HOW DEEP-FAKES CAN PROVE FATAL TO THE WORLD

Raunakk Jalan, Shubham Agarwal

Deep fakes are synthetic images or videos created in a manner such that it is very difficult to segregate it from a real one. Though fake images and videos have been on the internet for very long, the fact which makes deep fames stand out is that they use new technologies such as artificial intelligence and machine learning to create very deceptive media. Deep fakes have been used for swapping the faces of big personalities like Barack Obama where in a video he appeared to be cursing about Donald Trump. New technology has enabled even amateurs to create very convincing deep fakes. The threat of deep fakes is on the rise with fake news flooding the internet. Many many cases of defamation using deep fakes have occurred and it is becoming really hard for us to distinguish between what is true and what is not. Falsified videos created by AI—in particular, by deep neural networks (DNNs)— are a recent twist to the disconcerting problem of online disinformation.

Although manipulation of images and videos and other online material is common to us by now, the rapid development of AI technology in recent years has made the process to create deepfake videos much easier and faster. The technology of making deepfakes is at the disposal of ordinary users; there are quite a few software tools freely available on GitHub, including FakeApp, DFaker, faceswap-GAN, faceswap and DeepFaceLab—so it’s not hard to imagine the technology could be used in political campaigns, election seasons, election engineering, foreign interference in political democratic processes and other important events. However the proliferation of deep-fakes depends on many factors. One factor is cost. Creating deepfakes, although much easier than ever before, still requires time, resources and skill. Compared to other, cheaper approaches to disinformation (e.g., repurposing an existing image or video to a different context), deepfakes are still an expensive and inefficient technology. Another factor is that deepfake videos can usually be easily exposed by cross source fact-checking, and are thus unable to create long-lasting effects. But needless to say that we should be prudent and alert when it comes to deepfake videos as they can be used in an extensive disinformation campaigns, or deployed at a particular time (e.g., within a few hours of voting) to cause short term chaos and confusions.

However, while there are ways to curb the effects posed by deepfakes, they pose an existential threat to our democratic systems. Imagine deep-fakes being generated of Donald Trump asking people to continue taking hydroxychloroquine or to stop wearing masks (although he would probably do that himself) ; this would create mass chaos and lead to exponential growth of the anti- mask movement that has already become a pandemic by itself today. Deepfakes can not only have long lasting effects on politics but also on the fields of journalism, science , research, etc where any wrong statement is enough to generate waves of confusion and problems in the world. To curb the threat posed by deepfakes, detection technology will also need to keep up and adapt. Progress should be made to further develop technologies like video compression, social media laundering and other common post-processing operations, as well as intentional counter-forensics operations.

However, given the incredible propagation speed and reach of online media, even the most effective detection method will largely operate in a postmortem manner, applicable only after deepfake videos emerge and have had their nefarious effect on their targeted sphere. Although the pandora’s box has been opened by recent technological improvements, it is quintessential that technology is developed to recognise and take down deep-fakes within moments. AI applications are boundless in recognising deep-fakes and definitely need to be improved upon to ensure deepfakes do not have the opportunity to cause mass panic and hysteria in our political, economic, scientific, or social spheres.