Deepfakes

For Prelims: Deepfake technology, **Deep synthesis technology**, Artificial Intelligence technology, **blockchain technology**

For Mains: Impact of Deepfake Technology, Dealing with deepfakes, Ethical concerns,

Source: TH

Why in News?

A Deepfake **video** showing an Indian actress has sparked **outrage and concern over the misuse of** artificial intelligence (AI) to create realistic but fake videos, also known as deepfakes.

What are Deepfakes?

- **About:**
    - Deepfakes are **synthetic media that use AI** to manipulate or generate visual and audio content, usually with the intention of deceiving or misleading someone.
- **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.
- **Positive Applications of Deep Learning**:
    - Deep learning technology has enabled positive advancements, such as **restoring lost voices and recreating historical figures.**
    - Deep learning techniques have been applied in **comedy, cinema, music, and gaming to enhance artistic expression**.
    - Synthetic avatars of people with **physical or mental disabilities** will help express themselves online.
    - It enhances **medical training and simulation** by generating diverse and realistic medical images. It also creates **virtual patients and scenarios** for simulating medical conditions and procedures, improving training efficiency.
    - It can also be used to enhance the interaction and immersion of augmented reality (AR) and gaming applications.

- **Concerns Regarding the Deepfakes**:
  - Deepfakes are a problem because they can be used for various malicious purposes, such as
    - **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.
- **Detection:**
  - 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.

What are the Global Approaches Related to Deepfake Regulation?

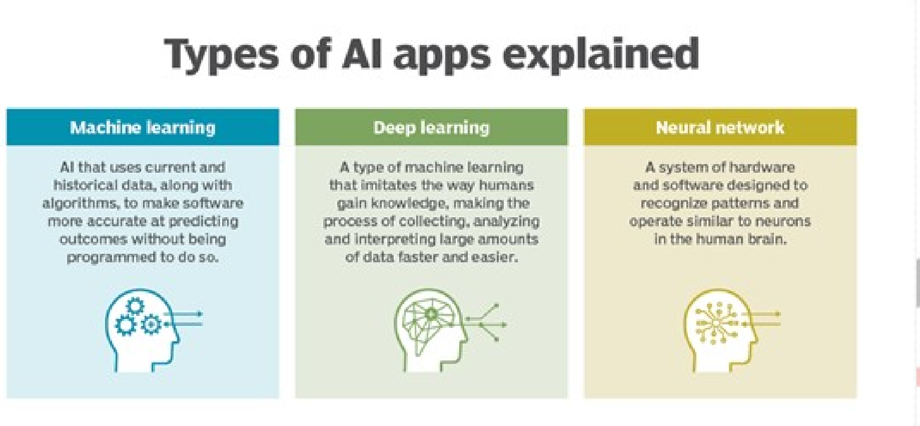
- **India:**
  - India **does not have specific laws or regulations that ban or regulate the use of deepfake technology.**
  - India has called for a global framework on the expansion of “ethical” AI tools.
  - Existing laws such as **Sections 67 and 67A of the** Information Technology Act (2000) have 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, 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 declaration at the summit** acknowledged the risks of intentional misuse and the loss of control over AI technologies.
  - **European Union:**
    - The European Union's Code of **Practice on Disinformation** requires **tech companies to counter deep fake**s 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**
  - **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 January 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

The Vision

"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.

//



Types of AI apps explained

Machine learning

AI that uses current and historical data, along with algorithms, to make software more accurate at predicting outcomes without being programmed to do so.

Deep learning

A type of machine learning that imitates the way humans gain knowledge, making the process of collecting, analyzing and interpreting large amounts of data faster and easier.

Neural network

A system of hardware and software designed to recognize patterns and operate similar to neurons in the human brain.

दृष्टि The Vision

Way Forward

- Developing and implementing comprehensive laws and regulations that **specifically target the creation and dissemination of deepfakes**, while balancing the freedom of speech and expression.
- Enhancing the **public awareness and media literacy of the potential risks** and impacts of deepfakes, and **encouraging critical thinking and verification** of the sources and content of media.
- Creating and adopting **technical solutions and standards that can detect, prevent**, and remove deepfakes, such as digital watermarks, and blockchain.
- Promoting **ethical and responsible use of deep learning technology** and synthetic media, and establishing codes of conduct and best practices for the creators and users of deepfakes.
- Fostering **collaboration and coordination among various stakeholders**, such as governments, media, civil society, academia, and industry, to address the challenges and opportunities posed by deepfakes.
-
- **Legal Insight**: Indian Laws Safeguarding Society against Deepfake Technology

https://www.drishtijudiciary.com/en

- Is Deepfake Technology Dangerous? In News Drishti IAS English

UPSC Civil Services Examination Previous Year Question (PYQ)

Mains

Q. What are the main socio-economic implications arising out of the development of IT industries in major cities of India? **(2022)**

PDF Refernece URL: https://www.drishtiias.com/printpdf/deepfakes-3

दृष्टि The Vision