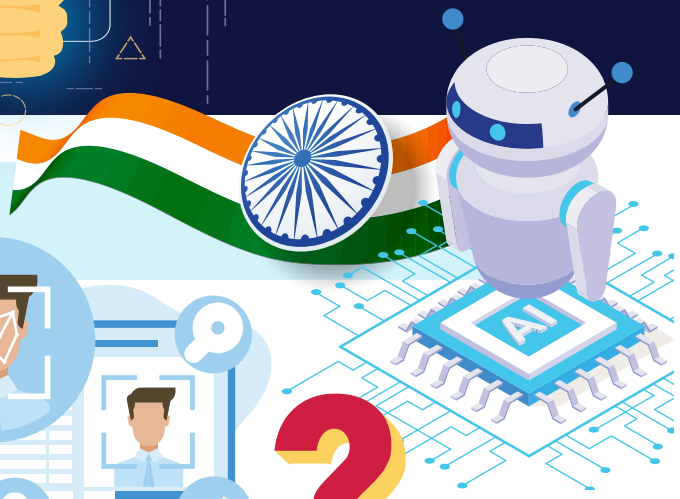
DEEPFAKES

IASGYAN
PREDICT THE UNPREDICTABLE
An initiative of APTI PLUS

Context

- Prime Minister Narendra Modi (November 17) said it's important to understand how artificial intelligence (AI) works. It could be used to create 'deepfakes' to purposefully spread false information or have malicious intent behind their use.
- He also urged the media to spread awareness about the issue.



What are deepfakes?

- Deepfakes are synthetic media that use AI to manipulate or generate visual and audio content, usually with the intention of deceiving or misleading someone.

How deepfakes are created?

- **Technology:** Generative Adversarial Networks (GANs)
- GANs involve two competing neural networks: **a generator and a discriminator**
- **Work of Generator:** to create fake images or videos that look realistic
- **Work of Discriminator:** to distinguish between the real and the fake ones
- The generator learns from the discriminator's feedback and improves its output until it can trick the discriminator.
- **Sources for Deepfakes:** a large amount of data, such as photos or videos, of the source and the target person, often collected from the internet or social media without their consent or knowledge.

What is Deep Synthesis?

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 application of Deepfakes	Negative application of Deepfakes
• It Restoring lost voices and recreating historical figures.	• Spreading propaganda, and fake news;
• Applied in comedy, cinema, music, and gaming to enhance artistic expression.	• Influencing elections and public opinion;
• Enhances medical training and simulation by generating diverse and realistic medical images	• Blackmailing and extortion of individuals or organizations;
• To enhance the interaction and immersion of augmented reality (AR) and gaming applications.	• Damaging the reputation and credibility of celebrities, politicians, activists, and journalists; and
• Synthetic avatars of people with physical or mental disabilities will help express themselves online.	• Creating non-consensual pornography and revenge porn.
	• Eroding trust in institutions, media, and democracy, and undermining the rule of law and human rights.
	• Can violate the privacy, dignity, and reputation of individuals, and harm the mental health and well-being of the victims, especially women.

Detection of Deepfake technology

- Unnatural eye movements
- Mismatches in colour and lighting
- Compare and contrast audio quality
- Strange body shape or movement
- Artificial facial movements

- Unnatural positioning of facial features
- Awkward posture or physique
- Verify before you share
- Stay Informed
- Use AI detection tools if possible

The Global Approaches Related to Deepfake Regulation

India:

- There are **no particular rules or regulations** in India that prohibit or restrict the use of deepfake technology.
- India has asked for a global framework for the expansion of **"ethical" AI tools.**
- Existing laws, such as **Sections 67 and 67A of the Information Technology Act (2000),** contain prohibitions that may apply to certain parts of deep fakes, such as defamation and the publication of explicit material.
- Defamation is punishable under **Section 500 of the Indian Penal Code (1860).**
- **The Digital Personal Data Protection Act** provides various safeguards against personal data misuse.
- **The Information Technology Rules, 2021,** require the removal of content impersonating others and digitally manipulated photographs within 36 hours.

Global:

- The recent first-ever AI Safety Summit 2023, attended by 28 key countries including the United States, China, and India, agreed on the need for global action to address AI's potential hazards.
- The summit's Bletchley Park Declaration addressed the hazards of intentional misuse and loss of control over AI technologies.

Europian 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 of 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.

Big Tech companies

Google

- Google has introduced tools for identifying synthetic content, including watermarking and metadata.