Harm Reduction Plan - Child Sexual Abuse Images

Version: 2021/Q1

Description of Harm:

The distribution of images of child sexual abuse is widely recognised as an inherently harmful activity.

These images are different from many other forms of harmful content as the commission of serious criminal acts is integral to their production.

The distribution of these images can also cause serious ongoing harm to the child victims of abuse being portrayed in them.

In response to these concerns, many legal codes around the world have made the distribution of this material a ‘strict liability offence’.

This means that the distribution itself is criminal, with very few technical defences, rather than requiring prosecutors to show that specific harms resulted from the distribution.

We follow a similar approach in respect of this material and our goal is as far as possible to eliminate its distribution on our platform.

We believe that there is very close alignment between the content we prohibit and UK law, so we are dealing almost exclusively with illegal content in this plan.

Indicators:

There is a well-developed understanding of the indicators for content that should be recognised as illegal child sexual abuse images within the global expert community, and we incorporate these expert assessment and classification models into our review systems.

There have been efforts by a number of parties to build a corpus of known abuse images through reporting mechanisms and we are both a contributor to and user of this database.

Our reviewers are trained to recognise new images and add them to the database.

We have provided the regulator with further information about our training processes, and have agreed with them that, as this necessarily involves illegal content, it is not appropriate to make this training material public.

Context:

There is a high level of awareness of the illegal nature of this content in the UK which provides a strong disincentive to people sharing it in public or open settings.

It is more likely to be shared in closed groups and in private communications as people try to keep their sharing secret from others.

We see some instances of people sharing this content in order to condemn it or as part of some kind of vigilante effort but this is not permitted by our rules and is likely against UK law.

We have shared with the regulator details of the number and type of referrals we make to the National Centre for Missing and Exploited Children (NCMEC) where we believe the users were based in the UK to help them understand the extent of this behaviour.

Detection Processes:

User Reports

We offer users the ability to report this material through our reporting flows that are accessed in context across our service. Users are presented with various options including “Nudity”, “Sexual Exploitation” and “Child Abuse” as they make their reports.

Expert Reports

In the UK we work especially closely with the Internet Watch Foundation (IWF) who have a dedicated channel for IWF to report content to us, and we use a fast-track process for reviewing their reports.

Government Reports

We may receive reports from the Child Exploitation and Online Protection command of the National Crime Agency (CEOP) and local UK Police forces about potentially illegal child sexual abuse content and have an expedited process in place for reviewing these.

Automated Detection

We use sophisticated automated tools to check all images and videos that are uploaded to our service against an extensive database of known child exploitation images.

This includes content in both public contexts and private contexts like direct messages.

We also use automated systems to identify and remove a wide variety of images of sexual activity on our service as our rules generally prohibit these (see Sexual Content Plan) and these systems may pick up images of child sexual abuse.

We have shared with the regulator details of how these automated systems work, their accuracy levels, and the protocols we follow to ensure that reports are handled correctly according to the type of content.

Review Processes:

Platform Review

It is important to note that we check all reports for violating content of all types irrespective of the reason selected by the reporter.

This approach is based on our experience that people very commonly select report categories that are not related to the actual problem with the content.

This means that our reviewers will identify and act on child sexual exploitation content if this appears in reports made alleging content for other reasons such as hate speech or bullying.

We recognise that reviewing child sexual abuse content can be very difficult for our review teams, whether employed directly or by our partners, and we have shared with the regulator the protocols we have put in place to support them in this work.

Automated Review

We have a preference for using automated tools to review this type of content where that can be done to a very high confidence level.

This allows us to act more quickly and reduces the exposure of our reviewers to this material.

Where we are dealing with known images then the process of detection and review is effectively unified - if the content matches known illegal content then the person distributing it has broken our rules irrespective of the context or any other factors.

We have shared with the regulator the details of the criteria for when we use this fully automated - detect-review-action - protocol and allowed them to audit its accuracy levels.

External Review

We are subject to a special legal regime in respect of the content covered by this plan under US law which means that we report all instances to an external body, the National Centre for Missing and Exploited Children (NCMEC).

NCMEC reviews the reports they receive from us, and other platforms, and refers these on to law enforcement agencies around the world under a set of cooperation agreements.

As well as this standard reporting, we work with child protection agencies in the UK and may share information about specific instances of distribution where there is a UK locus.

This may happen on a reactive basis, ie in response to requests from UK authorities through the relevant legal channels, or on a proactive basis, where our reviewers believe there is a threat of imminent harm and wish to make UK authorities aware of this.

We have shared with the regulator the protocols we use for data sharing with UK authorities and we have agreed that it is not appropriate to make these public because of the risk that this knowledge might assist offenders.

Actions Taken:

We believe that the serious nature of this harm requires us to take the strongest possible actions when a user distributes any content within this category.

This means that we will remove the content and close a user's account on a first offence.

We may reinstate an account on appeal only if we have made a technical error, but do not accept that there can be any mitigating circumstances for the distribution of this content.

As described in the Review section, content of this type is also routinely reviewed by external agencies who may initiate criminal proceedings against users.

See Also:

This plan deals with content that is defined as child sexual abuse imagery.

There are also issues related to images of children that are not in themselves illegal as they do not show indicators of sexual activity but are nevertheless viewed and shared by people with a sexual interest in children.

A separate section of our harm reduction plan deals with this type of content.

There is also a specific issue related to young people sharing sexual images with each other, a phenomenon commonly known as 'sexting', which is covered in its own plan.