Meta oversight board wants deepfake rules for AI Age

July 25 2024

Credit: AI-generated image

Meta's oversight board on Thursday called on the tech titan to bring its rules regarding porn deepfakes out of the "photoshop" days and into the era of artificial intelligence.

The independent board, which is referred to as a top court for Meta

content moderation decisions, said it came up with the recommendation after reviewing two cases involving deepfake images of high-profile women in India and the United States.

In one case a deepfake shared on Instagram was left up despite a complaint and in the other the faked image was not allowed on the Meta platform. Both decisions prompted appeals to the board.

The board decided that the deepfakes involved in both cases violated a Meta rule against a practice called "derogatory sexualized photoshop", which needs to be made easier for people to understand, it said.

The board said that Meta defined the term as involving manipulated images sexualized in ways likely to be unwanted by those pictured.

Photoshop software for image editing was first released in 1990, and was so widely used it became a common reference for image tweaking.

But, referring to "photoshop" in a rule regarding porn deepfakes is "too narrow" with technology like generative AI available to create images or videos with simple text prompts, the board concluded.

The oversight board suggested Meta make clear it does not allow AI-created or manipulated non-consensual sexual content.

While Meta has agreed to abide by board decisions regarding specific content moderation decisions, it takes policy suggestions as recommendations it can adopt if it sees fit.

© 2024 AFP

Citation: Meta oversight board wants deepfake rules for AI Age (2024, July 25) retrieved 26 July 2024 from https://techxplore.com/news/2024-07-meta-oversight-board-deepfake-ai.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.