Instagram Is Using AI to Alert Bullies to Offensive Captions

Whether or not that serves as a deterrent for abusive posting on Instagram is another story.

December 16, 2019 2 min read

This story originally appeared on Engadget

Presented by
Instagram’s anti-bullying efforts now extend to discouraging abusive behavior in posts, not just comments. It’s introducing a feature that warns people when their post captions might be offensive, using AI to detect language similar to posts that have previously been reported. In theory, this makes bullies second-guess their vitriol, stick to kinder language and learn about the social network’s rules. You can ignore the warning if Instagram made a mistake.

Related: Instagram Stories: 18 Marketers to Follow for Incredible Inspiration

The anti-bullying notice is available now in “select countries,” and should be available elsewhere in the “coming months.”

Whether or not this proves effective is uncertain. Instagram said that its efforts to reduce bullying in comments have been “promising,” but that doesn’t guarantee similar performance for the posts themselves. Someone caught up in the heat of the moment might hit “share anyway,” consequences be damned. And there will be moments where a vicious tone may not represent

bullying at all – calling a politician stupid may not be constructive, but it’s not bullying. Still, this could be helpful if it leads even a handful of people to mend their ways.

https://www.entrepreneur.com/article/343895