



## Perspective: Combating Deepfakes

**For Prelims:** Deepfake Technology, Deep Synthesis Technology, Artificial Intelligence (AI), Blockchain Technology, IT Act, 2000, IT Rules, 2021, Section 66D of IT Act, 2000

**For Mains:** Impact of Deepfake Technology, Dealing with Deepfakes, Ethical concerns arising out of its misuse.

### What is the Context?

**Deepfake** has emerged as a serious threat to democracy and social institutions across the world. Propagation of deepfake content via social media platforms has aggravated this challenge.

- The **Ministry of Electronics and Information Technology** has, from time to time, advised social media platforms to exercise due diligence and take expeditious action against deepfakes.

### What are Deepfakes?

- **About:**
  - Deepfakes are synthetic media that use **Artificial Intelligence (AI)** to manipulate or generate visual and audio content, usually with the intention of deceiving or misleading someone.
  - Deep fakes first came into notice in **2017 when a Reddit user** posted explicit videos of celebrities. After that several instances have been reported.
- **Deepfake Creation:**
  - Deepfakes are created using a technique called **generative adversarial networks (GANs)**, which involve two competing neural networks: **a generator and a discriminator**.
    - The generator tries to **create fake images or videos that look realistic, while the discriminator tries to distinguish** between the real and the fake ones.
    - The **generator learns from the feedback of the discriminator and improves** its output until it can fool the discriminator.
  - Deepfakes **require a large amount of data, such as photos or videos**, of the source and the target person, which are often collected from the internet or social media without their consent or knowledge.
  - Deepfakes are a **part of deep synthesis**, which uses technologies, including **deep learning and augmented reality**, to generate text, images, audio and video to create virtual scenes.

### 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:**

- Deep learning technology has enabled **positive advancements, such as restoring lost voices and recreating historical figures** to deliver engaging lessons.
  - 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.

▪ **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 Present Concerns With the Deepfakes?

▪ **Procedural Concerns:**

- **Deepfakes are a problem because they can be used for various malicious purposes:**
  - **Spreading propaganda**, and fake news;
  - **Influencing elections** and public opinion;
  - Blackmailing and **extortion individuals** or organizations;
  - Damaging the reputation and credibility of celebrities, politicians, activists, and journalists; and
  - **Creating non-consensual pornography** and revenge porn.
  - Deepfakes can cause **various harms, such as eroding trust in institutions, media, and democracy**, and undermining the rule of law and human rights.
  - Deepfake technology can **violate the privacy, dignity, and reputation of individuals**, and harm the mental health and well-being of the victims, especially women, who are often the targets of such malicious manipulation.

▪ **Legal Concerns:**

- In the US, deep fake legality is complex. **Victims can claim defamation, but removing content may be seen as censorship**, potentially violating the First Amendment, which protects freedom of religion, expression, assembly, and petitioning.
- However, [the Right to be Forgotten](#), allows a user to request companies such as Facebook and Google, that have collected his/her data to take it down.

▪ **Ethical Concerns:**

- Deepfakes have been employed for **malicious purposes, including revenge porn and**

### **hacking facial recognition systems.**

- They undermine trust in the media and **blur the lines between fact and fiction.**
- Misinformation propagated by **deepfakes can be mistaken as true**, leading to potential social unrest.

## **What are the Global Approaches to Combat Deepfakes?**

### ▪ **India:**

- India does not have specific laws or regulations that ban or regulate the use of deepfake technology.
- India has called for a global framework on the expansion of **“ethical” AI** tools.
- Existing laws such as **Sections 67 and 67A of the Information Technology Act, 2000** have provisions that may be applied to certain aspects of deep fakes, such as defamation and publishing explicit material.
- Section 500 of the **Indian Penal Code (1860)** provides punishment for defamation.
- The **Digital Personal Data Protection Act, 2023** provides some protection against the misuse of personal data.
- The **Information Technology Rules, 2021**, mandate the removal of content impersonating others and artificially morphed **images within 36 hours.**
- India needs to develop a comprehensive legal framework specifically targeting deepfakes, considering the potential implications for **privacy, social stability, national security, and democracy.**

### ▪ **Global:**

- The recent **world’s first ever AI Safety Summit 2023** involving 28 major countries, including the US, China, and India, agreed on the need for global action to address AI’s potential risks.
  - The **‘Bletchley Park Declaration’** at the summit acknowledged the risks of intentional misuse and the loss of control over AI technologies.
- The **Global Partnership on Artificial Intelligence (GPAI)** summit was held in New Delhi in December 2023. The summit concluded with the adoption of the **New Delhi Declaration on artificial intelligence.**
  - The declaration built consensus among GPAI members on advancing **safe, secure, and trustworthy AI** and commitment to supporting the sustainability of GPAI projects.

### ▪ **European Union:**

- The **European Union’s Code of Practice on Disinformation** requires tech companies to counter deep fakes and fake accounts within six months of signing up to the Code.
  - If found non-compliant, tech companies can **face fines up to 6%** of their annual global turnover.
- The European Union has passed the **world’s first comprehensive laws** to regulate the use of artificial intelligence. **The Artificial Intelligence Act (AI Act) aims to introduce a common regulatory and legal framework for artificial intelligence.**
  - The draft regulation aims to ensure that AI systems placed on the European market and used in the EU are safe and respect fundamental rights and EU values.

### ▪ **United States:**

- The U.S. introduced the bipartisan **Deepfake Task Force Act** to assist the **Department of Homeland Security** in countering deepfake technology.

### ▪ **China:**

- China introduced **comprehensive regulation on deep synthesis**, effective from 2023.
- Aimed at curbing disinformation, the regulation requires clear labelling and traceability of deep synthesis content.
- The Regulations impose obligations on the providers and users of so-called **“deep synthesis technology”**.

### ▪ **Tech Companies:**

- Big tech companies like **Meta and Google** have announced measures to address the issue of deep fake content.
  - However, there are still vulnerabilities in their systems that allow the dissemination of such content.
- **Google** has introduced tools for identifying synthetic content, including watermarking and

metadata.

- **Watermarking embeds information directly into content**, making it resistant to editing, while metadata provides additional context to original files.

## 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**. Providing users with skills and tools to verify and report suspicious content, like
  - **Look for visual and audio inconsistencies** in the media.
  - Use **reverse image search** to find the original source or similar images.
  - Use **AI-based tools** to analyze the **quality, consistency, and authenticity** of the images or videos.
  - Using **digital watermarking** or [blockchain](#) to verify the source and integrity of the media.
  - **Educate oneself and others** about deepfake technology and its implications.

### ▪ **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.

### ▪ **Blockchain-based Deepfake Verification:**

- Use [blockchain technology](#) to create an unchangeable record of who created a piece of digital media and ensure transparency in verifying its authenticity.
- This decentralized approach allows individuals to trace the origin and modification history of media, discouraging the creation and dissemination of malicious deepfakes.



## **Prelims**

**Q. With the present state of development, Artificial Intelligence can effectively do which of the following? (2020)**

1. Bring down electricity consumption in industrial units
2. Create meaningful short stories and songs
3. Disease diagnosis
4. Text-to-Speech Conversion
5. Wireless transmission of electrical energy

**Select the correct answer using the code given below:**

- (a)** 1, 2, 3 and 5 only  
**(b)** 1, 3 and 4 only  
**(c)** 2, 4 and 5 only  
**(d)** 1, 2, 3, 4 and 5

**Ans: (b)**

## **Mains**

**Q. "The emergence of the Fourth Industrial Revolution (Digital Revolution) has initiated e-Governance as an integral part of government". Discuss. (2020)**

PDF Refernece URL: <https://www.drishtiias.com/printpdf/perspective-combating-deepfakes>

