ITMS
INFORMATION SYSTEMS SECURITY DIVISION
ISSD
PHILIPPINE NATIONAL POLICE
INFORMATION TECHNOLOGY MANAGEMENT SERVICE
INFORMATION SYSTEMS SECURITY DIVISION

Deepfake
(CSB20-07)

Deepfake technology enables anyone with a computer and an Internet connection to create realistic-looking photos and videos of people saying and doing things that they did not actually say or do.

The first use case to which deepfake technology has been widely applied is pornography. A handful of websites dedicated specifically to deepfake pornography have emerged, collectively garnering hundreds of millions of views over the past two years. Deepfake pornography is almost always non-consensual, involving the artificial synthesis of explicit videos that feature famous celebrities or personal contacts.

From these dark corners of the web, the use of deepfakes has begun to spread to the political sphere, where the potential for confusion is even greater.

HOW ARE THEY MADE

A generative adversarial network (Gan) pits two AI algorithms against each other. Starting with a given dataset (a collection of photos of human faces), the generator begins generating new images that, in terms of pixels, are mathematically similar to the existing images. Meanwhile the discriminator is fed photos without being told whether they are from the original dataset or from the generator's output; its task is to identify which photos have been synthetically generated.

HOW TO IDENTIFY A DEEPFAKE

There are different types and levels of deepfakes and certain tools needed to discern a deepfake from a real live selfie. Some deepfakes are coarser and lower quality, and they can be quickly produced with free apps. More convincing or higher-quality deepfakes require more significant effort, skill, money and time.
The human eye can detect deepfakes by closely observing the images for slight imperfections such as:

- Face discolorations
- Lighting that isn't quite right
- Badly synced sound and video
- Blurriness where the face meets the neck and hair

Algorithms can detect deepfakes by analyzing the images and revealing small inconsistencies between pixels, coloring, or distortion. It's also possible to use AI to detect deepfakes by training a neural network to spot changes in facial images that have been artificially altered by software. The most robust forms of liveness detection rely on machine learning, AI, and computer vision to examine dozens of miniscule details from a selfie video such as hair and skin texture, micromovements, and reflections in a subject's eye.

IMPACT

- Distorting democratic discourse;
- Manipulating elections;
- Eroding trust in organizations;
- Weakening journalism;
- Worsening social divisions;
- Undermining public safety;
- Inflicting hard-to-repair damage on the reputation of prominent individuals.

REFERENCE

- https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/#2ddfdd657494;
- https://www.darkreading.com/endpoint/authentication/the-rise-of-deepfakes-and-what-that-means-for-identity-fraud/a/d-id/1337633