Deepfake smiles matter less—the psychological and neural impact of presumed AI-generated faces

November 27 2023

How do we distinguish between real and generated images of human emotions, and how can scientific research understand the ways in which our brains differentiate between real and artificially created emotional cues? Credit: Image generated via Jasper Illustrator by Anna Eiserbeck as a response to the prompt "Smiling face, vector art"

In our digital time, where artificial intelligence (AI) crafts deceptively realistic human faces, the emergence of deepfake technology may blur the boundaries between reality and digital fabrication. These AI-generated faces, though technologically astounding, carry a weight of societal implications that demand a thorough examination.

A recent study published in *Scientific Reports* and conducted by Science of Intelligence (SCIoI) scientists Anna Eiserbeck, Martin Maier, Julia Baum, and Rasha Abdel Rahman, delves into the psychological and neural repercussions tied to the perception of AI-generated faces, especially focusing on the emotional expressions they portray.

Products of generative artificial intelligence, such as deceptively real-looking photos and videos of people, known as deepfakes, are becoming increasingly common. Until now, however, it was unclear how knowing that a face might or might not be real affects how we perceive it and respond to it emotionally.

In their new study, the researchers analyzed facial expression ratings and brain responses to smiling, angry, and neutral faces that participants assumed were either real or computer-generated. The results show that a computer-generated smile matters less to us on several levels: it is perceived as less intense, elicits a weaker emotional response in the brain, and appears to give us pause.

Angry faces, on the other hand, remain equally threatening, whether we believe them to be genuine or not. These fundamental new findings have implications for how we as a society will deal with deepfakes, both when they are used for good and for ill.

Deepfakes and the human brain: A study of perception and emotional evaluation

The study, involving 30 participants and utilizing EEG technology, explored the effects of the belief that a portrayed individual is either real or deepfake on psychological and neural measures of face perception. In the words of SCIoI researcher Martin Maier, "When confronted with smiling faces marked as deepfakes, participants showed reduced perceptual and emotional responses, and a slower evaluative process as opposed to when the faces were marked real. Intriguingly, this impact was not mirrored in the perception of negative expressions, which remained consistent whether believed to be real or fake."

The findings highlight a complex interplay of emotional valence and presumed authenticity, and mark the first time a distinction has been drawn in the psychological impact between positive and negative expressions portrayed by deepfakes. "When real and fake faces are otherwise indistinguishable, perception and emotional responses may crucially depend on the prior belief that what you are seeing is, in fact, real or fake," added Rasha Abdel Rahman, principal investigator of the study.

In order to reach these conclusions, the researchers looked at how the brain's response to images of faces evolves over time, focusing on three stages: early visual perception (up to 200 milliseconds after a face was shown, before we are even aware of seeing it), reflexive emotional processing (at 200—350 milliseconds, reflecting our immediate emotional reactions), and higher-level evaluative processing (at 350 milliseconds and later, marking a more thoughtful consideration).

They used a method called Event-Related Potentials (ERPs) to track these stages. The findings showed that when people looked at smiles they thought were created by deepfake technology, their typical early visual and emotional responses were weaker. Understanding this has direct implications for different situations in which we may encounter deepfakes: when used, for instance, to bring back younger versions of

movie characters, the hope is that the emotional expressions of artificially generated characters appear just as lively and genuine as a real actor.

In these situations, the study results suggest, knowing that the character is artificially generated may compromise its impact, especially for positive emotions. When used in the context of misinformation campaigns, the results suggest that artificially generated negative contents may stick, even though observers may suspect that the images are fake.

Implications and future directions in deepfake technology

The findings of this study serve as a cornerstone in understanding the behavioral and neural dynamics of human interaction with AI-generated faces, while the discoveries underscore the necessity for a nuanced approach in devising policies and strategies to navigate the growing sphere of deepfake technology. The results also provide a starting point for further explorations into other domains of AI-generated content such as text, visual art, or music.

As deepfake technology continues to evolve, nurturing a profound understanding of its psychological and neural impact becomes central to both optimally using its potential to benefit society, and fortify societal resilience against the various challenges it poses.