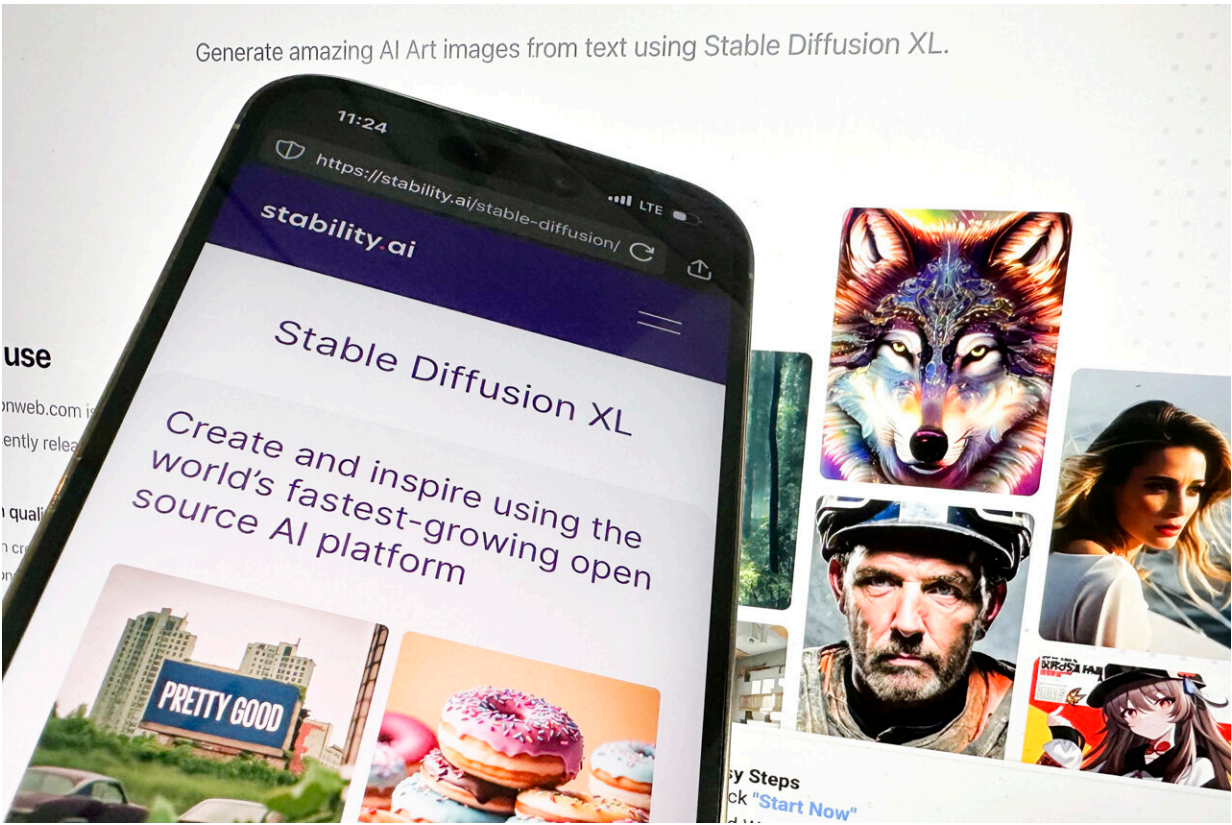


# AI-generated child sexual abuse images could flood the internet. A watchdog is calling for action

October 25 2023, by MATT O'BRIEN and HALELUYA HADERO



In this picture, the desktop and mobile websites for Stable Diffusion are seen, Tuesday, Oct. 24, 2023, in New York. Computer-generated child sexual abuse images made with artificial intelligence tools like Stable Diffusion are starting to proliferate on the internet and are so realistic that they can be indistinguishable from photographs depicting actual children, according to a new report. Credit: AP Photo/John Minchillo

The already-alarming proliferation of child sexual abuse images on the internet could become much worse if something is not done to put controls on artificial intelligence tools that generate deepfake photos, a watchdog agency warned on Tuesday.

In a written report, the U.K.-based Internet Watch Foundation urges governments and technology providers to act quickly before a flood of AI-generated images of child sexual abuse overwhelms law enforcement investigators and vastly expands the pool of potential victims.

"We're not talking about the harm it might do," said Dan Sexton, the watchdog group's chief technology officer. "This is happening right now and it needs to be addressed right now."

In a first-of-its-kind case in South Korea, a man was sentenced in September to 2 1/2 years in prison for using artificial intelligence to create 360 virtual child abuse images, according to the Busan District Court in the country's southeast.

In some cases, kids are using these tools on each other. At a school in southwestern Spain, police have been investigating teens' alleged use of a phone app to make their fully dressed schoolmates appear nude in photos.

The report exposes a dark side of the race to build generative AI systems that enable users to describe in words what they want to produce—from emails to novel artwork or videos—and have the system spit it out.

If it isn't stopped, the flood of deepfake child sexual abuse images could bog investigators down trying to rescue children who turn out to be virtual characters. Perpetrators could also use the images to groom and

coerce new victims.

Sexton said IWF analysts discovered faces of famous children online as well as a "massive demand for the creation of more images of children who've already been abused, possibly years ago."

"They're taking existing real content and using that to create new content of these victims," he said. "That is just incredibly shocking."

Sexton said his charity organization, which is focused on combating online child sexual abuse, first began fielding reports about abusive AI-generated imagery earlier this year. That led to an investigation into forums on the so-called dark web, a part of the internet hosted within an encrypted network and accessible only through tools that provide anonymity.

What IWF analysts found were abusers sharing tips and marveling about how easy it was to turn their home computers into factories for generating sexually explicit images of children of all ages. Some are also trading and attempting to profit off such images that appear increasingly lifelike.

"What we're starting to see is this explosion of content," Sexton said.

While the IWF's report is meant to flag a growing problem more than offer prescriptions, it urges governments to strengthen laws to make it easier to combat AI-generated abuse. It particularly targets the European Union, where there's a debate over surveillance measures that could automatically scan messaging apps for suspected images of child sexual abuse even if the images are not previously known to law enforcement.

A focus of the group's work is to prevent previous sex abuse victims from being abused again through the redistribution of their photos.

The report says technology providers could do more to make it harder for the products they've built to be used in this way, though it's complicated by the fact that some of the tools are hard to put back in the bottle.

A crop of new AI image-generators was introduced last year and wowed the public with their ability to conjure up whimsical or photorealistic images on command. But most of them are not favored by producers of child sex abuse material because they contain mechanisms to block it.

Technology providers that have closed AI models, with full control over how they're trained and used—for instance, OpenAI's image-generator DALL-E—have been more successful at blocking misuse, Sexton said.

By contrast, a tool favored by producers of child sex abuse imagery is the open-source Stable Diffusion, developed by London-based startup Stability AI. When Stable Diffusion burst onto the scene in the summer of 2022, a subset of users quickly learned how to use it to generate nudity and pornography. While most of that material depicted adults, it was often nonconsensual, such as when it was used to create celebrity-inspired nude pictures.

Stability later rolled out new filters that block unsafe and inappropriate content, and a license to use Stability's software comes with a ban on illegal uses.

In a statement released Tuesday, the company said it "strictly prohibits any misuse for illegal or immoral purposes" across its platforms. "We strongly support law enforcement efforts against those who misuse our products for illegal or nefarious purposes," the statement reads.

Users can still access older versions of Stable Diffusion, however, which are "overwhelmingly the software of choice ... for people creating

explicit content involving children," said David Thiel, chief technologist of the Stanford Internet Observatory, another watchdog group studying the problem.

The IWF report acknowledges the difficulty of trying to criminalize AI image-generating tools themselves, even those "fine-tuned" to produce abusive material.

"You can't regulate what people are doing on their computers, in their bedrooms. It's not possible," Sexton added. "So how do you get to the point where they can't use openly available software to create harmful content like this?"

Most AI-generated child sexual abuse images would be considered illegal under existing laws in the U.S., U.K. and elsewhere, but it remains to be seen whether law enforcement has the tools to combat them.

A British police official said the report shows the impact already witnessed by officers working to identify victims.

"We are seeing children groomed, we are seeing perpetrators make their own imagery to their own specifications, we are seeing the production of AI imagery for commercial gain—all of which normalizes the rape and abuse of real children," said a statement from Ian Critchley, child protection lead for the National Police Chiefs' Council.

The IWF's report is timed ahead of a global AI safety gathering next week hosted by the British government that will include high-profile attendees including U.S. Vice President Kamala Harris and tech leaders.

"While this report paints a bleak picture, I am optimistic," IWF CEO Susie Hargreaves said in a prepared written statement. She said it is important to communicate the realities of the problem to "a wide

audience because we need to have discussions about the darker side of this amazing technology."

© 2023 The Associated Press. All rights reserved. This material may not be published, broadcast, rewritten or redistributed without permission.

Citation: AI-generated child sexual abuse images could flood the internet. A watchdog is calling for action (2023, October 25) retrieved 29 April 2024 from

<https://techxplore.com/news/2023-10-ai-generated-child-sexual-abuse-images.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.