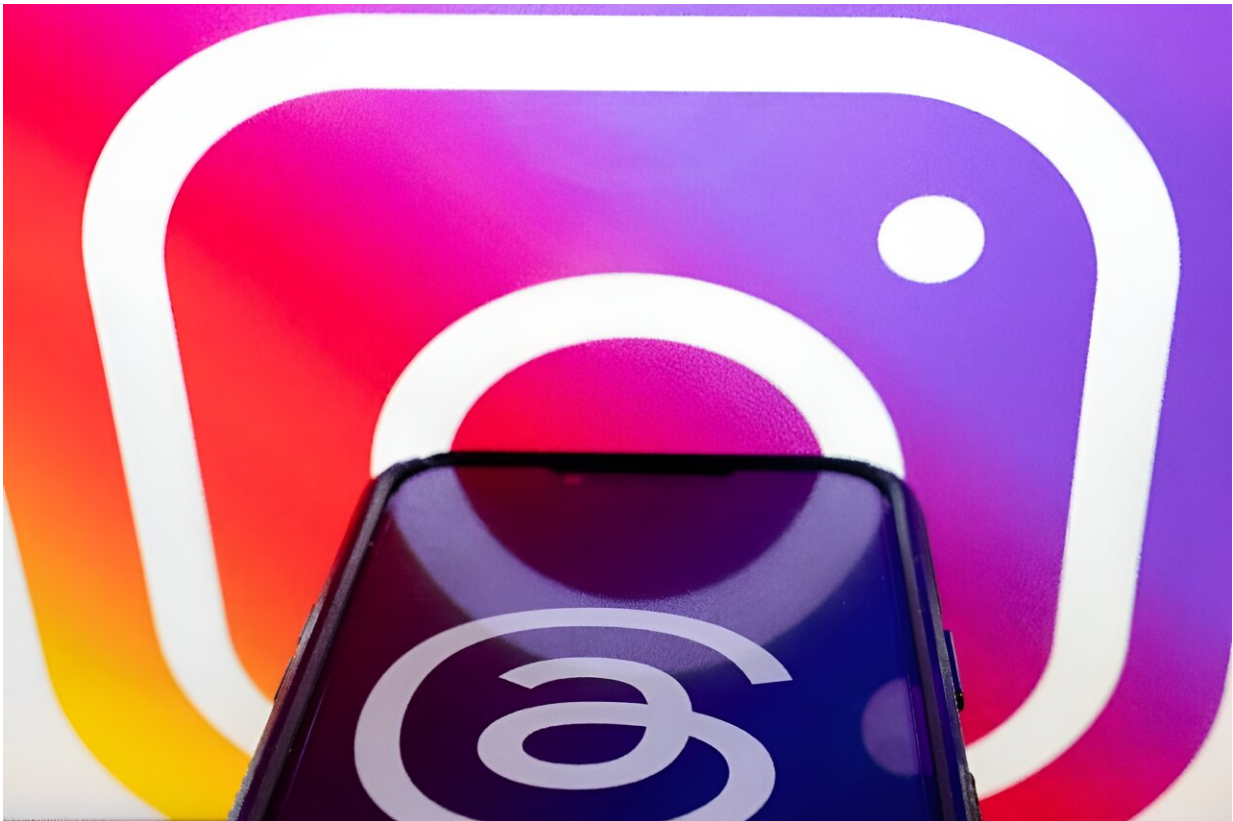# Meta wants industry-wide labels for AI-made images

February 6 2024, by Glenn CHAPMAN

As tech firms strive for ways to automatically detect images, video or audio altered or created using artificial intelligence tools, Meta is also urging internet users to look at posts critically to make sure what they are seeing isn't bogus.

Meta on Tuesday said it is working with other tech firms on standards that will let it better detect and label artificial intelligence-generated

images shared with its billions of users.

The Silicon Valley social media titan expects to have a system in place in a matter of months to identify and tag AI created images posted on its Facebook, Instagram and Threads platforms.

Meta and other platforms are under pressure to keep tabs on AI-generated content with fears that bad actors will ramp up disinformation, with elections due this year in countries representing half the world's population.

"It's not perfect, it's not going to cover everything; the technology is not fully matured," Meta head of global affairs Nick Clegg told AFP.

While Meta has implemented visible and invisible tags on images created using its own AI tools since December, it also wants to work with other companies "to maximize the transparency the users have," Clegg added.

"That's why we've been working with industry partners to align on common technical standards that signal when a piece of content has been created using AI," the company said in a blog post.

This will be done with companies Meta already works with on AI standards, including OpenAI, Google, Microsoft, Midjourney and other firms involved in the fierce race to lead the nascent sector, Clegg said.

But while companies have started including "signals" in images made using their AI tools, the industry has been slower to start putting such identifying markers into audio or video created with AI, according to Clegg.

Clegg admits that this large-scale labeling, using invisible markers, "won't totally eliminate" the risk of false images being produced, but

argues that "it would certainly minimize" their proliferation "within the limits of what technology currently allows."

In the meantime, Meta advised people to look at online content critically, checking whether accounts posting it are trustworthy and looking for details that look or sound unnatural.

Politicians and women have been prime targets for so-called "deepfake" images, with AI-created nudes of superstar singer Taylor Swift recently going viral on X, formerly Twitter.

The rise of generative AI has raised fears that people could use ChatGPT and other platforms to sow political chaos via disinformation or AI clones.

OpenAI last month announced it would "prohibit any use of our platform by political organizations or individuals."

Meta already asks that advertisers disclose when AI is used to create or alter imagery or audio in political ads.

The company's Oversight Board, which independently reviews content moderation decisions, on Monday warned that Meta's policy on deepfake content is in urgent need of updating.

The warning was in a decision about a manipulated video of US President Joe Biden that was not created with AI.

The Board said that Meta's policy in its current form was "incoherent, lacking in persuasive justification and inappropriately focused on how content has been created."

© 2024 AFP