Tennessee v. Meta
Case No. 23-1364-IV

DX2956

Log in

New Tools to Help Protect Against Sextortion and Intimate Image Abuse

April 11, 2024

Takeaways:

We're testing new features to help protect young people from sextortion and intimate image abuse, and to make it more difficult for potential scammers and criminals to find and interact with teens.

We're also testing new ways to help people spot potential sextortion scams, encourage them to report and empower them to say no to anything that makes them feel uncomfortable.

We've started sharing more signals about sextortion accounts to other tech companies through Lantern, helping disrupt this criminal activity across the internet.

Financial sextortion is a horrific crime. We've spent years working closely with experts, including those experienced in fighting these crimes, to understand the tactics scammers use to find and extort victims online, so we can develop effective ways to help stop them.

Today, we're sharing an overview of our latest work to tackle these crimes. This includes new tools we're testing to help protect people from sextortion and other forms of intimate image abuse, and to make it as hard as possible for scammers to find potential targets – on Meta's apps and across the internet. We're also testing new measures to support young people in recognizing and protecting themselves from sextortion scams.

These updates build on our longstanding work to help protect young people from unwanted or potentially harmful contact. We default teens into stricter message settings so they can't be messaged by anyone they're not already connected to, show Safety Notices to teens who are already in contact with potential scam accounts, and offer a dedicated option for people to report DMs that are threatening to share private images. We also supported the National Center for Missing and Exploited Children (NCMEC) in developing Take It Down, a platform that lets young people take back control of their intimate images and helps prevent them being shared online – taking power away from scammers.

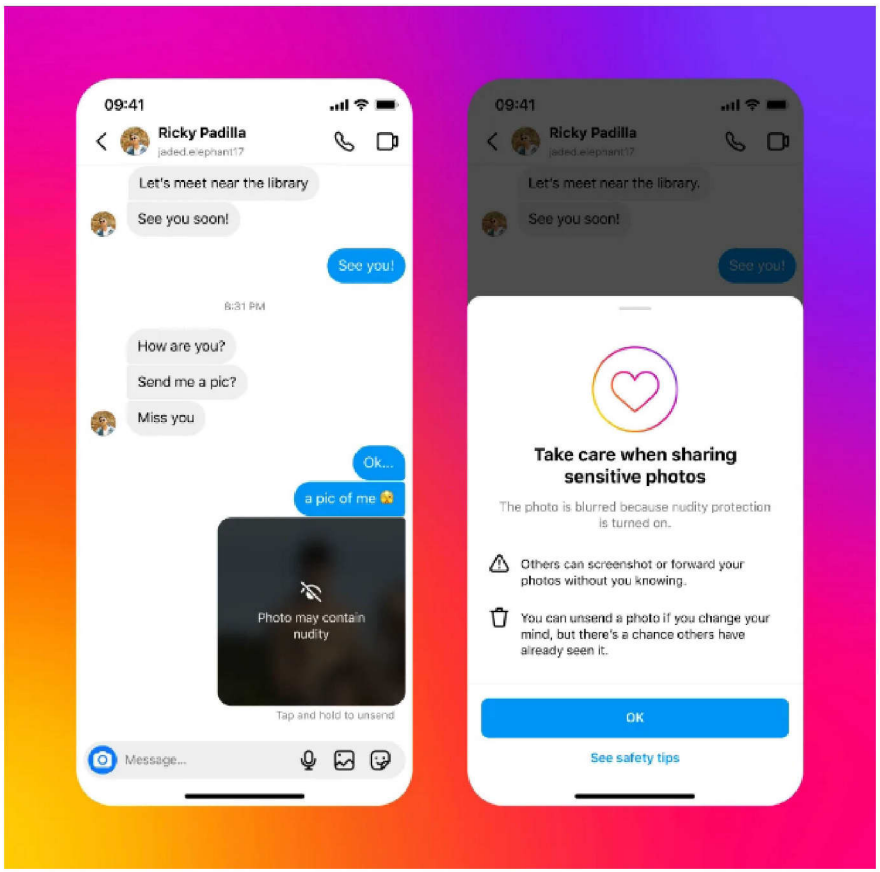
Introducing Nudity Protection in DMs

While people overwhelmingly use DMs to share what they love with their friends, family or favorite creators, sextortion scammers may also use private messages to share or ask for intimate images. To

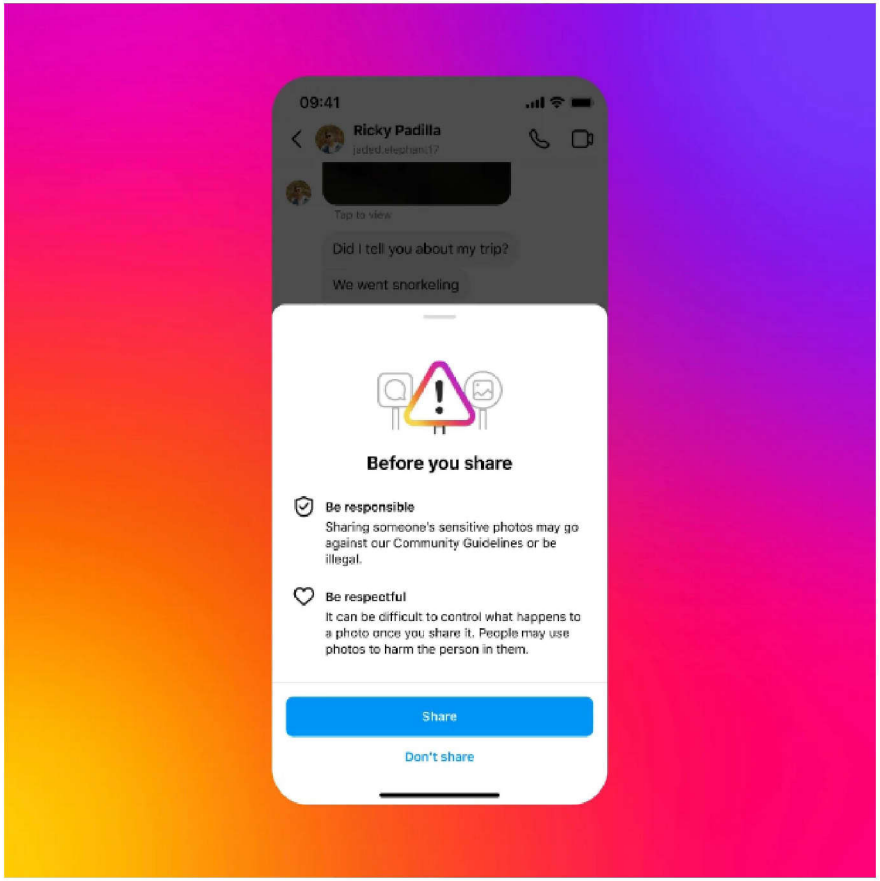
help address this, we'll soon start testing our new nudity protection feature in Instagram DMs, which blurs images detected as containing nudity and encourages people to think twice before sending nude images. This feature is designed not only to protect people from seeing unwanted nudity in their DMs, but also to protect them from scammers who may send nude images to trick people into sending their own images in return.

Nudity protection will be turned on by default for teens under 18 globally, and we'll show a notification to adults encouraging them to turn it on.

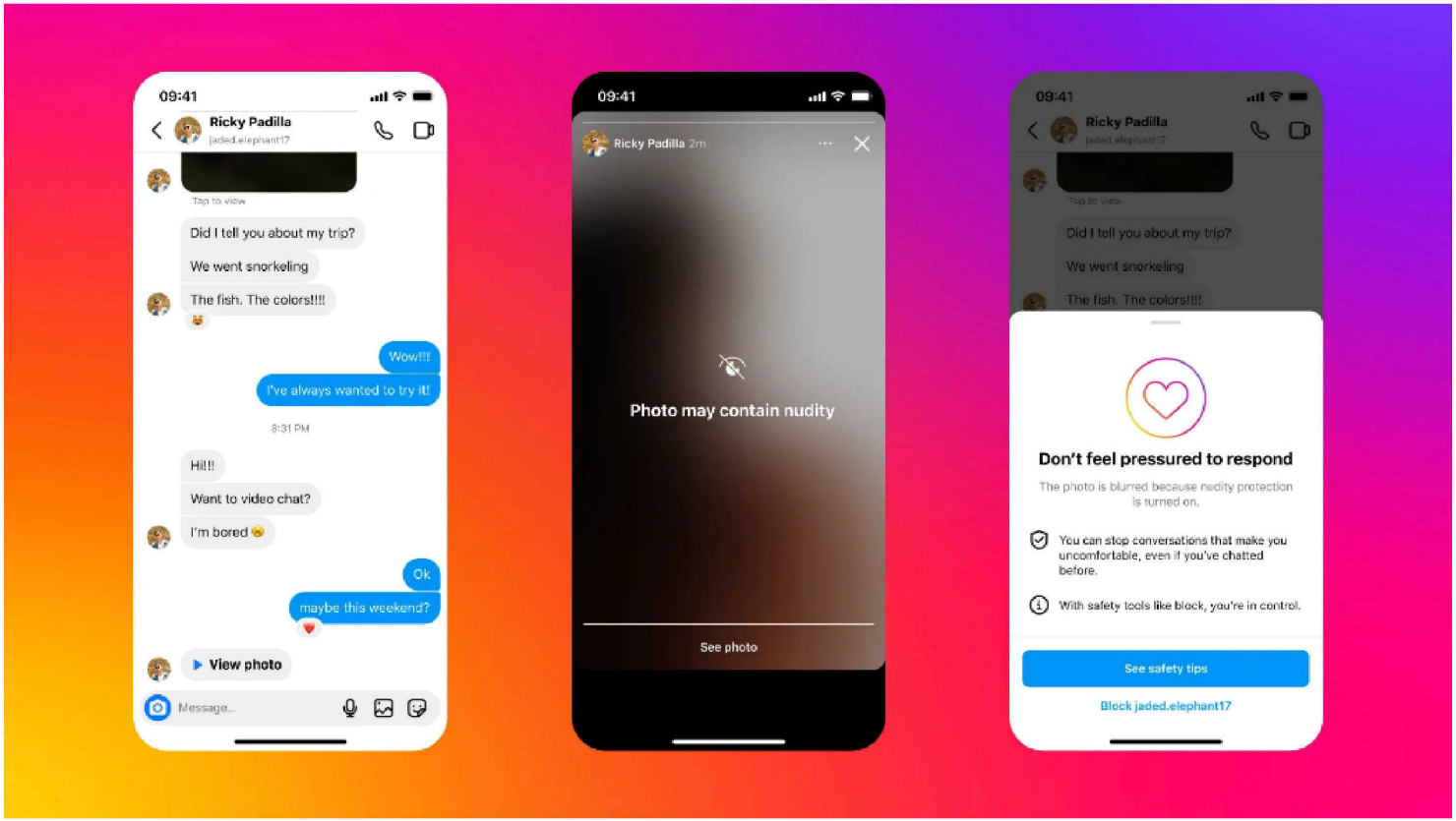
When nudity protection is turned on, people sending images containing nudity will see a message reminding them to be cautious when sending sensitive photos, and that they can unsend these photos if they've changed their mind.



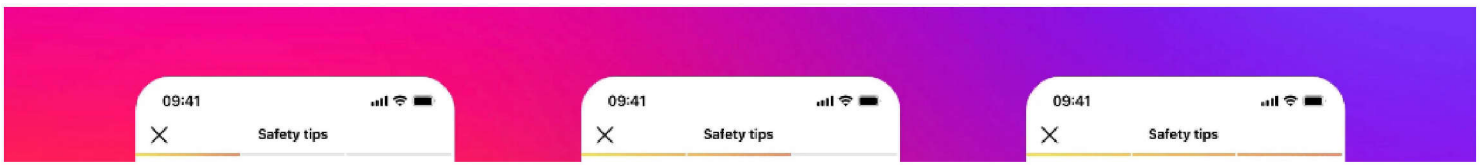
Anyone who tries to forward a nude image they've received will see a message encouraging them to reconsider.

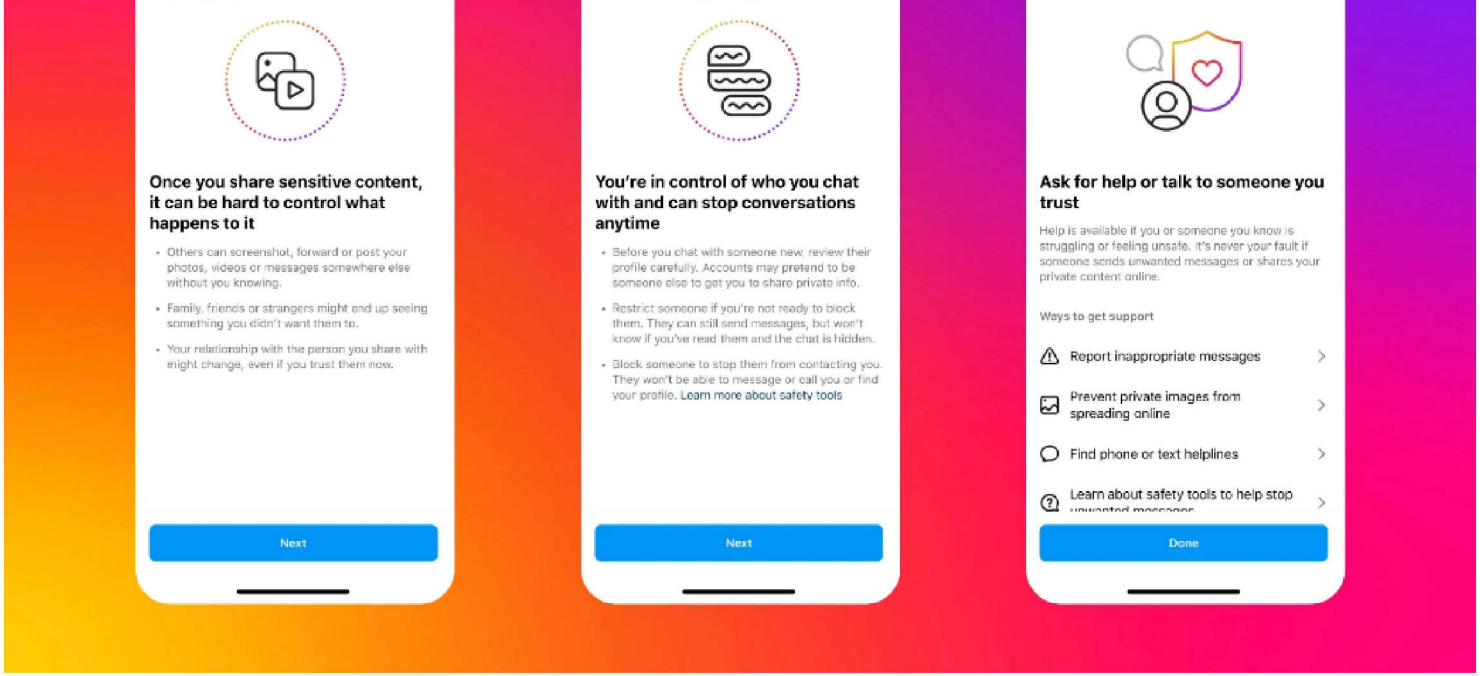


When someone receives an image containing nudity, it will be automatically blurred under a warning screen, meaning the recipient isn't confronted with a nude image and they can choose whether or not to view it. We'll also show them a message encouraging them not to feel pressure to respond, with an option to block the sender and report the chat.



When sending or receiving these images, people will be directed to safety tips, developed with guidance from experts, about the potential risks involved. These tips include reminders that people may screenshot or forward images without your knowledge, that your relationship to the person may change in the future, and that you should review profiles carefully in case they're not who they say they are. They also link to a range of resources, including Meta's Safety Center, support helplines, StopNCII.org for those over 18, and Take It Down for those under 18.





Nudity protection uses on-device machine learning to analyze whether an image sent in a DM on Instagram contains nudity. Because the images are analyzed on the device itself, nudity protection will work in end-to-end encrypted chats, where Meta won't have access to these images – unless someone chooses to report them to us.

"Companies have a responsibility to ensure the protection of minors who use their platforms. Meta's proposed device-side safety measures within its encrypted environment is encouraging. We are hopeful these new measures will increase reporting by minors and curb the circulation of online child exploitation."

— John Shehan, Senior Vice President, National Center for Missing & Exploited Children.

"As an educator, parent, and researcher on adolescent online behavior, I applaud Meta's new feature that handles the exchange of personal nude content in a thoughtful, nuanced, and appropriate way. It reduces unwanted exposure to potentially traumatic images, gently introduces cognitive dissonance to those who may be open to sharing nudes, and educates people about the potential downsides involved. Each of these should help decrease the incidence of sextortion and related harms, helping to keep young people safe online."

— Dr. Sameer Hinduja, Co-Director of the Cyberbullying Research Center and Faculty Associate at the Berkman Klein Center at Harvard University.

Preventing Potential Scammers from Connecting

with Teens

We take severe action when we become aware of people engaging in sextortion: we remove their account, take steps to prevent them from creating new ones and, where appropriate, report them to the NCMEC and law enforcement. Our expert teams also work to investigate and disrupt networks of these criminals, disable their accounts and report them to NCMEC and law enforcement – including several networks in the last year alone.

Now, we're also developing technology to help identify where accounts may potentially be engaging in sextortion scams, based on a range of signals that could indicate sextortion behavior. While these signals aren't necessarily evidence that an account has broken our rules, we're taking precautionary steps to help prevent these accounts from finding and interacting with teen accounts. This builds on the work we already do to prevent other potentially suspicious accounts from finding and interacting with teens.

One way we're doing this is by making it even harder for potential sextortion accounts to message or interact with people. Now, any message requests potential sextortion accounts try to send will go straight to the recipient's hidden requests folder, meaning they won't be notified of the message and never have to see it. For those who are already chatting to potential scam or sextortion accounts, we show Safety Notices encouraging them to report any threats to share their private images, and reminding them that they can say no to anything that makes them feel uncomfortable.

For teens, we're going even further. We already restrict adults from starting DM chats with teens they're not connected to, and in January we announced stricter messaging defaults for teens under 16 (under 18 in certain countries), meaning they can only be messaged by people they're already connected to – no matter how old the sender is. Now, we won't show the "Message" button on a teen's profile to potential sextortion accounts, even if they're already connected. We're also testing hiding teens from these accounts in people's follower, following and like lists, and making it harder for them to find teen accounts in Search results.

New Resources for People Who May Have Been

Approached by Scammers

We're testing new pop-up messages for people who may have interacted with an account we've removed for sextortion. The message will direct them to our expert-backed resources, including our Stop Sextortion Hub, support helplines, the option to reach out to a friend, StopNCII.org for those over 18, and Take It Down for those under 18.

We're also adding new child safety helplines from around the world into our in-app reporting flows. This means when teens report relevant issues – such as nudity, threats to share private images or sexual exploitation or solicitation – we'll direct them to local child safety helplines where available.

Fighting Sextortion Scams Across the Internet

In November, we announced we were founding members of Lantern, a program run by the Tech Coalition that enables technology companies to share signals about accounts and behaviors that violate their child safety policies.

This industry cooperation is critical, because predators don't limit themselves to just one platform – and the same is true of sextortion scammers. These criminals target victims across the different apps they use, often moving their conversations from one app to another. That's why we've started to share more sextortion-specific signals to Lantern, to build on this important cooperation and try to stop sextortion scams not just on individual platforms, but across the whole internet.

RELATED ARTICLES

Check out more announcements

#INSTAGRAM #ANNOUNCEMENTS

Close Friends Only: Megan Thee Stallion and GloRilla on Instagram

#ANNOUNCEMENTS

Survey: What are Threads' Users Texting About?

#ANNOUNCEMENTS

Close Friends Only: Reneé Rapp & Rachel Sennott on Instagram

View More →

Leadership
Brand
Brand Kit
Working at Instagram
Politics

Reels
Stories
Messenger
Shopping
Search & Explore

Safety Tools
Privacy Tools
Account Security

Programs
Anti-Bullying
Parents

Business
Creators

Meta
Family Center
Help Center↗



English (US)Instagram from MetaAPIPrivacyTermsSitemap