Deepfakes: Opportunities, Threats, and Regulation

This editorial is based on "Rashmika Mandanna's deepfake: Regulate AI, don't ban it" which was published in The Indian Express on 06/11/2023. It discusses the recent viral video of actor Rashmika Mandanna, which was revealed to be a deepfake, and the need for a holistic approach to the regulation of such technologies.

For Prelims : Deepfake, **Artificial Inteliigence (AI)**, IT Act, 2000 **and IT Rules**, Section 66D of IT Act**,**

For Mains: Deepfakes: Uses, Challenges, Rules set by the Government and Way Forward

Recently, a fact-checker webiste posted that a viral video of an actor entering a lift was a deepfake. The video sparked much debate, with other actors calling for the legal regulation of deepfake videos. In response, the Minister of State for Electronics and Information Technology (IT) talked about regulations under the IT Act, 2000, which could tackle the spread of such videos. However, a holistic approach to the regulation of deepfakes should focus on the interplay between platform and Artificial Inteliigence (AI) regulation, and ways to incorporate safeguards for emerging technologies more broadly.

What is Deepfake?

- **Deepfake** is a term that refers to synthetic media that have been digitally manipulated to replace one person's likeness convincingly with that of another.
- **Deepfakes** are created using powerful techniques from machine learning and AI, such as **deep learning** and **Generative Adversarial Networks (GANs).**
- Deepfake technology can be used for various purposes, such as entertainment, education, art, and activism.
  - However, **it can also pose serious ethical and social challenges**, such as creating fake news, spreading misinformation, violating privacy, and harming reputation.
  - It may be **used to generate fake videos**, it can also be used to impersonate friends or loved ones to trick individuals into sending money to scammers.

What are the Uses of Deepfake Technology?

- **Film Dubbing**: Deepfake technology can be used to create realistic lip-syncing for actors who speak different languages, **making the film more accessible and immersive for global audiences.**
  - For example, a video was created to launch a petition to end malaria, where celebrities like David Beckham, Hugh Jackman, and Bill Gates spoke in different languages using deepfake technology.
- **Education:** Deepfake technology can **help teachers deliver engaging lessons by bringing historical figures to life in the classroom**, or creating interactive simulations of different

scenarios.

    - For example, a deepfake video of Abraham Lincoln giving his Gettysburg Address could be used to teach students about the American Civil War.

- **Art**: Deepfake technology can be **used as a creative tool for artists to express themselves,** experiment with different styles, or collaborate with other artists.
    - For example, a deepfake video of Salvador Dali was created to promote his museum in Florida, where he interacted with visitors and commented on his artworks.
- **Autonomy and Expression**: Deepfake technology can empower people to control their own digital identity, protect their privacy, or express their identity in different ways.
    - For example, **a deepfake app called Reface** allows users to swap their faces with celebrities or characters in videos or gifs, for fun or personalization.
- **Amplification of the Message and its Reach**: Deepfake technology can help **amplify the voice and impact of people who have important messages to share**, especially those who face discrimination, censorship, or violence.
    - For example, a deepfake video of a journalist who was killed by the Saudi government was created to deliver his final message and call for justice.
- **Digital Reconstruction and Public Safety:** Deepfake technology can help **reconstruct missing or damaged digital data**, such as restoring old photos or videos, or enhancing low-quality footage.
    - It can also help **improve public safety by creating realistic training materials** for emergency responders, law enforcement, or military personnel.
    - For example, a deepfake video of a school shooting was created to train teachers on how to react in such a situation.
- **Innovation**: Deepfake technology can **spur innovation in various fields and industries, such as entertainment, gaming, or marketing**. It can enable new forms of storytelling, interaction, diagnosis, or persuasion.
    - For example, a deepfake video of Mark Zuckerberg was created to demonstrate the potential of synthetic media and its implications for society.

What are the Challenges of Deepfake Technology?

- **Spreading False information:** Deepfakes can be used to **purposefully spread false information or misinformation**, which can create confusion about important issues.
    - For example, deepfake videos of politicians or celebrities can be used to influence public opinion or sway elections.
- **Harassment and Intimidation:** Deepfakes can be **designed to harass, intimidate, demean, and undermine people.**
    - For example, deepfake technology can fuel other unethical actions like creating revenge porn, where women are disproportionately harmed.
        - Deepfake porn can also violate the privacy and consent of the victims, and cause psychological distress and trauma.
    - Deepfake technology **can be used to create blackmail or ransom materials**, such as fake videos of someone committing a crime, having an affair, or being in danger.
        - For example, a deepfake video of a politician was used to demand money in exchange for not releasing it to the public.
- **Fabricating Evidence**: Deepfakes can be **used to fabricate evidence**, which can be used to defraud the public or harm state security. Deepfake evidence can also be used to manipulate legal proceedings or investigations.
    - For example, deepfake audio or video can be used to impersonate someone’s identity or voice, and make false claims or accusations.
- **Reputation Tarnishing**: Deepfakes can be used to create an image of a person that does not exist, **creating a video of someone saying or doing** something they have never done, or synthesizing a person’s voice in an audio file, which can be used to tarnish someone’s reputation.
    - For example, deepfake media can be used to damage the credibility or trustworthiness of a person or an organisation, and cause reputational or financial losses.
- **Financial Frauds:** Deepfake technology can be used to impersonate executives, employees, or customers, and manipulate them into revealing sensitive information, transferring money, or making false decisions.
    - For example, a deepfake audio of a CEO was used to trick an employee into wiring USD

243,000 to a fraudulent account.

What are the Rules set by the Government to Curb Deepfakes?

- **IT Act, 2000 and IT Rules, 2021**: Both the IT Act and IT Rules have clear instructions which **place the onus on social media intermediaries to ensure such deep-fake videos or photos are taken down as soon as possible**. In case of failure, there are provisions for imprisonment of up to three years of fine of Rs 1 lakh.
    - Section 66D of IT Act: Section 66D of the IT Act, 2000 states that anyone **who cheats by personating using a communication device or computer resource can be punished** with Imprisonment of up to three years and a fine of up to one lakh rupees.
    - Rule 3(1)(b)(vii): This Rules mandates that social media intermediaries must **ensure that the users of their platform do not host any content which impersonates another person.**
    - Rule 3(2)(b): It requires such **content to be taken down within 24 hour**s of receipt of a complaint against such content.

//

The Rule Book

3(1)(b)(vii) of the IT Rules mandates that social media intermediaries must ensure that the users do not host content which impersonates another person

Rule 3(2)(b) requires such content to be taken down within 24 hours of receipt of a complaint against such content

What Should be done to Address the Menace of Deepfakes?

- **Learning from Other Countries**: The life cycle of deepfakes can be divided into three parts – **creation, dissemination and detection.** AI regulation can be used to mitigate the creation of unlawful or non-consensual deepfakes.
    - One of the ways in which **countries such as China** are approaching such regulation is to require providers of deepfake technologies to obtain consent of those in their videos, verify the identities of users, and offer recourse to them.
    - The **Canadian approach** to prevent harm from deepfakes includes mass public awareness campaigns and possible legislation that would make creating and distributing deepfakes with malicious intent illegal.
- **Adding Watermarks to all AI-generated Videos**: Adding watermarks to AI-generated videos is essential for effective detection and attribution. Watermarks reveal the content's origin and ownership, serving various purposes. They aid in attribution by clarifying the content's creator or source, especially when shared in different contexts.
    - **Visible watermarks also act as a deterrent against unauthorized use**, making it clear that the content can be traced back to its source.
    - Furthermore, **watermarks support accountability by providing evidence of the original creator's rights**, simplifying the enforcement of copyright and intellectual property protections for AI-generated content.
- **Deterring Users to Upload Inappropriate Content**: Online platforms should take steps to educate and inform users about their content policies, and perhaps implement measures to deter the upload of inappropriate content.
- **Developing and Improving Deepfake Detection Technologies**: This can involve using more sophisticated algorithms, as well as developing new methods that can identify deepfakes based on their context, metadata, or other factors.
- **Strengthening Digital Governance and Legislation**: This can involve creating clear and consistent laws and policies that define and prohibit the malicious use of deepfakes, as well as providing effective remedies and sanctions for the victims and perpetrators of digital harm.
- **Enhancing media Literacy and Awareness**: This can involve educating the public and the media about the existence and potential impact of deepfakes, as well as providing them with the skills and tools to verify and report suspicious content.
- **Promoting Ethical and Responsible use of Deepfake Technology**: This can involve establishing and enforcing codes of conduct and standards for the creators and users of deepfake technology, as well as encouraging its positive and beneficial applications.

Drishti Mains Question:

Q. Discuss the potential uses and threats of deepfake technology. In light of this, examine the measures governments and technology companies can take to mitigate the negative consequences of deepfakes.

PDF Refernece URL: https://www.drishtiias.com/printpdf/deepfakes-opportunities-threats-and-regulation