Internet Watch Foundation Policy Briefing

King's Speech Debate 15 November 2023: Reducing violent crime and raising confidence in policing and the criminal justice system

Key asks:

1. To highlight the need for paedophile manuals and the exchanging of hints and tips related to the creation of AI imagery to be made illegal through an amendment to the Government's Criminal Justice Bill.
2. To call on the Government to address the lack of regulatory oversight of technologies and data sets before they are made open-source or available to markets.
3. To call on the Government to ensure that there is continued parliamentary scrutiny of the Online Safety Act and regime, highlighting the need for Ofcom to work with third parties, such as the Internet Watch Foundation, to make the most of the expertise they possess.
4. To call on the Government to ensure Ofcom makes full use of the powers available to them, including the implementation of provisions regarding End-to-End Encryption.
5. To question the Government on when we can expect more information regarding the Government's Pornography Review and who will be the Chair of the Review.

About the Internet Watch Foundation

The Internet Watch Foundation (IWF) is a not-for-profit organization whose mission is to eradicate child sexual abuse imagery online. We do this by working closely in partnership with the internet industry, law enforcement, and governments globally to detect, remove, and block illegal child sexual abuse material (CSAM) from the internet.

Artificial Intelligence

The IWF published a research report into how artificial intelligence (AI) is increasingly being used to create child sexual abuse imagery online. Depictions of child sexual abuse, including artificial ones, normalise sexual violence against children. We know there is a link between viewing child sexual abuse imagery and going on to commit contact offenses against children.

The key findings ofaylo this report are as follows:

- **11,108 images were selected for assessment by IWF analysts. Of these images, 2,562 images were assessed as criminal pseudo-photographs, and 416 assessed as criminal prohibited images.**

Depictions of child sexual abuse, including artificial ones, normalise sexual violence against children. We know there is a link between viewing child sexual abuse imagery and going on to commit contact offenses against children.

The increase in the creation and proliferation of AI-generated CSAM **poses significant risks to law enforcement's ability to identify children** who need safeguarding. Law Enforcement will be unable to identify whether a victim is 'real' or synthetically generated.

All or almost all AI CSAM found was generated using **Stable Diffusion**. This is because images can be generated locally and there are no prompt restrictions. Characteristics of children are added to the 'positive' prompts, and characteristics of adults are added to the 'negative' prompts to increase the overall impression of an image of a child. With this type of software, there are no limitations or restrictions on prompts that can be used. This is a consequence of Stable Diffusion as an **open-source model**.

IWF analysts have discovered an **online "manual"** written by offenders with the aim of helping other criminals train the AI and refine their prompts to return ever more realistic results. **The 2015 Serious Crime Act does not cover pseudo images of children,** meaning possessing an online manual for AI-generated content is not currently illegal.

Key recommendations:

1. Confusion over legality of images under different laws – Protection of Children Act and Coroners and Justice Act.
2. Differences in how Non-Photographic Imagery and Computer-Generated Imagery are treated internationally (e.g. Japan) and need for ongoing international collaboration.
3. Need for law enforcement training on AI images.
4. Lack of regulatory oversight of technologies and data sets before they are made open-source or available to markets.
5. Engagement with academic and open-source communities to keep technology safe.
6. The need for paedophile manuals and the exchanging of hints and tips related to the creation of AI imagery to be made illegal.

Online pornography regulation review

The IWF also supports the Government's announcement to review the laws around pornography.

We have recently announced that we are working with Aylo, formerly known as Mindgeek, a technology company, which operates several brands, including Pornhub, which offer legal, adult-themed content to a global audience to **explore a potential blueprint for the adult industry in how they can fight the spread of child sexual abuse online**, in a two-year pilot project.

This work builds on the **successful collaboration between the IWF, Lucy Faithful Foundation, and Aylo** in developing a chatbot that deploys on Pornhub and alerts users to when they are using keywords that may return child sexual abuse material and diverts them to where they can get help and support. The IWF welcomes the opportunity of feeding the findings of our work to the Government's review.

End-to-end encryption (E2EE)

We are calling companies to ensure that if they already use, or are about to deploy, end-to-end encryption on their messaging services, they also introduce safeguards to prevent the spread of online child sexual abuse material.

- The IWF has seen a **doubling** of the most extreme forms (Category A) of child sexual abuse in the last two years.
- Children's self-generating content depicting their own abuse now accounts for **three-quarters** of everything we remove from the internet.

The IWF is not against strong encryption, but we believe that companies should be proving through their risk assessments that they have pursued other avenues to protect user privacy before pursuing E2EE.

We believe that they should have measures in place to detect and report CSAM in the same way they do now in non-encrypted channels if E2EE is to be pursued.

The IWF was awarded a grant as part of the government's **Safety Tech Challenge Fund**, whereby we worked with the University of Edinburgh, Cyan Forensics, and Crisp Thinking to develop a plug-in to be integrated within encrypted social platforms. The plug-in is designed to detect CSAM by matching content against **known illegal material** using Hash Matching.

Apple's 2021 *FAQ on Expanded Protections for Children*[1], includes an explanation from Apple of how their CSAM detection in iCloud Photos would operate by not breaking E2EE, preserving the privacy of users' messages, and only using accurate data provided by two child safety organisations. We believe these would have met the tests set down by the ICO's blog on lessons learned from the safety tech challenge fund[2] and demonstrate that it is possible for safety and privacy to co-exist.