IWF
Internet
Watch
Foundation

October 2023

# How AI is being abused to create child sexual abuse imagery

**Prompt:** from fantasy to photo-realistic reality

PUBLIC VERSION

Full Report:
https://www.iwf.org.uk/media/q4zll2ya/iwf-ai-csam-report_public-oct23v1.pdf

# Executive summary

Child sexual abuse images generated using artificial intelligence is a new and growing area of concern.

**The key findings of this report are as follows:**

In total, **20,254 AI-generated images were found** to have been posted to one dark web CSAM forum **in a one-month period.**

Of these, **11,108 images were selected for assessment by IWF analysts.** These were the images that were judged most likely to be criminal.

(The remaining 9,146 AI-generated images either did not contain children or contained children but were clearly non-criminal in nature.)

**12 IWF analysts dedicated a combined total of 87.5 hours to assessing these 11,108 AI-generated images.**

Any images assessed as criminal were criminal under one of two UK laws, as described in section 5. These are:

- The Protection of Children Act 1978 (as amended by the Criminal Justice and Public Order Act 1994). This law criminalises the taking, distribution and possession of an "indecent photograph or pseudo-photograph of a child".
- The Coroners and Justice Act 2009. This law criminalises the possession of "a prohibited image of a child". These are non-photographic – generally cartoons, drawings, animations or similar.

**2,562 images were assessed as criminal pseudo-photographs, and 416 assessed as criminal prohibited images.**

**Other findings:**

1. AI-generated content currently comprises a small proportion of normal IWF activities, though one of its defining features is its potential for rapid growth.
2. Perpetrators can legally download everything they need to generate these images, then can produce as many images as they want – offline, with no opportunity for detection. Various tools exist for improving and editing generated images until they look exactly like the perpetrator wants.
3. Most AI CSAM found is now realistic enough to be treated as 'real' CSAM. The most convincing AI CSAM is visually indistinguishable from real CSAM, even for trained IWF analysts. Text-to-image technology will only get better and pose more challenges for the IWF and law enforcement agencies.
4. There is now reasonable evidence that AI CSAM has increased the potential for the re-victimisation of known child sexual abuse victims, as well as for the victimisation of famous children and children known to perpetrators. The IWF has found many examples of AI-generated images featuring known victims and famous children.
5. AI CSAM offers another route for perpetrators to profit from child sexual abuse. The first examples of this new commerciality have been identified by the IWF.
6. Creating and distributing guides to the generation of AI CSAM is not currently an offence, but could be made one. The legal status of AI CSAM models (files used for generating images) is a more complicated question.

# Introduction to this report

This year, the Internet Watch Foundation (IWF) has been investigating its first reports of child sexual abuse material (CSAM) generated by artificial intelligence (AI).

Initial investigations uncovered a world of text-to-image technology.

In short, you type in what you want to see; the software generates the image.

The technology is fast and accurate – images usually fit the text description very well. Many images can be generated at once – you are only really limited by the speed of your computer. You can then pick out your favourites; edit them; direct the technology to output exactly what you want.

**These images can be so convincing that they are indistinguishable from real images.**

The most convincing AI CSAM images, then, can be called photorealistic. For IWF analysts, looking at this sort of AI CSAM is exactly like looking at 'real' images of the sexual abuse of children. Except these images have been generated by algorithms.

Images show the rape of babies and toddlers; famous pre-teen children being sexually abused; BDSM (bondage and discipline, dominance and submission, and sadomasochism) content featuring tweens and teenagers. And more.

Effectively articulating the criminality of AI CSAM can be a challenge – there are groups who seek to lessen the severity of these images: they 'don't have real children', or 'don't hurt anyone'.

**UK law, however, is clear: AI CSAM is criminal.**

**Images that are not realistic –** that appear like cartoons or drawings – are "actionable" by our analysts (criminal, and therefore able to be removed from the internet under UK law) under laws on prohibited (non-photographic) images of children.

**Images that are realistic –** that appear to be photographs – are actionable under laws on indecent pseudo-photographs of children. (For precise laws, see section 5).

Amid all the focus on realism, photorealism, and hyperrealism, and complex debates about legality – simply stated – this technology allows perpetrators to generate dozens, even hundreds of child sexual abuse images at the click of a button.

Crucially, you can download AI technology (at just a couple of gigabytes) and run it on your device offline. So, once you have the technology, you can generate as many child sexual abuse images as you like – 'in the dark', with little or no risk of detection.

The genie is out of the bottle. Offline child sexual abuse image generation is our reality.

## What does this mean for IWF?

Currently, AI CSAM represents a small portion of the vast numbers of 'real' CSAM we find. (Over 255,000 webpages last year, representing hundreds of thousands or even millions of images.) Time will tell whether this trickle becomes a flood.

Some websites have been set up that are dedicated to sharing AI-generated images, but we are also starting to see AI-generated images mixed in with 'real' images. These images can be especially difficult for analysts to detect as AI-generated – to tell 'real' from 'fake'.

As the technology continues to improve, and perpetrators generally get better at generating realistic images, this challenge will only get harder.

These websites are still reported, and removal is pursued. UK law is clear, but if websites lie in other jurisdictions, removal can be more complicated.

IWF tags all these images to identify them as AI-generated, which helps law enforcement and victim identification (VID) efforts.

Questions remain. How can safeguards be built into this technology, even if offline image generation is possible? Is AI image detection possible and practicable? Is the law fit for purpose, or should it be changed? Will mass quantities of AI-generated images enfeeble hash lists?

Lots of discussion about the risks of AI – discussion that spurs moves to regulate AI companies – centres around hypothetical or long-term risks like creation of synthetic viruses, cyberattacks or, at the extreme, the risks in creating a 'superintelligence', or postulated artificial general intelligence (AGI).

**AI CSAM is different because it is happening now. Images are being shared online now. It is a current problem that requires action.**

At the same time, solutions developed and implemented now have the potential to mitigate this problem.

With all technological advance comes benefits as well as risks. Though this report focuses on current abuse of AI technology to generate CSAM, it is important to bear in mind the widespread potential for benefits from AI across society, from applications in science, research, and healthcare, to applications in the creative and entertainment industries.

Nonetheless, left unchecked, this technology will cause harm to children.

It harms known victims of child sexual abuse, whose likenesses are being used to generate more images of them in new scenarios.

It harms new victims of child sexual abuse, whose potential investigators might spend time and resources pursuing the rescue of children who turn out to be virtual characters.

These images provide new possibilities for perpetrators to use to groom and coerce children. They even allow the most technically proficient perpetrators to make money from abuse.

And this is the worst in terms of quality of output that AI technology will ever be. It only has the potential to get better: to produce more lifelike images; to better enable the grooming and abuse of children.

Overall, AI CSAM poses a significant risk to IWF's mission to remove child sexual abuse material from the internet.

## What is the IWF and why has it produced this report?

The IWF is a not-for-profit organisation, funded by tech companies, government, global funders and the public, whose remit is to remove CSAM from the internet.

The IWF Hotline, which finds, assesses, and seeks removal of this criminal content, has two main sources for its work: reports from the public (and external partners), and proactive searching for content.

This year, the Hotline has received its first reports of AI CSAM, mostly from members of the public. Reporting numbers were – and remain – small relative to the number of other CSAM reports.

Nonetheless, subsequent proactive searches for AI CSAM found widespread evidence for a large and growing problem. Images and intelligence obtained from these proactive searches have informed IWF media pieces that have raised awareness of this problem and the enormous potential for abuse. Consultations with government and civil society about how to address this problem are ongoing, and discussions with industry in the early stages.