

Meta turns to AI to protect minors from 'sextortion' on Instagram

April 11 2024, by Daxia ROJAS



Over 40 US states are suing Meta, saying it has 'profited from children's pain'

Meta said on Thursday it was developing new tools to protect teenage users from "sextortion" scams on its Instagram platform, which has been accused by US politicians of damaging the mental health of youngsters.

Gangs run sextortion scams by persuading people to provide explicit images of themselves and then threatening to release them to the public unless they receive money.

Meta said it was testing an AI-driven "nudity protection" tool that would find and blur images containing nudity that were sent to minors on the app's messaging system.

"This way, the recipient is not exposed to unwanted intimate content and has the choice to see the image or not," Capucine Tuffier, who is in charge of child protection at Meta France, told AFP.

The US company said it would also offer advice and safety tips to anyone sending or receiving such messages.

Some 3,000 [young people](#) fell victim to sexploitation scams in 2022 in the United States, according to the authorities there.

Separately, more than 40 US states began suing Meta in October in a case that accuses the company of having "profited from children's pain".

The legal filing alleged Meta had exploited young users by creating a business model designed to maximize time they spend on the platform despite harm to their health.

'On-device machine learning'

Meta announced in January it would roll out measures to protect under-18s that included tightening content restrictions and boosting parental supervision tools.

The firm said on Thursday that the latest tools were building on "our long-standing work to help protect young people from unwanted or

potentially harmful contact".

"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," the company said.

It added that the "nudity protection" tool used "on-device machine learning", a kind of Artificial Intelligence, to analyze images.

The firm, which is also constantly accused of violating the data privacy of its users, stressed that it would not have access to the images unless users reported them.

Meta said it would also use AI tools to identify accounts sending offending material and severely restrict their ability to interact with young users on the platform.

Whistle-blower Frances Haugen, a former Facebook engineer, publicized research in 2021 carried out internally by Meta—then known as Facebook—which showed the company had long been aware of the dangers its platforms posed for the mental health for young people.

© 2024 AFP

Citation: Meta turns to AI to protect minors from 'sextortion' on Instagram (2024, April 11) retrieved 2 May 2024 from

<https://techxplore.com/news/2024-04-meta-ai-minors-sextortion-instagram.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.