

Instagram lets users filter insults in private message requests

21 April 2021



Users will be able to create their own list of blocked words for Instagram's messaging service

Instagram started offering users on Wednesday the option to filter incoming direct message requests for hurtful language, a step aimed at showing the Facebook-owned social network is serious about reducing online harassment.

"Because DMs are private conversations, we don't proactively look for hate speech or bullying the same way we do elsewhere," Instagram said in a blog post.

But requests to open new private chats are "where people usually receive abusive messages," it added.

Running on the user's device to preserve privacy, rather than Instagram's servers, the new filters will block requests to open a direct messaging conversation that contain offensive words, expressions or emojis.

As well as a set of pre-defined filters, users will be able to add their own terms to their personal block list, just as they already can for comments on their

posts.

At first, users in seven countries—Britain, France, Ireland, Germany, Australia, New Zealand, and Canada—will be given the option, with more to follow "over the next few months".

"These are the countries where public figures, especially [football players](#), have been victims of abuse and offensive language in direct messages," Instagram public policy manager in France Clotilde Briend told AFP.

"Most hate speech comes in these direct message requests. We'll give everyone the option of reporting these messages to us, or not," she added.

Facebook is also considering rolling the filter out to its other chat properties Messenger and WhatsApp, Briend said.

With more than one billion users, Instagram has like other social networks spent years trying to fight harassment, [hate speech](#) and disinformation.

In 2019, it set up [artificial intelligence](#)-based warnings to people trying to post insulting messages, and recently said it would use machine learning to work out [users'](#) real age.

In March, it confirmed it would push ahead with a version of the platform geared for pre-teens, as the existing platform bans under-13s.

Childrens' groups and rights advocates from around the world have called on Facebook to drop the plans.

© 2021 AFP

APA citation: Instagram lets users filter insults in private message requests (2021, April 21) retrieved 29 November 2021 from <https://techxplore.com/news/2021-04-instagram-users-filter-insults-private.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.