AUGUST 22, 2022

eff.org

Google's Scans of Private Photos Led to False Accusations of Child Abuse

ΕΛΛΗΝΙΚΆ ESPAÑOL

Internet users' private messages, files, and photos of everyday people are increasingly being examined by tech companies, which check the data against government databases. While this is not a new practice, the public is being told this massive scanning should extend to nearly every reach of their online activity so that police can more productively investigate crimes related to child sexual abuse images, sometimes called CSAM.

We don't know much about how the public gets watched in this way. That's because neither the tech companies that do the scanning, nor the government agencies they work with, share details of how it works. But we do know that the scanning is far from perfect, despite claims to the contrary. It makes mistakes, and those mistakes can result in false accusations of child abuse. We don't know how often such false accusations happen, or how many people get hurt by them.

The spread of CSAM causes real harms, and tech companies absolutely should work on new ways of fighting it. We have suggested some good ways of doing so, like building better reporting tools, privacy-respecting

warning messages, and metadata analysis.

An article published yesterday in the New York Times reports on how Google made two of these false accusations, and the police follow-up. It also highlights Google's refusal to correct any of the damage done by its erroneous scans, and the company's failed human review processes. This type of scanning is increasingly ubiquitous on tech products we all use, and governments around the world want to extend its reach even further, to check even our most private, encrypted conversations. The article is especially disturbing, not just for the harm it describes to the two users Google falsely accused, but also as a warning of potentially many more such mistakes to come.

Google's AI System Failed, And Its Employees Failed Too

In February of last year, Google's algorithms wrongly flagged photos taken by two fathers in two different states as being images of child abuse. In both cases, the fathers—one in San Francisco, one in Houston—had small children with infections on their genitals, and had taken photos of the area at the request of medical professionals.

Google's algorithms, and the employees who oversee them, had a different opinion about the photos. Without informing either parent, Google reported them to the government. That resulted in local police departments investigating the parents.

The company also chose to perform its own investigation. In the case of Mark, the San Francisco father, Google employees looked at not just the photo that had been flagged by their mistaken AI, but his entire collection of family and friend photos.

Both the Houston Police Department and the San Francisco Police Department quickly cleared the fathers of any wrongdoing. But Google refused to hear Mark's appeal or reinstate his account, even after he brought the company documentation showing that the SFPD had determined there was "no crime committed." Remarkably, even after the New York Times contacted Google and the error was clear, the company continues to refuse to restore any of Mark's Google accounts, or help him get any data back.

Google's False Accusations Cause Real Harm

Google has a right to decide which users it wants to host. But it was Google's incorrect algorithms, and Google's failed human review process, which caused innocent people to be investigated by the police in these cases. It was also Google's choice to destroy without warning and without due process these fathers' email accounts, videos, photos, and in one case, telephone service. The consequences of the company's error are not trivial.

We don't know how many other people Google has wrongly accused of child abuse, but it's likely many more than these two. Given the massive scope of the content it scans, it could be hundreds, or thousands.

Mark and Cassio, the two fathers wrongly flagged by Google, were accused within one day of each other in February 2021. That could be coincidental timing, or it could suggest that one or more flaws in Google's system—either flaws in the AI software, or flaws in the human review process—were particularly manifest at that time.

Google's faulty CSAM scans caused real harm in these cases, and it's not hard to imagine how they could be more harmful in other cases. Once both Google employees and police officers have combed through an accused parent's files, there could be consequences that have nothing to do with CSAM. Police could find evidence of drug use or other wrongdoing, and choose to punish parents for those unrelated crimes, without having suspected them in the first place. Google could choose to administer its own penalties, as it did to Mark and Cassio.

Despite what had happened to them, both Mark and Cassio, the Houston father, felt empowered to speak out to a reporter. But systems like this could report on vulnerable minorities, including LGBT parents in locations where police and community members are not friendly to them. Google's system could wrongly report parents to authorities in autocratic countries, or locations with corrupt police, where wrongly accused parents could not be assured of proper due process.

Governments Want More Unaccountable CSAM Scans

Google isn't the only company doing scans like this. But evidence is mounting that the scans are simply not accurate. A Facebook study on 150 accounts that were reported to authorities for alleged CSAM found that 75% of the accounts sent images that were "non-malicious" and were sending images for reasons "such as outrage or poor humor." LinkedIn found 75 accounts that were reported to EU authorities in the second half of 2021, due to files that it matched with known CSAM. But upon manual review, only 31 of those cases involved confirmed CSAM. (LinkedIn uses PhotoDNA, the software product specifically recommended by the U.S. sponsors of the EARN IT Bill.)

In the past few years, we've seen governments push for more scanning. Last year, Apple proposed a form of on-device scanning on all of its devices that would search user photos and report matches to authorities. That program was scuttled after a public outcry. This year in the U.S., the Senate Judiciary Committee considered and passed the EARN IT Act, which would have opened the door for states to compel companies to use CSAM scanners. (The EARN IT Act hasn't been considered in a floor debate by either house of Congress.) The European Union is considering a new CSAM detection law as well. The EU proposal would not only search for known and new abuse images, it would use AI to scan text messages for "grooming," in an attempt to judge abuse that might happen in the future.

Earlier this month, EU Commissioner Ylva Johnasson wrote a blog post asserting that the scanners they propose to use have accuracy rates "significantly above 90%." She asserts "grooming" detection will be 88% accurate, "before human review."

These accuracy rates are nothing to brag about. If billions of private messages in the EU are scanned with a false positive rate of "above 90%," it will result in millions of falsely flagged messages. This avalanche of false positives will be a humanitarian disaster even in wealthy democracies with rule of law—to say nothing of the autocracies and backsliding democracies, which will demand similar systems. Defenders of these systems point to the very real harms of CSAM, and some argue that false positives–the kind that result in erroneous reports like those in the article–are acceptable collateral damage.

What we're being asked to accept here is nothing less than "bugs in our pockets." Governments want companies like Google and Apple to constantly scan every digital space we have, including private spaces.

But we're seeing the results when companies like Google second-guess their own users' family lives—and even second-guess the police.

The Solution is Real Privacy

At EFF, we've been fighting against spying on peoples' digital lives for more than 30 years. When police want to look at our private messages or files, they should follow the 4th Amendment and get a warrant. Period.

As for private companies, they should be working to limit their need and ability to trawl our private content. When we have private conversations with friends, family, or medical professionals, they should be protected using end-to-end encryption. In end-to-end encrypted systems, the service provider doesn't have the option of looking at the message, even if they wanted to. Companies should also commit to encrypted backups, something EFF has requested for some time now.

The answer to a better internet isn't racing to come up with the best scanning software. There's no way to protect human rights while having AI scan peoples' messages to locate wrongdoers. The real answer is staring us right in the face: law enforcement, and elected leaders, that work to *coexist* with strong encryption and privacy, not break them down.