

Online Bullying



Instagram has introduced a new anti-bullying feature that notifies people when their comment may be considered offensive before it's posted. The intervention gives people a chance to reflect and undo their comment and prevents the recipient from receiving the harmful comment.

'Our Commitment to Lead the Fight Against Online Bullying'

Our mission is to connect you with the people and things you love, which only works if people feel comfortable expressing themselves on Instagram. We know bullying is a challenge many face, particularly young people. We are committed to leading the industry in the fight against online bullying, and we are rethinking the whole experience of Instagram to meet that commitment.

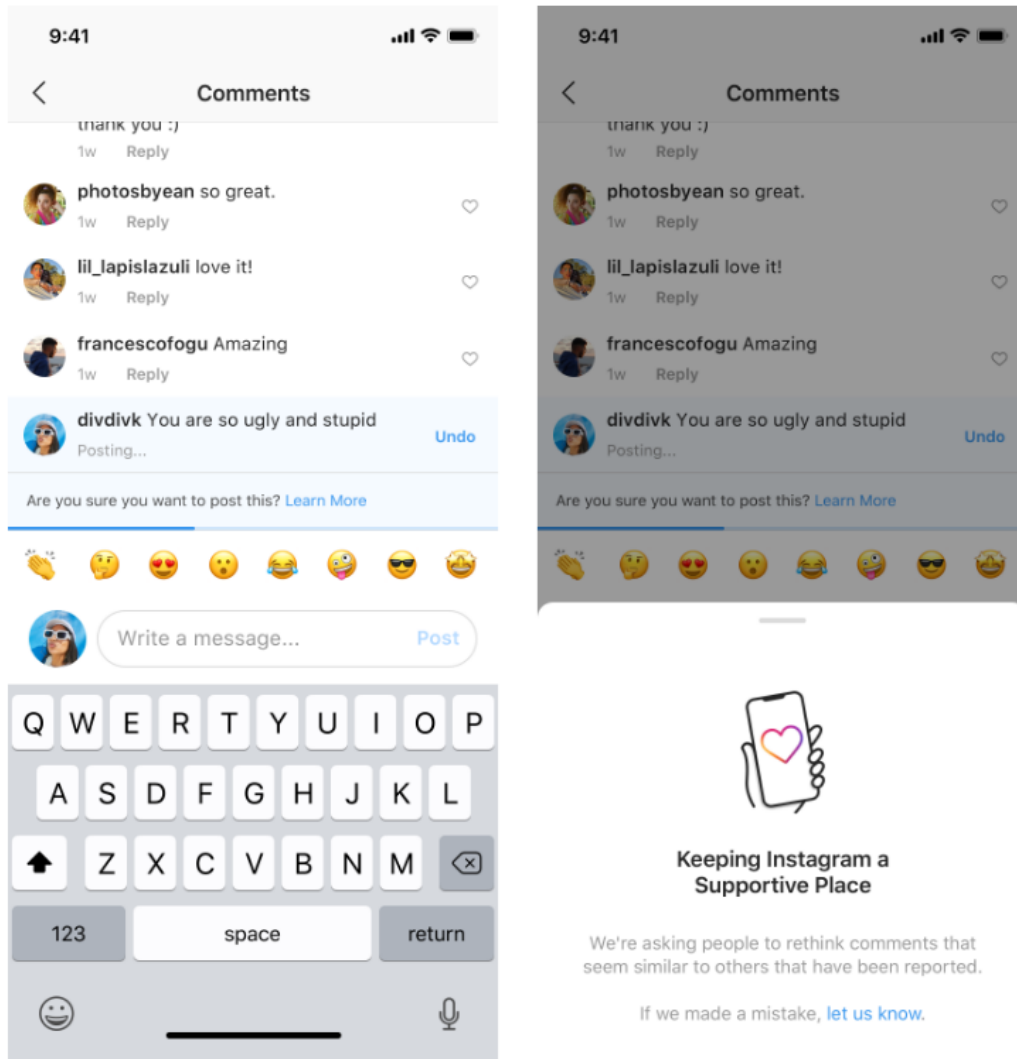
We can do more to prevent bullying from happening on Instagram, and we can do more to empower the targets of bullying to stand up for themselves. Today we're announcing one new feature in both areas. These tools are grounded in a deep understanding of how people bully each other and how they respond to bullying on Instagram, but they're only two steps on a longer path.

Encouraging Positive Interactions

Online bullying is a complex issue. For years now, we have used artificial intelligence to detect bullying and other types of harmful content in [comments](#), [photos and videos](#). As our community grows, so does our investment in technology. This is especially crucial for teens since they are less likely to report online bullying even when they are the ones who experience it the most.

In the last few days, we started rolling out a new feature powered by AI that notifies people when their comment may be considered offensive before it's posted. This intervention gives

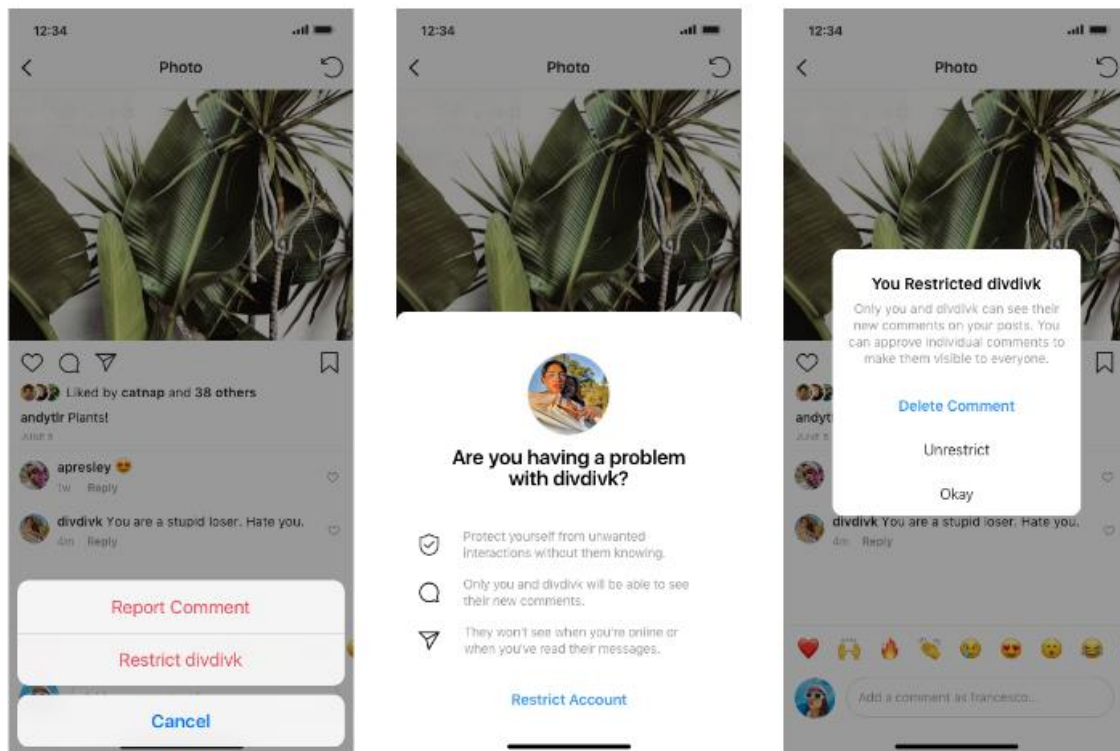
people a chance to reflect and undo their comment and prevents the recipient from receiving the harmful comment notification. From early tests of this feature, we have found that it encourages some people to undo their comment and share something less hurtful once they have had a chance to reflect.



Protecting Your Account From Unwanted Interactions With Restrict

While identifying and removing bullying on Instagram is important, we also need to empower our community to stand up to this kind of behaviour. We've heard from young people in our community that they're reluctant to block, unfollow, or report their bully because it could escalate the situation, especially if they interact with their bully in real life. Some of these actions also make it difficult for a target to keep track of their bully's behaviour.

We wanted to create a feature that allows people to control their Instagram experience, without notifying someone who may be targeting them. Soon, we will begin testing a new way to protect your account from unwanted interactions called Restrict. Once you Restrict someone, comments on your posts from that person will only be visible to that person. You can choose to make a restricted person's comments visible to others by approving their comments. Restricted people won't be able to see when you're active on Instagram or when you've read their direct messages.



It's our responsibility to create a safe environment on Instagram. This has been an important priority for us for some time, and we are continuing to invest in better understanding and tackling this problem. I look forward to sharing more updates soon.

Adam Mosseri, Head of Instagram

Liam Dodd
WKS CEOP/ E-safety Coordinator