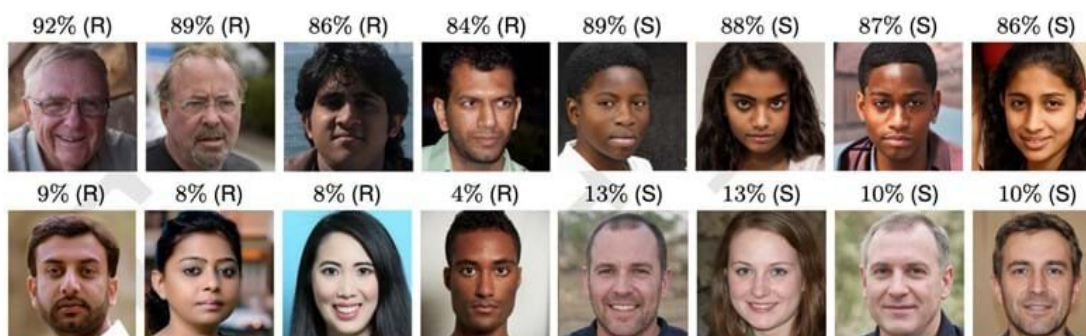# AI-generated faces found more trustworthy than real faces: Researchers warn of 'deep fakes'

February 16 2022

The most (top row) and least (bottom row) accurately classified real (R) and synthetic (S) faces. Credit: DOI: 10.1073/pnas.2120481119

People cannot distinguish between a face generated by Artificial Intelligence—using StyleGAN2—and a real face say researchers, who are calling for safeguards to prevent "deep fakes."

AI-synthesized text, audio, image, and video have already been used for so-called "revenge porn," fraud and propaganda.

Dr. Sophie Nightingale from Lancaster University and Professor Hany Farid from the University of California, Berkeley, conducted

experiments in which participants were asked to distinguish state of the art StyleGAN2 synthesized faces from real faces and what level of trust the faces evoked.

The results revealed that synthetically generated faces are not only highly photo realistic, but nearly indistinguishable from real faces and are even judged to be more trustworthy.

Writing in the journal, they say that: "Our evaluation of the photo realism of AI-synthesized faces indicates that synthesis engines have passed through the uncanny valley and are capable of creating faces that are indistinguishable—and more trustworthy—than real faces."

The researchers warn of the implications of people's inability to identify AI-generated images.

"Perhaps most pernicious is the consequence that in a digital world in which any image or video can be faked, the authenticity of any inconvenient or unwelcome recording can be called into question."

- In the first experiment, 315 participants classified 128 faces taken from a set of 800 as either real or synthesized. Their accuracy rate was 48%, close to a chance performance of 50%.
- In a second experiment, 219 new participants were trained and given feedback on how to classify faces. They classified 128 faces taken from the same set of 800 faces as in the first experiment—but despite their training, the [accuracy rate](link) only improved to 59%.

The researchers decided to find out if perceptions of trustworthiness could help people identify artificial images.

"Faces provide a rich source of information, with exposure of just

milliseconds sufficient to make implicit inferences about individual traits such as trustworthiness. We wondered if synthetic faces activate the same judgements of trustworthiness. If not, then a perception of trustworthiness could help distinguish real from synthetic faces."

A third study asked 223 participants to rate the trustworthiness of 128 faces taken the same set of 800 faces on a scale of 1 (very untrustworthy) to 7 (very trustworthy).

The average rating for synthetic faces was 7.7% MORE trustworthy than the average rating for real faces which is statistically significant.

"Perhaps most interestingly, we find that synthetically-generated faces are more trustworthy than real faces."

Black faces were rated as more trustworthy than South Asian faces but otherwise there was no effect across race.

Women were rated as significantly more trustworthy than men.

"A smiling face is more likely to be rated as trustworthy, but 65.5% of the real faces and 58.8% of synthetic faces are smiling, so facial expression alone cannot explain why synthetic faces are rated as more trustworthy."

The researchers suggest that synthesized faces may be considered more trustworthy because they resemble average faces—which themselves are deemed more trustworthy.

To protect the public from "deep fakes," they also proposed guidelines for the creation and distribution of synthesized images.

"Safeguards could include, for example, incorporating robust

watermarks into the image- and video-synthesis networks that would provide a downstream mechanism for reliable identification. Because it is the democratization of access to this powerful technology that poses the most significant threat, we also encourage reconsideration of the often-laissez-faire approach to the public and unrestricted releasing of code for anyone to incorporate into any application.

"At this pivotal moment, and as other scientific and engineering fields have done, we encourage the graphics and vision community to develop guidelines for the creation and distribution of synthetic-media technologies that incorporate ethical guidelines for researchers, publishers, and media distributors."