Chapter 10
Deepfake Technology and the Quagmire of Artificial Intelligence:
An Analysis of National and International Legal Frameworks

Pallavi Tiwari
https://orcid.org/0009-0004-3050-8531
Maharashtra National Law University, Nagpur, India

Ujwal Prabhakar Nandekar
https://orcid.org/0000-0003-2545-6736
Symbiosis Centre for Advanced Legal Studies and Research (SCALSAR), Symbiosis Law School, Symbiosis International University, Pune, India

ABSTRACT

Artificial Intelligence has proved to be a boon and a bane in all sectors of society. It has raised concerns regarding the protection of individuals' right to privacy and personality rights, and issues related to the protection of the personal data of individuals. Deepfakes are combinations of fake videos or images of individuals which are morphed or distorted to produce unethical results. The present research aims to conduct an empirical and doctrinal study to look into the issue of deepfake technology, how the public perceives it, and how it is a sensitive issue for laymen in India and internationally. The study involves an empirical analysis involving stakeholders from different dimensions of society, including individuals from the age group of 18-55(+) years and can be from any domain affected by such usage of

DOI: 10.4018/979-8-3373-3063-1.ch010

deepfake pictures and videos. Further to this, with the help of a theoretical analysis with the aid of an analytical study of existing laws in India and the understanding of the international legal regime to protect against misuse of artificial intelligence.

INTRODUCTION

As per recent news, a gang defrauded SP Oswal, 'the chairman and managing director of Vardhman Group', of Rs 7 crore by pretending to be representatives of the CBI and the former Chief Justice of India (CJI), DY Chandrachud, in a fictitious online proceeding (Sharma, 2024). This is just one incident of violating rights and conducting fraudulent activities using Artificial Intelligence. AI is used in deepfakes to make photos and movies appear frightfully real. However, these videos are fake and can be used to disseminate false information with negative effects.

Deepfakes are artificial intelligence (AI)-generated forgeries or copies of extremely lifelike audio, video, or images. Machine learning (ML) and Generative Adversarial Networks (GANs) are two technologies that produce deepfakes. Systems may learn and get better from experience through the data they gather thanks to machine learning (ML), a subset of artificial intelligence (Hall, 2018). Facial expression manipulation and face swapping are popular in deepfake videos. Artificial neural networks, which are computer programs that identify patterns in data, are the foundation of deepfakes. Creating a deepfake image or video usually entails "training" an artificial neural network with hundreds or thousands of photos to recognise and recreate patterns, most often faces (Thombre, 2021).

The term "deepfakes," which combines the terms "deep learning" and "fake," was invented on Reddit in 2017 when AI replaced the faces of famous people in some videos (Somers, 2020). Article 3(60) of the European Union's Artificial Intelligence Act, which was formally adopted by the European Parliament on March 13, 2024, defines "deep fake" as artificially produced or transformed audio, video, or image output that would misleadingly appear genuine or true, and that bears similarities to real people, locations, things, or other occurrences or things (EU AI Act, 2024).

Two machine learning models are used by deepfakes. Using a collection of data from the available sample videos or photographs, one of those generates fakes by analysing how a face might blink, smile, or grin, among other ways a face can convey emotion. When the second machine learning model is unable to determine whether the available video or image of the targeted individual is a hoax, the deepfake-generated product is most likely sufficiently convincing to the human eye. The other machine learning model attempts to determine whether the available sample detected from an autoencoder is a hoax and detects fraud to the utmost capability. Significant