General Studies-3; Topic: Science and Technology- developments and their applications and effects in everyday life. Basics of cyber security.

Countering Deepfakes

Introduction

- Deepfakes are the digital media (video, audio, and images) manipulated using Artificial Intelligence.
- Deepfakes are a new tool to spread computational propaganda and disinformation at scale and with speed.
- From creating fake pornographic videos to making politicians appear to say things they did not, the potential for damage to individuals, organisations and societies is vast.

What is Deepfake and how is it made?

- Advances in **Artificial Intelligence (AI) and Machine Learning (ML)** have enabled computer systems to create synthetic videos, a.k.a. deepfakes.
- A deepfake video can show a person saying or doing something that they never said or did.

Present Status

- **Synthetic media is growing exponentially with malicious actors targeting individuals** most connected to the entertainment business, followed by politics.
- As of June 2020, deepfake videos identified online doubled to 49,081 in just six months since January, according to an analysis.
- The rising number of synthetic content mimicks people in the entertainment industry.
- In India, a day before Delhi election in February, two videos of BJP Delhi unit President urging citizens to vote for BJP, in english and haryanvi, were sent to 15 million voters via 5800 WhatsApp groups. It was reported that the videos were deepfakes.
- Hardware has become cheaper and more powerful and software more accessible and capable.
- The incorporation of AI in software has made it “dramatically” easier to manipulate media.

What Deepfakes Does?

- Deepfakes can inflict **damage to individuals, institutions, businesses and democracy.**
- They make it possible to fabricate media — swap faces, lip-syncing, and puppeteer — mostly without consent and bring threat to psychology, security, political stability, and business disruption.
- Deepfakes can depict a person indulging in antisocial behaviours and saying vile things.

Impact of Deepfakes

- **Targeting Women**
  - Deepfake pornography exclusively targets women.
  - It inflicts emotional, reputational, and in some cases, violence towards the individual.
- It can have severe implications on a persons reputation, sabotaging their professional and personal life.
- A deepfake can also aid in **altering the democratic discourse** and undermine trust in institutions and impair diplomacy.
- Leaders can also use them to increase populism and consolidate power.
- Deepfakes can become a very effective tool to sow the seeds of polarisation, amplifying division in society, and suppressing dissent.

Concerns / Challenges

- Even if the victim could explode the fake, it may come too late to remedy the initial harm.

- Deepfakes can be deployed to extract money, confidential information, or exact favours from individuals.
- Deepfakes can cause short- and long-term social harm and accelerate the already declining trust in news media.
- The ability for social media to make things viral compounds the problem.
- It can be used by insurgent groups and terrorist organisations, to represent their adversaries as making inflammatory speeches or engaging in provocative actions to stir up anti-state sentiments among people.
- According to research, individuals in South Korea, India, and Japan make up a significant proportion of targets.

Way Forward

- To defend the truth and secure freedom of expression, we need a **multi-stakeholder and multi-modal approach.**
- Legislative regulations, technology intervention, and media literacy can provide effective and ethical countermeasures to mitigate the threat of malicious deepfakes.
- Artificial Intelligence can help detect deepfake videos.
- Media literacy for consumers and journalists is the most effective tool to combat disinformation and deepfakes.