DON'T BELIEVE EVERYTHING YOU SEE

The term *deepfake* refers to synthetic media created with deep learning, a subset of machine learning (which itself is a subcategory of AI) that processes unstructured data and automates feature extraction. Popular deepfakes often feature celebrity faces transposed onto digital content to create realistic but fake new media (see @deeptomcruise on TikTok).

Luckily, news services are alert for these scams. A large amount of imagery and computing power is required to generate a deepfake. This makes them mainly inaccessible to the average person—beyond the use of filters on social media photos.

In June 2020, OpenAI, a San Francisco research lab cofounded by Elon Musk, launched Generative Pretrained Transformer 3 (GPT-3), a machine learning algorithm for natural language processing that teaches a computer to process language. The system was trained using a massive amount of sample texts and computing power. It was presented with a sentence and then asked to predict the next word. Afterward, the accuracy of its prediction was evaluated. The predictions were senseless at first but improved over time until a language model was developed that can generate text.

GPT-3 can create beautiful news stories, essays, and poetry. Soon after its release, interesting extensions emerged. For example, automatic website generators appeared with which users can describe in writing what their webpages should look like, and the system generates the corresponding code.

DALL-E 2 (OpenAI) is built on GPT-3 but is supplemented by a network that can recognize objects in photos, allowing it to put a name to what it sees. An image of virtual seats in the shape of an avocado that was created using DALL-E 2 went viral in 2021.[1] The system is not a glorified search engine. It doesn't produce existing photos in response to a description. Instead, it generates a new image.

The name DALL-E 2 refers to the possible surreality of the images the system produces. The name is an amalgam of the Spanish surrealist artist Salvador Dalí and the Walt Disney Pictures/Pixar Animation Studios' cartoon robot character Wall-E from the film of the same name. The more abstract the question, the more interesting the result—a Dalí-style painting of Einstein taking a selfie, a teddy bear on a skateboard in Times Square, or sneakers decorated with Frida Kahlo-esque motifs. The user asks; DALLE-2 delivers. Unfortunately, its capability could be easily abused.

In May 2022, Matt Bell, an American blogger who writes on science, technology, culture, and other topics, demonstrated the ease with which the human brain can be fooled by images when he conducted the Turing test on his followers.[2] He posted a series of photos from a diving vacation to his Facebook page, but the last four pictures in the post were created with DALL-E 2. Later, Bell surveyed his followers and found that 83% of respondents had cheerfully reacted to the post with likes, hearts, and wow faces without detecting anything unusual about the photos. The experiment showed how easily images, especially online, can trick the human brain.

DALLE-2 will likely become a source of inspiration for artists and designers. They do not have to fear for their jobs in the immediate future because interesting output still requires creative input. Bell's use of DALLE-2, however, highlights its potential to be misused. You have been warned. Don't always believe what you see. ■

[1] Heaven WD. This avocado armchair could be the future of AI. *MIT Technology Review*. January 5, 2021. Accessed November 8, 2022. https://www.technologyreview.com/2021/01/05/1015754/avocado-armchair-future-ai-openai-deep-learning-nlp-gpt3-computer-vision-common-sense

[2] Bell M. My deepfake DALL-E 2 vacation photos passed the Turing test. May 12, 2022. Accessed November 7, 2022. https://www.mattbell.us/my-fake-dall-e-2-vacation-photos-passed-the-turing-test

ERIK L. MERTENS, MD, FEBOPHTH | CHIEF MEDICAL EDITOR

Physician CEO, Medipolis-Antwerp Private Clinic, Antwerp, Belgium