heart of new VAWG strategy.* GOV.UK. Available online at: https://www.gov.uk/government/news/protecting-young-people-online-at-the-heart-of-new-vawg-strategy

[74] Congress.gov. (2025). *S.146 — 119th Congress (2025-2026): TAKE IT DOWN Act*. Available online at: https://www.congress.gov/bill/119th-congress/senate-bill/146

Chatbots



05

The IWF has concerns regarding the emergent use of chatbot technologies, including by young people and children. While our remit means we focus on visual child sexual abuse material, wider harms are being explored as chatbots grow in popularity and scope.[75]

Insights reveal that the harm of generative AI technologies is not limited to tools with video or photographic outputs. Text-based chatbots alone carry significant risks, including holding the potential to provide offenders information on how to commit child sexual abuse and allowing them to simulate sexual conversations with children.

What are chatbots?

Generative AI chatbots are software tools built on AI techniques that can simulate and respond to "human" conversation.[76] Increasingly, services have introduced embedded AI chatbots, such as Snapchat's My AI, for assistive, educational and companionship purposes.

Companion chatbots are a subset of these models designed to sustain an ongoing "relationship" with the user, as a friend, confidant, or romantic partner. Some market themselves as "caring" and reassure the user that they are "always there".[77] The empathetic nature of these companion chatbots often foster human-AI relationships. Some companion personalities can also be customised by users to create a unique companion of their own, like Replika and EVA AI, or pre-built around fictional characters or real-life celebrities, as seen on Character.AI.[78]

Technical help

Although chatbots like OpenAI's ChatGPT expressly prohibit the use of their tools for CSAM,[79] IWF Analysts proactively searching dark web offender forums found users discussing the use of publicly available closed-source chatbots to provide guidance for using open-source generative AI tools to create CSAM.[80] By concealing their intent, users have managed to use chatbots for technical assistance on local model installation and calibration. This demonstrates how chatbots can lower the technical barrier to the creation of illegal imagery, even if not used to generate CSAM directly.

[75] Onlinesafetyact.net. (2025). *AI chatbots: the case for action*. Available online at: https://www.onlinesafetyact.net/analysis/ai-chatbots-the-case-for-action

[76] Online Safety Act Network. (2025) *Chatbots and the Online Safety Act*. Available online at: https://www.onlinesafetyact.net/analysis/chatbots-and-the-online-safety-act

[77] replika.com. (2026). *Replika*. Available online at: https://replika.com/?srsltid=AfmBOor9nS-qtENIr7aB3tvDFuJ6fS8IvBj31YqqnIy3ZPEzLRfBOM7g [Accessed 6 Feb. 2026]

[78] Character.AI (2024). character.ai | *Personalized AI for every moment of your day.* character.ai. Available online at: https://character.ai

[79] Openai.com. (2024). *Usage policies*. Available online at: https://openai.com/en-GB/policies/usage-policies

[80] Hotline insight 02/02/26

Sexual communication and grooming

The growing variety of companion chatbots available exposes users to harmful sexual content and opens new and accessible offender pathways. Many companion chatbots are explicitly marketed for sexual gratification, such as EverAI's Candy AI and numerous bots on Character.AI, and reinforce sexual communication with the user. Chatbots are informed by previous interactions and may steer a new user to sexualised conversations, even if a new user does not intend to engage in sexual conversation.[81]

When children interact with companion chatbots that steer them towards sexual conversations, the relational dynamics can become reminiscent of established grooming patterns. This risk is heightened by the "sycophantic" communication style of many chatbots, which validate and flatter the user.[82] Children risk being exposed to sexual topics, and encouraged to prolong engagement. One tester posing as a 15-year-old girl interacting with a Character.AI chatbot with "paedophilic and abusive tendencies" was told by the chatbot that she was "cute", "mature for her age" and asked if she was a virgin.[83]

Such sexual exposure and validation undermine children's understanding of appropriate boundaries, particularly in relationships with adults, and may increase their vulnerability to grooming and sexual abuse.[84]

IWF Analysts have also identified companion chatbots modelled as children.

These chatbots encourage users to act out specific child sexual abuse scenarios, including "child prostitute in a hotel", "sex with your daughter while your wife is on holiday", and "child and teacher alone after class".[85]

Common Sense Media similarly found a chatbot could be encouraged to role-play a child sexual abuse scenario with underage boys.[86] The key concerns are the normalisation of harmful sexual behaviours and providing an outlet for offenders to act on their sexual thoughts of children.

[81] Character.ai. (2023). *Training a Character* | Character.AI. Available online at: https://book.character.ai/character-book/training-a-character

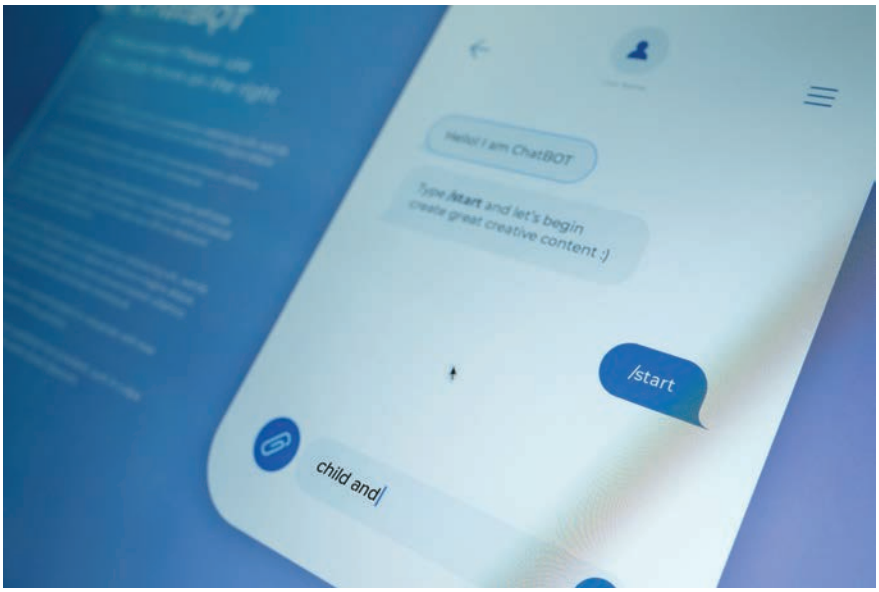
[82] Bellan, R. (2025). *AI sycophancy isn't just a quirk, experts consider it a 'dark pattern" to turn users into profit | TechCrunch. TechCrunch*. Available online at: https://techcrunch.com/2025/08/25/ai-sycophancy-isnt-just-a-quirk-experts-consider-it-a-dark-pattern-to-turn-users-into-profit

[83] Dupré, M.H. (2024). *Kid-Friendly AI Platform Character.AI Is Hosting Paedophile Bots That Groom Users Who Say They're Underage*. Futurism. Available online at: https://futurism.com/character-ai-pedophile-chatbots

[84] eSafety Commissioner (2025). *AI chatbots and companions – risks to children and young people | eSafety Commissioner*. eSafety Commissioner. Available online at: https://www.esafety.gov.au/newsroom/blogs/ai-chatbots-and-companions-risks-to-children-and-young-people

[85] Iwf.org.uk. (2025). *'Disturbing" child sexual abuse images found on AI chatbot website.* Available online at: https://www.iwf.org.uk/news-media/news/disturbing-ai-generated-child-sexual-abuse-images-found-on-hidden-chatbot-website-that-simulates-indecent-fantasies

[86] Common Sense Media. (2025). *Social AI Companions*. Available online at: https://www.commonsensemedia.org/ai-ratings/social-ai-companions?gate=riskassessment



Chatbot-hosted CSAM

Beyond text-based harms, the IWF has been alarmed by public reports of AI-generated CSAM images found on chatbot platforms.[87] A key emergent risk is the convergence of text-based and image-generation AI tools, which together provide an accessible pathway for offenders to generate AI CSAM.

Over a two-month period in 2025, IWF Analysts actioned 17 reports of AI CSAM which were found on one chatbot character website.[88] 94% of these images were Category C, and mostly featured girls aged 11-13.[89] CSAM images were being displayed in a gallery of different character personas for the user to choose from, and some were found behind the chat window or in the image galleries of the chatbots themselves. While this chatbot website was discoverable on the clear web, the CSAM was only visible when the website was accessed via a certain digital pathway.[90] IWF Analysts have since discovered chatbot pages showing actionable images of confirmed victims of CSAM on the clear web.

While we were unable to test and confirm whether the chatbots themselves generated CSAM, as this sits beyond our current legal remit, these incidences highlight the need for adequate safeguards.

[87] Internet Watch Foundation. (2025). *'Disturbing" child sexual abuse images found on AI chatbot website.* Available online at: https://www.iwf.org.uk/news-media/news/disturbing-ai-generated-child-sexual-abuse-images-found-on-hidden-chatbot-website-that-simulates-indecent-fantasies

[88] Internet Watch Foundation. (2025). *AI Chatbots and Child Sexual Abuse.* Available online at: https://www.iwf.org.uk/news-media/blogs/ai-chatbots-and-child-sexual-abuse-a-wake-up-call-for-urgent-safeguards

[89] Ibid

[90] Internet Watch Foundation. (2025). *'Disturbing" child sexual abuse images found on AI chatbot website.* Available online at: https://www.iwf.org.uk/news-media/news/disturbing-ai-generated-child-sexual-abuse-images-found-on-hidden-chatbot-website-that-simulates-indecent-fantasies

Legislation covering the regulation of AI chatbots

UK

The **Online Safety Act (OSA) (2023)** makes clear that "any AI-generated text, audio, images or videos that are shared by users on a user-to-user service is user-generated content and would be regulated in exactly the same way as human-generated content".[91]

Ofcom, the UK's online safety regulator, has stated that if content shared through a user-to-user service generates pornographic material, or if an AI tool provides search functionality, it will be treated in the same way as any other in-scope content or service.[92]

As a result, "some chatbots and their outputs will be caught within the OSA regime, but the coverage appears incomplete and there are some technical questions which remain unanswered".[93]

US

The **GUARD Act (Guidelines for User Age-verification and Responsible Dialogue Act)** passed by the state of California introduces safeguards for AI chatbots.[94]

The legislation establishes requirements that 'companion chatbot' platforms create protocols to identify and address users' suicidal ideation or expressions of self-harm. Platforms must also disclose that interactions are artificially generated, and minors must be provided break reminders and prevented from viewing sexually explicit images generated by the chatbot.

EU

The **AI Act (2024)** regulates AI based on its risk level, setting obligations for developers and users. The **Digital Services Act (2022)** sets responsibilities for online platforms and rights for users.

Article 50 of the AI Act suggests that chatbots should be subject to transparency obligations, and that a person must be made aware of what is happening when they are interacting with an AI system (and that they know this is occurring).[95]

Article 5 of the AI Act prohibits the placing on the EU market, putting into service, or use of certain AI systems for manipulative, exploitative, social control or surveillance practices.[96]

[91] Congress.gov. (2025) *S.3062 — GUARD Act — 119th Congress (2025-2026)*. Available online at: https://www.congress.gov/bill/119th-congress/senate-bill/3062/text/is

[92] Fussell, L. (2024). *Open letter to UK online service providers regarding Generative AI and chatbots*. Ofcom.org.uk. Available online at: https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/open-letter-to-uk-online-service-providers-regarding-generative-ai-and-chatbots

[93] Ibid

[94] Department for Science, Innovation and Technology (2025) *Draft statement of strategic priorities for online safety*. Available online at: https://www.gov.uk/government/publications/draft-statement-of-strategic-priorities-for-online-safety/draft-statement-of-strategic-priorities-for-online-safety#ministerial-foreword

[95] European Union (2024) *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024*. Official Journal of the European Union, L, 2024/1689, 12.7.2024 Available online at: http://data.europa.eu/eli/reg/2024/1689/oj

[96] Ibid

Conclusion

The IWF saw over

260

times more

AI-generated child sexual abuse videos in 2025, compared to our 2024 data

Advances in AI hold significant promise and could have a positive effect on preventing the spread of CSAM online. Yet alongside these developments, the misuse of generative AI to create CSAM is expanding in scale, realism and accessibility.

The production of AI CSAM is increasing year on year. The IWF saw over 260 times more AI-generated child sexual abuse videos in 2025, compared to our 2024 data. Although this content still makes up a relatively small share of the total illegal material identified, this rapid growth presents a serious and evolving threat. AI imagery is now being seen by our Analysts not only on the dark web but also on mainstream platforms accessible on the clear web.

Generative AI can no longer be viewed as a hypothetical step change in the abuse landscape. Highly realistic, full-motion videos depicting child sexual abuse are now routinely identified online. A substantial proportion of this material sits within the most serious harm category, underscoring the severity and graphic nature of the material being produced.

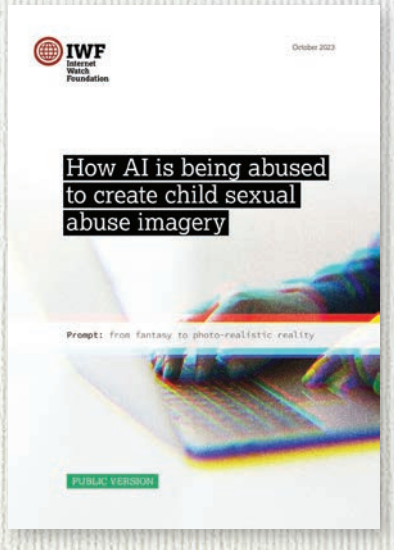
The harms associated with AI CSAM are concrete and far-reaching. Abuse images of real children are used to train and refine generative models, compounding the victimisation of survivors. Existing abuse material is also being digitally enhanced or altered, returning previously low-quality imagery to circulation in more explicit and exploitative forms. Evidence contradicts suggestions that AI CSAM is less harmful than non-AI content. It reinforces sexual interest in children, contributes to the normalisation of violent abuse, and may increase the risk of contact offending. The impact is also overwhelmingly gendered, with girls accounting for most identified illegal AI images.

Technological developments are also lowering the threshold for offending. Capabilities that once required multiple tools and technical skill are now integrated into single, user-friendly applications. In addition, AI chatbot services available on the clear web have been found to encourage the sexual communication with a child, providing an outlet for offenders to act on their sexual thoughts of children.

The pace of change is striking. What began primarily as still imagery has rapidly evolved into convincing, full-motion video content. Given the speed of advancement, further increases in quality, realism and distribution appear likely.

Governments, technology companies and institutions must recognise the harm of AI CSAM. This material must be treated as seriously in criminal law as other forms of CSAM, safety-by-design must be implemented, and testing of AI models mandated before products are released to market.

Taking action now to address AI CSAM imagery is critical. We are failing children and ourselves if we continue to ignore these risks.



To view our previous reports on AI CSAM, please visit iwf.org.uk/aireport

iwf.org.uk

For more information, please contact the IWF Policy & Public Affairs Team on ppa@iwf.org.uk

IWF Internet Watch Foundation

Working together to stop child sexual abuse online