Expanded Protections for Children

Frequently Asked Questions

August 2021

v1.1

Overview

At Apple, our goal is to create technology that empowers people and enriches their lives — while helping them stay safe. We want to protect children from predators who use communication tools to recruit and exploit them, and limit the spread of Child Sexual Abuse Material (CSAM). Since we announced these features, many stakeholders including privacy organizations and child safety organizations have expressed their support of this new solution, and some have reached out with questions. This document serves to address these questions and provide more clarity and transparency in the process.

What are the differences between communication safety in Messages and CSAM detection in iCloud Photos?

These two features are not the same and do not use the same technology.

Communication safety in Messages is designed to give parents and children additional tools to help protect their children from sending and receiving sexually explicit images in the Messages app. It works only on images sent or received in the Messages app for child accounts set up in Family Sharing. It analyzes the images on-device, and so does not change the privacy assurances of Messages. When a child account sends or receives sexually explicit images, the photo will be blurred and the child will be warned, presented with helpful resources, and reassured it is okay if they do not want to view or send the photo. As an additional precaution, young children can also be told that, to make sure they are safe, their parents will get a message if they do view it.

The second feature, CSAM detection in iCloud Photos, is designed to keep CSAM off iCloud Photos without providing information to Apple about any photos other than those that match known CSAM images. CSAM images are illegal to possess in most countries, including the United States. This feature only impacts users who have chosen to use iCloud Photos to store their photos. It does not impact users who have not chosen to use iCloud Photos. There is no impact to any other on-device data. This feature does not apply to Messages.

Communication safety in Messages

Who can use communication safety in Messages?

Communication safety in Messages is only available for accounts set up as families in iCloud. Parent/guardian accounts must opt in to turn on the feature for their family group. Parental notifications can only be enabled by parents/guardians for child accounts age 12 or younger.

Does this mean Messages will share information with Apple or law enforcement?

No. Apple never gains access to communications as a result of this feature in Messages. This feature does not share any information with Apple, NCMEC or law enforcement. The communications safety feature in Messages is separate from CSAM detection for iCloud Photos — see below for more information about that feature.

Does this break end-to-end encryption in Messages?

No. This doesn't change the privacy assurances of Messages, and Apple never gains access to communications as a result of this feature. Any user of Messages, including those with communication safety enabled, retains control over what is sent and to whom. If the feature is enabled for the child account, the device will evaluate images in Messages and present an intervention if the image is determined to be sexually explicit. For accounts of children age 12 and under, parents can set up parental notifications which will be sent if the child confirms and sends or views an image that has been determined to be sexually explicit. None of the communications, image evaluation, interventions, or notifications are available to Apple.

Does this feature prevent children in abusive homes from seeking help?

The communication safety feature applies only to sexually explicit photos shared or received in Messages. Other communications that victims can use to seek help, including text in Messages, are unaffected. We are also adding additional support to Siri and Search to provide victims — and people who know victims — more guidance on how to seek help.

Will parents be notified without children being warned and given a choice?

No. First, parent/guardian accounts must opt-in to enable communication safety in Messages, and can only choose to turn on parental notifications for child accounts age 12 and younger. For child accounts age 12 and younger, each instance of a sexually explicit image sent or received will warn the child that if they continue to view or send the image, their parents will be sent a notification. Only if the child proceeds with sending or viewing an image after this warning will the notification be sent. For child accounts age 13–17, the child is still warned and asked if they wish to view or share a sexually explicit image, but parents are not notified.

CSAM detection

Does this mean Apple is going to scan all the photos stored on my iPhone?

No. By design, this feature only applies to photos that the user chooses to upload to iCloud Photos, and even then Apple only learns about accounts that are storing collections of known CSAM images, and only the images that match to known CSAM. The system does not work for users who have iCloud Photos disabled. This feature does not work on your private iPhone photo library on the device.

Will this download CSAM images to my iPhone to compare against my photos?

No. CSAM images are not stored on or sent to the device. Instead of actual images, Apple uses unreadable hashes that are stored on device. These hashes are strings of numbers that represent known CSAM images, but it isn't possible to read or convert those hashes into the CSAM images they are based on. This set of image hashes is based on images acquired and validated to be CSAM by at least two child safety organizations. Using new applications of cryptography, Apple is able to use these hashes to learn only about iCloud Photos accounts that are storing collections of photos that match to these known CSAM images, and is then only able to learn about photos that are known CSAM, without learning about or seeing any other photos.

Why is Apple doing this now?

One of the significant challenges in this space is protecting children while also preserving the privacy of users. With this new technology, Apple will learn about known CSAM photos being stored in iCloud Photos where the account is storing a collection of known CSAM. Apple will not learn anything about other data stored solely on device.

Existing techniques as implemented by other companies scan all user photos stored in the cloud. This creates privacy risk for all users. CSAM detection in iCloud Photos provides significant privacy benefits over those techniques by preventing Apple from learning about photos unless they both match to known CSAM images and are included in an iCloud Photos account that includes a collection of known CSAM.

How will CSAM detection in iCloud Photos handle photos of my kids in the bathtub, or other innocent images that involve child nudity?

CSAM detection for iCloud Photos is designed to find matches to known CSAM images. The system uses image hashes that are based on images acquired and validated to be CSAM by at least two child safety organizations. It is not designed for images that contain child nudity that are not known CSAM images.

Does turning off iCloud Photos disable CSAM detection?

Yes. When iCloud Photos is deactivated, no images are processed. CSAM detection is applied only as part of the process for storing images in iCloud Photos.

Can Apple unlock my iPhone using this new system?

No. CSAM detection for iCloud Photos does not provide any access to data stored only on the device. Nothing about the security guarantees of your device are changed with CSAM detection for iCloud Photos, including the protections provided by passcode-locking your device.

Does this impact existing photos or only new photos that are uploaded?

When the feature rolls out, CSAM detection for iCloud Photos will apply to both photos that are already in a user's account as well as new photos that they upload.

Security for CSAM detection for iCloud Photos

Can the CSAM detection system in iCloud Photos be used to detect things other than CSAM?

Our process is designed to prevent that from happening. CSAM detection for iCloud Photos is built so that the system only works with CSAM image hashes provided by NCMEC and other child safety organizations. This set of image hashes is based on images acquired and validated to be CSAM by at least two child safety organizations. There is no automated reporting to law enforcement, and Apple conducts human review before making a report to NCMEC. As a result, the system is only designed to report photos that are known CSAM in iCloud Photos. In most countries, including the United States, simply possessing these images is a crime and Apple is obligated to report any instances we learn of to the appropriate authorities.

Could governments force Apple to add non-CSAM images to the hash list?

No. Apple would refuse such demands and our system has been designed to prevent that from happening. Apple's CSAM detection capability is built solely to detect known CSAM images stored in iCloud Photos that have been identified by experts at NCMEC and other child safety groups. The set of image hashes used for matching are from known, existing images of CSAM and only contains entries that were independently submitted by two or more child safety organizations operating in separate sovereign jurisdictions. Apple does not add to the set of known CSAM image hashes, and the system is designed to be auditable. The same set of hashes is stored in the operating system of every iPhone and iPad user, so targeted attacks against only specific individuals are not possible under this design. Furthermore, Apple conducts human review before making a report to NCMEC. In a case where the system identifies photos that do not match known CSAM images, the account would not be disabled and no report would be filed to NCMEC.

We have faced demands to build and deploy government-mandated changes that degrade the privacy of users before, and have steadfastly refused those demands. We will continue to refuse them in the future. Let us be clear, this technology is limited to detecting CSAM stored in iCloud and we will not accede to any government's request to expand it.

Can non-CSAM images be "injected" into the system to identify accounts for things other than CSAM?

Our process is designed to prevent that from happening. The set of image hashes used for matching are from known, existing images of CSAM that have been acquired and validated by at least two child safety organizations. Apple does not add to the set of known CSAM image hashes. The same set of hashes is stored in the operating system of every iPhone and iPad user, so targeted attacks against only specific individuals are not possible under our design. Finally, there is no automated reporting to law enforcement, and Apple conducts human review before making a report to NCMEC. In the unlikely event of the system identifying images that do not match known CSAM images, the account would not be disabled and no report would be filed to NCMEC.

Will CSAM detection in iCloud Photos falsely report innocent people to law enforcement?

No. The system is designed to be very accurate, and the likelihood that the system would incorrectly identify any given account is less than one in one trillion per year. In addition, any time an account is identified by the system, Apple conducts human review before making a report to NCMEC. As a result, system errors or attacks will not result in innocent people being reported to NCMEC.