Deepfake

Posted at: 16/05/2024

Context:

Deepfakes has emerged as a major concern during election season.

Background:

AI-powered misinformation campaigns can influence voter behaviour by spreading false narratives or amplifying divisive content.

About Deepfake :

1. Deepfakes are a type of synthetic media where one person's likeness is convincingly replaced with that of another. The term "deepfake" is a portmanteau of "deep learning" and "fake"
2. A deepfake is an artificial image, audio or video generated by a special kind of machine learning called "deep" learning (hence the name)
3. Deepfake technology leverages tools and techniques from machine learning and artificial intelligence, including facial recognition algorithms andartificial neural networks such as variational autoencoders (VAEs) and generative adversarial networks (GANs).
4. It is used to manipulate videos, images, and audios.
5. This technology can be used to generate fake news and commit financial fraud among other wrongdoings.It overlays a digital composite over an already-existing video, picture, or audio.
6. Deepfake technology can seamlessly stitch anyone in the world into a video or photo they never actually participated in.

How does deepfake technology work?

1. The technology involves modifying or creating images and videos using a machine learning technique called generative adversarial network (GAN).
2. The AI-driven software detects and learns the subjects' movements and facial expressions from the source material and then duplicates these in another video or image.
3. To ensure that the deepfake created is as close to real as possible, creators use a large database of source images. This is why more deepfake videos are created of public figures, celebrities and politicians.
4. The dataset is then used by one software to create a fake video, while a second software is used to detect signs of forgery in it.
5. Through the collaborative work of the two software, the fake video is rendered until the second software package can no longer detect the forgery. This is known as "unsupervised learning", when machine-language models teach themselves. The method makes it difficult for other software to identify deepfakes.

Examples:

1. In January 2024, during the New Hampshire primary of the Democratic Party in the US, a robocall mimicking President Joe Biden’s voice falsely advised voters not to participate, claiming it would affect their eligibility for the general election.
2. In Slovakia, an AI-generated voice, mimicking that of a liberal candidate, discussing plans to raise alcohol prices and rig the election was widely circulated on Facebook.
3. In Bangladesh, deepfake videos of opposition politicians Rumin Farhana in a bikini and Nipun Roy in a swimming pool surfaced on social media ahead of the national elections.