

Google AI flagged parents' accounts for potential abuse over nude photos of their sick kids

A concerned father says that after using his Android smartphone to take photos of an infection on his toddler's groin, Google flagged the images as child sexual abuse material (CSAM), according to a report from *The New York Times*. The company closed his accounts and filed a report with the National Center for Missing and Exploited Children (NCMEC) and spurred a police investigation, highlighting the complications of trying to tell the difference between potential abuse and an innocent photo once it becomes part of a user's digital library, whether on their personal device or in cloud storage.

Concerns about the consequences of blurring the lines for what should be considered private were aired last year when Apple announced its Child Safety plan. As part of the plan, Apple would locally scan images on Apple devices before they're uploaded to iCloud and then match the images with the NCMEC's hashed database of known CSAM. If enough matches were found, a human moderator would then review the content and lock the user's account if it contained CSAM.

The accounts were taken away due to content that "might be illegal"

The Electronic Frontier Foundation (EFF), a nonprofit digital rights group, slammed Apple's plan, saying it could "open a backdoor to your private life" and that it represented "a decrease in privacy for all iCloud Photos users, not an

improvement.”

Apple eventually placed the stored image scanning part on hold, but with the launch of iOS 15.2, it proceeded with including an optional feature for child accounts included in a family sharing plan. If parents opt-in, then on a child’s account, the Messages app “analyzes image attachments and determines if a photo contains nudity, while maintaining the end-to-end encryption of the messages.” If it detects nudity, it blurs the image, displays a warning for the child, and presents them with resources intended to help with safety online.

The main incident highlighted by *The New York Times* took place in February 2021, when some doctor’s offices were still closed due to the COVID-19 pandemic. As noted by the *Times*, Mark (whose last name was not revealed) noticed swelling in his child’s genital region and, at the request of a nurse, sent images of the issue ahead of a video consultation. The doctor wound up prescribing antibiotics that cured the infection.

According to the *NYT*, Mark received a notification from Google just two days after taking the photos, stating that his accounts had been locked due to “harmful content” that was “a severe violation of Google’s policies and might be illegal.”

Like many internet companies, including Facebook, Twitter, and Reddit, Google has used hash matching with Microsoft’s PhotoDNA for scanning uploaded images to detect matches with known CSAM. In 2012, it led to the arrest of a man who was a registered sex offender and used Gmail to send images of a young girl.

In 2018, Google announced the launch of its Content Safety API AI toolkit that can “proactively identify never-before-seen CSAM imagery so it can be reviewed and, if confirmed as CSAM, removed and reported as quickly as possible.” It uses the tool for its own services and, along with a video-targeting CSAI

Match hash matching solution developed by YouTube engineers, offers it for use by others as well.

Google “Fighting abuse on our own platforms and services”:

We identify and report CSAM with trained specialist teams and cutting-edge technology, including machine learning classifiers and hash-matching technology, which creates a “hash”, or unique digital fingerprint, for an image or a video so it can be compared with hashes of known CSAM. When we find CSAM, we report it to the National Center for Missing and Exploited Children (NCMEC), which liaises with law enforcement agencies around the world.

A Google spokesperson told the *Times* that Google only scans users’ personal images when a user takes “affirmative action,” which can apparently include backing their pictures up to Google Photos. When Google flags exploitative images, the *Times* notes that Google’s required by federal law to report the potential offender to the CyberTipLine at the NCMEC. In 2021, Google reported 621,583 cases of CSAM to the NCMEC’s CyberTipLine, while the NCMEC alerted the authorities of 4,260 potential victims, a list that the *NYT* says includes Mark’s son.

Mark ended up losing access to his emails, contacts, photos, and even his phone number, as he used Google Fi’s mobile service, the *Times* reports. Mark immediately tried appealing Google’s decision, but Google denied Mark’s request. The San Francisco Police Department, where Mark lives, opened an investigation into Mark in December 2021 and got ahold of all the information he stored with Google. The investigator on the case ultimately found that the incident “did not meet the elements of a crime and that no crime occurred,” the *NYT* notes.

“Child sexual abuse material (CSAM) is abhorrent and we’re committed to preventing the spread of it on our platforms,”

Google spokesperson Christa Muldoon said in an emailed statement to *The Verge*. “We follow US law in defining what constitutes CSAM and use a combination of hash matching technology and artificial intelligence to identify it and remove it from our platforms. Additionally, our team of child safety experts reviews flagged content for accuracy and consults with pediatricians to help ensure we’re able to identify instances where users may be seeking medical advice.”

While protecting children from abuse is undeniably important, critics argue that the practice of scanning a user’s photos unreasonably encroaches on their privacy. Jon Callas, a director of technology projects at the EFF called Google’s practices “intrusive” in a statement to the *NYT*. “This is precisely the nightmare that we are all concerned about,” Callas told the *NYT*. “They’re going to scan my family album, and then I’m going to get into trouble.”

<https://www.theverge.com/2022/8/21/23315513/google-photos-csam-scanning-account-deletion-investigation>