

Detecting Deepfakes Through Personality Assessment

Originally published at <https://metaphysic.ai/detecting-deepfakes-through-personality-assessment/>

August 15, 2022



[Create your Hyperreal AI-generated True Self today!](#)

In the original, and the various cinematic incarnations of Jack Finney's 1954 science-fiction novel *The Body Snatchers*, the fact that aliens are 'taking over' the earthly population is signaled to their loved ones by some imperceptible changed quality that cuts through the perfect simulation of their form.

This instinctive rejection of the mannerisms of a 'faked' person, or one who is not behaving as they normally do, is one of the possible roads forward to developing reliable deepfake detection systems that cannot easily be outpaced by new innovations and improvements in facial synthesis.

In cases where we are familiar with the person (i.e. a celebrity or a loved one), we can expect to have a high degree of 'instinctive' familiarity with their mannerisms – something that is challenging to quantify and train into a machine learning model, or a new deepfake detector algorithm. However, interest in this approach is growing.

Will the Real Donald Trump Stand Out?

A [new research collaboration](#) between researchers in the US and Canada offers hope that such 'behavioral signatures' could eventually prove to be a resilient algorithmic method for determining the authenticity of deepfakes that are designed to influence or subvert political or popular opinion. Titled *Study of detecting behavioral signatures within DeepFake videos*, the new work conducts a study to determine whether ordinary people who are familiar with a well-known person – in this case, Donald Trump – will be able to pick out videos where his behavior has been subtly altered, and where either his speech or his mannerisms (or both) have been artificially generated.



In the experiments for the new project, faked videos of Donald Trump were generated based on his own utterances, the speech of others, and his own words as performed by actors. A study group was challenged to distinguish the 'real' Donald Trump. Source: <https://arxiv.org/ftp/arxiv/papers/2208/2208.03561.pdf>

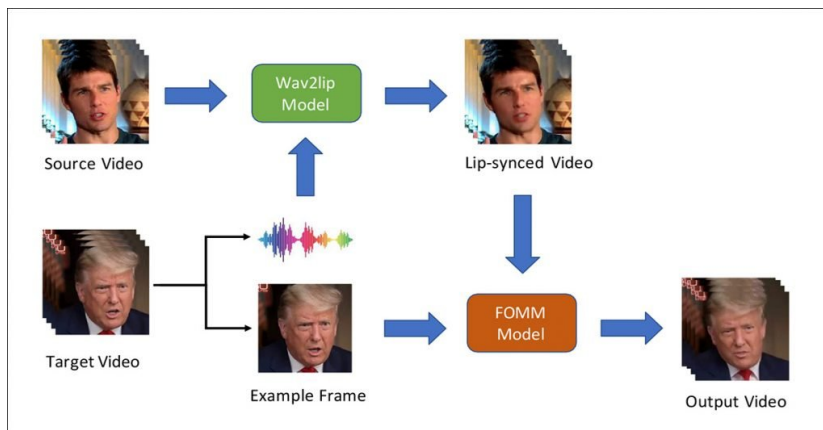
Considering that experts have [recently warned](#) that the 2022 Volodymyr Zelenskyy [deepfake](#) could be 'the tip of the iceberg' in a climate which has feared political deepfakes [for years](#), this strand of research could prove critical in terms of maintaining credibility in the news arena.

In the case of the new study, the good news is that the participants were able, in the majority of cases, to recognize 'genuine' video of Donald Trump speaking, even though all the videos from which they were asked to choose featured high-quality Trump simulations, some using non-congruous speech from personalities such as Tom Cruise, Barack Obama, Taylor Swift, and Emma Watson – or else Trump's own words reperformed by hired actors, and fed back into the original footage, retaining the words but subtly altering the mannerisms.

Let's take a look at how the tests were formulated.

The Trump Triumvirate

Since the objective of the study was to see if people could identify the 'real' Trump footage, it was important to ensure a level playing field. Thus, the 'genuine' video presented in the trials was actually an exact neural recreation of the source footage, with no alien mannerisms imposed, but designed to exactly match the quality of the 'adulterated' clips.



Conceptual architecture for the new paper's method.

For the other clips, the [Wav2Lip model](#) was used to interpose lip-synced simulated speech into the adulterated clips, while First Order Motion Model ([FOMM](#), see embedded video below) provided the actual face reenactment.

<https://www.youtube.com/watch?v=u-0cQ-grXBQ>

Test 1 – ‘Invasion of the Speech-Snatchers’

In the first deepfaked clip generated with these frameworks, the researchers used different speakers to put words into Trump’s mouth, taking care to select phrases that would not undermine the plausibility of the video. Here, phrases from Obama, Cruise, Watson and Swift were interposed via voice synthesis.

Test 2 – ‘As I said recently...’

For this test, all the words that Trump speaks were his own, taken directly from the video source material – but transplanted into a *different clip* of Trump.

The primary difference between the original and generated video in this case is that the head movements that the viewer sees were intended to support a different utterance.

Test 3 – ‘As others have said...’

For this test, voice actors were hired to exactly reproduce the speech of Donald Trump, carefully matching their mouthing of his words to the actual movement of the former president’s lips in the target clip. In order not to obtain results that might have been visually different from the other clips, the actors’ recitals were nonetheless incorporated into the fake video via Wav2Lip, as with the other examples, with the video component again provided by FOMM. Since no vocal content substitution occurred, no voice synthesis was necessary.

The objective of this third approach was to see if the public would recognize that the actors’ ‘alien’ head movements and mannerisms were not actually Trump’s own affectations, in the simulated video.

Create your Hyperreal AI-generated True Self today!

The Study

For the study, the hired participants were first asked if they were familiar with Donald Trump, and had seen him talking before. Those who confirmed both cases were then invited to view the neural ‘real’ video, together with single examples of fake videos from the three different approaches, and invited to answer the following questions:

1. *How much is the person in the video (real/synthetic) like Trump?*

Rating question: 1-7 (1: not like him, 7: exactly like him)

2. *Which person in the above two videos looks more natural?*

3. *Which person in the above two videos are you more engaged with?*

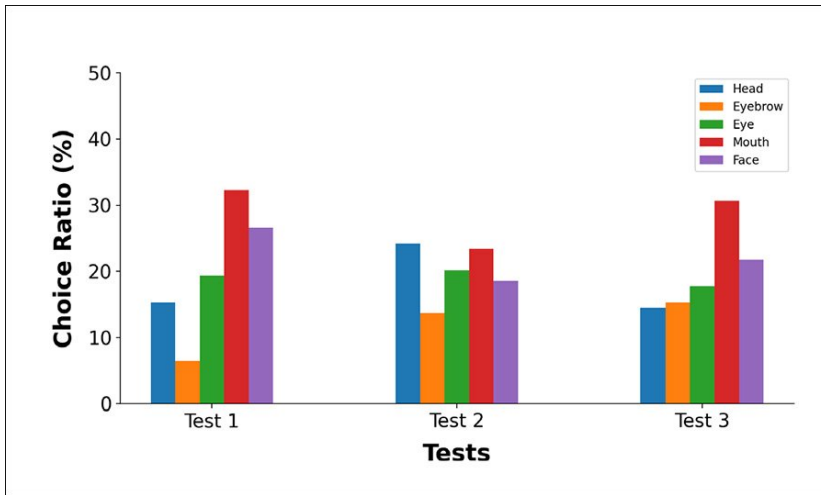
4. *What movement is most important for you to get engaged with the person that you chose in the last question?*

Available choices: *head movements, eyebrow movements, eye movements, mouth movements, facial expressions, other.*

In addition to the ‘genuine’ video, each participant was shown an additional video from the pool of faked clips. Therefore each participant was exposed to six videos, which they were required to rate.

Results

For the ‘speech snatchers’ test (where the words of other popular personalities were put into Trump’s mouth), the majority of the participants, around 60%, showed a preference for the original Trump video, in terms of naturalness and engagement, and the ‘true’ video was generally rated higher.



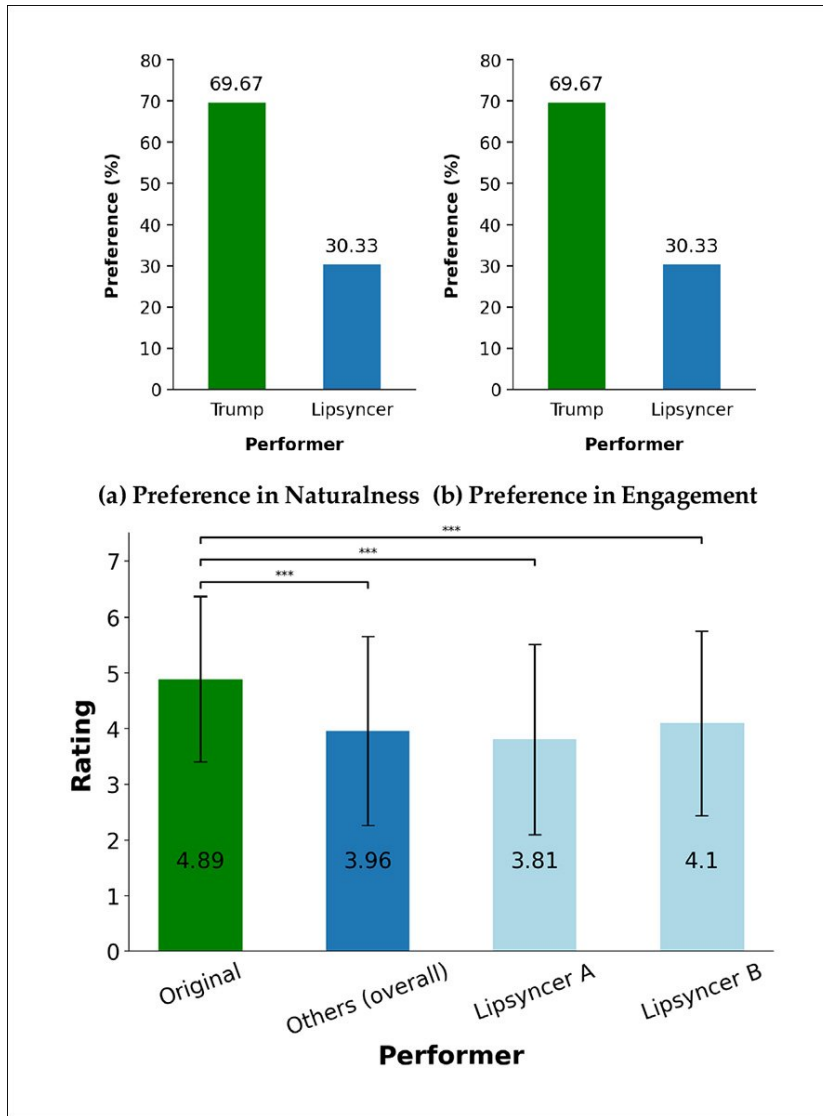
A visualization of the importance of specific movements assigned by the study participants across the range of videos shown them. In most cases, mouth movement is assigned the most importance, with eyebrow movement and general head movement among the lowest considerations.

For the second test, where Trump's own words are transposed into a different clip of himself, the participants were still able to detect that the 'real' clip truly reflected Trump. Most were influenced by the non-appropriate head movements that accompanied the transposed speech, which were originally intended to support a different contention.



Trump's head movements for transposed speech were recognized by the study group as inauthentic to the content. The two right-most columns are from simulations where the speech is not original.

For the third and final test, where actors reproduce Trump's own words, but where their own facial and head mannerisms are interposed into the fake video, the original and authentic video still obtained a 70% preference from the user study.



Trump beats the lip-syncers in the third and final round of tests, but, as with recent US elections, it's not a landslide victory.

For these tests, 'mouth' and 'facial expressions' were the dominant factors in the participants' choices.

The researchers state:

'Our study indicates that both the speaking behavior style and the correspondence between speaking behavior and utterance play vital roles in the non-appearance aspect in the identification of a person.'

'This provides evidence that the distinct speaking style of a person and the correspondence between speaking behavior and utterance can serve as important clues for DeepFake detection even for synthetic video with perfect visual appearance in the future.'

Analyzing which factors play into user choices, the researchers conclude that head movement plays a big role in whether the viewer correlates what is being said to the way that it's being said, while facial expression and mouth movement tend to hallmark different speakers in general.

The researchers conclude:

'The results of our study show that both behavioral signature and correspondences with utterance can significantly affect humans' judgements of the naturalness of a video.'

'This provides evidence of the necessity for leveraging behavioral signature and the correspondence between behavior and utterance in Deepfake Detection, which are overlooked by models that examine visual quality alone.'

Indications for the Future

It's worth considering that though the original clips in each test case lead the boards, they don't lead by an overwhelming margin. A significant percentage of the study group was successfully fooled by the adulterated clips. In terms of political manipulation, which frequently aims itself at marginal or swing voters, that's a [notable proportion](#).

A behavior-based deepfake detection system would need a number of central innovations involving expression recognition, and perhaps a form of upper-body movement recognition that concentrates on the head in order to obtain 'signature' movements of an individual.

Head movement by itself [conveys emotional state](#) quite [separately](#) from facial expressions, and the results of the new paper's experiments seem to indicate that disjoint between such movements and speech content could be a useful long-term identifier for deepfaked or adulterated content. Though [autism](#) and [anxiety](#), among other conditions, can be recognized through head pose evaluation, and increased effort in speaking is also [associated](#) with more vigorous head movements, other facets of identity have come to dominate the attention of the computer vision research sector – especially facial recognition.

However, since a computer user's mouse movements alone can [reveal the identity](#) of an individual, it would seem possible in theory that characteristic head movement patterns could be recorded for prominent figures (such as politicians) considered vulnerable to deepfake attacks, as another line of defense against AI-facilitated impersonation.

Create your Hyperreal AI-generated True Self today!