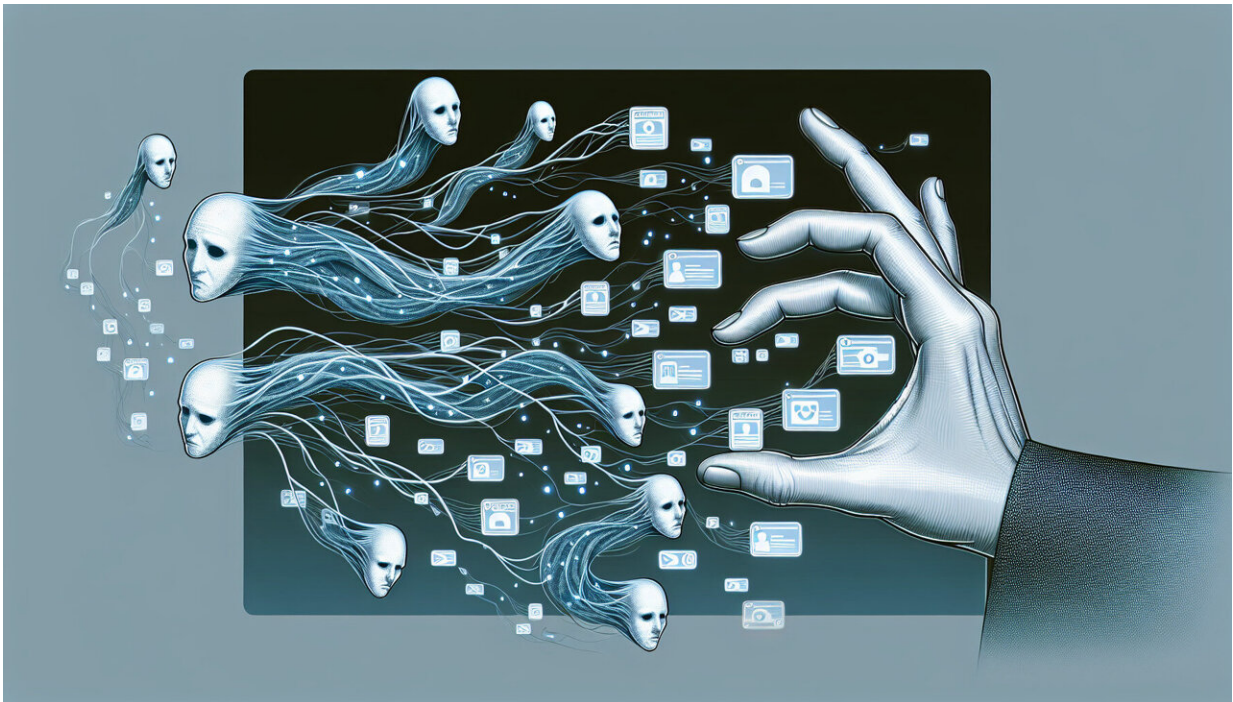


Researchers analyze the characteristics of AI-generated deepfakes

May 24 2024



Credit: AI-generated image

Most of the deepfakes (videos with fake hyper-realistic recreations) generated by artificial intelligence (AI) that spread through social media feature political representatives and artists and are often linked to current news cycles.

This is one of the conclusions of research by the Universidad Carlos III de Madrid (UC3M) that analyzes the formal and content characteristics of viral misinformation in Spain arising from the use of AI tools for illicit purposes. This advance represents a step towards understanding and mitigating the threats generated by hoaxes in our society.

In the [study](#), recently published in the journal *Observatorio (OBS*)*, the research team studied this fake content through the verifications of Spanish fact-checking organizations, such as EFE Verifica, Maldita, Newtral and Verifica RTVE.

"The objective was to identify a series of common patterns and characteristics in these viral deepfakes, provide some clues for their identification and make some proposals for [media literacy](#) so that citizens can tackle misinformation," explains one of the authors, Raquel Ruiz Incertis, a researcher in UC3M's Communication Department, where she is pursuing a Ph.D. in European communication.

The researchers have developed a typology of deepfakes, which makes it easier to identify and neutralize them. According to the results of the study, some [political leaders](#) (such as Trump or Macron) were the main protagonists of content referring to drug use or morally reprehensible activities. There is also a considerable proportion of pornographic deepfakes that harm women's integrity, particularly exposing famous singers and actresses. They are generally shared from unofficial accounts and spread quickly via instant messaging services, the researchers say.

The proliferation of deepfakes, or the frequent use of images, videos or audio manipulated with AI tools, is a highly topical issue. "This type of prefabricated hoax is especially harmful in sensitive situations, such as in pre-election periods or in times of conflict like the ones we are currently experiencing in Ukraine or Gaza. This is what we call 'hybrid wars': the war is not only fought in the physical realm, but also in the digital realm,

and the falsehoods are more significant than ever," says Ruiz Incertis.

The applications of this research are diverse, from national security to the integrity of election campaigns. The findings suggest that the proactive use of AI on [social media platforms](#) could revolutionize the way we maintain the authenticity of information in the digital age.

The research highlights the need for greater media literacy and proposes educational strategies to improve the public's ability to discern between real and manipulated content. "Many of these deepfakes can be identified through reverse image searches on search engines such as Google or Bing. There are tools for the public to check the accuracy of content in a couple of clicks before spreading content of dubious origin. The key is to teach them how to do it," says Ruiz Incertis.

It also provides other tips for detecting deepfakes, such as paying attention to the sharpness of the edges of the elements and the definition of the image background: if the movements are slowed down in the videos or whether there is any facial alteration, body disproportion or strange play of light and shadows, everything indicates that it could be AI-generated content.

In addition, the study's authors also see the need for legislation that obliges platforms, applications and programs (such as Midjourney or Dall-e) to establish a "watermark" that identifies them and allows the user to know at a glance that the image or video has been modified or created entirely with AI.

The research team has used a multidisciplinary approach, combining data science and qualitative analysis, to examine how fact-checking organizations apply AI in their operations. The main methodology is a content analysis of around thirty publications taken from the websites of the aforementioned fact-checkers where this AI-manipulated or

manufactured content is disproved.

More information: Miriam Garriga et al, Artificial intelligence, disinformation and media literacy proposals around deepfakes, *Observatorio (OBS*)* (2024). [DOI: 10.15847/obsOBS18520242445](https://doi.org/10.15847/obsOBS18520242445)

Provided by Carlos III University of Madrid

Citation: Researchers analyze the characteristics of AI-generated deepfakes (2024, May 24) retrieved 17 June 2024 from <https://techxplore.com/news/2024-05-characteristics-ai-generated-deepfakes.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.