Responsible AI For All

For Prelims: NITI Aayog, AI, Facial Recognition Technology, Right to Privacy, Puttaswamy Judgement, Personal Data Protection Law, Digi Yatra Programme

For Mains: Facial Recognition Technology - Pros and Cons, Socio-economic implications of AI, AI and Ethics

What is the Context?

In 2018, NITI Aayog **released the National Strategy on Artificial Intelligence (NSAI),** that inter alia highlighted the roadmap to adopt **Artificial Intelligence** in five public sectors in a manner that is safe and dispenses benefits to all citizens. The strategy document coined the **"AI for All"** mantra, to be the governing benchmark for future AI design, development, and deployment in India. A part of this strategy was to **ensure the safe and responsible use of AI (RAI).**

The RAI principles come in the background of a growing call for **developing governance and regulatory frameworks to mitigate potential risks of AI,** while maximising its benefits for the largest number of people. **Facial recognition technology (FRT)** has been taken as the first use case for examining the RAI principles and operationalisation mechanisms.

FRT has garnered domestic and international debate around its potential benefits of efficient and timely execution of existing processes in different sectors. However, it also poses risks to basic human and fundamental rights like individual privacy**,** equality, free speech and freedom of movement.

//

What are the Recent Developments in AI?

- **Development of AI:** In an increasingly technology centric society, the surge in designing and development of artificial intelligence (AI) driven tech is becoming ubiquitous.
    - While the **origins of AI are traceable to the 2nd half of the 20th century,** the past decade has witnessed a rapid resurgence.
        - This is attributable majorly to **Big Data analytics** - data collection, aggregation and processing, machine learning, deep learning, neural networks, natural language processing, etc.
- **AI and Ethics:** The other side of this technological revolution is a growing apprehension on the **socio-political and economic implications of AI**, specifically the concerns about co-existence of these emerging technologies and core principles of modern democracies.
    - Consequently, **AI ethics and the safe and responsible application of AI** are becoming front and centre of the technology revolution.
    - **Constitutional morality was envisioned as the cornerstone for AI ethics' principles in India,** thus, propelling our constitutional rights and ethos to the paramount consideration for deploying AI in a responsible manner.

What is Facial Recognition Technology (FRT)?

- **About**: It is a collective term referring to **different kinds of technologies that are designed to identify or trace individuals using visual images** (pictures/videos).
    - FRT uses key features of the face and their respective distances from one another to morph

a virtual facial map.
        - This ecosystem is dependent on the availability of facial data as the FRT programs, prior to their rollout, are engaged in intensive training and machine learning processes through large amounts of training datasets.
    - It has the benefits that any automation brings, which is to **expedite manual efforts with more efficiency in processes.**
    - The use of FRT has witnessed a significant debate globally around its ethical, legal, and constitutional ramifications.
- **Types:**
    - **1:1 FRT System:** FRT is mainly targeted at **verifying a specific person's facial data** (which is captured live) with a specific facial image data **from a gallery dataset, such as a face lock for unlocking phones.**
        - It exercises identification through verification between two specific faces, and **greater control over the quality** of facial images.
    - **1:n FRT System:** This FRT is **primarily used in identification** - to process a large number of faces captured in either image or video format to specifically identify a particular person's face.
        - It is mostly used in **Live Facial Recognition Technology (LFRT)** applicable to law enforcement, and other mass monitoring and surveillance purposes.
- **Applications of FRT:**
    - **Security Uses:** Typically include the use of FRT **for general law and order considerations - investigation, identification** of missing persons, identifying persons of interest to the law enforcement, **monitoring** of crowds etc.
    - **Non-Security Uses:** It **more likely involves 1:1 use of FRT**; international uses of FRT to provide greater ease of access to **airport facilities,** educational systems using FRT to **generate unique IDs,** authentication to provide **access to products, services, & public benefits** and recording **biometric attendance** of workers.

What are the Risks Associated with FRT?

- **Design Based Risks:**
    - **Inaccuracy due to technical factors** (ageing, plastic surgery, disfigurement, pose variation, occlusion, quality of image)
    - **Inaccuracy due to bias caused by underrepresentation** (skin-tone or gender based disparities)
    - **Inaccuracy due to lack of training of human operators**
    - **Inaccuracy due to glitches**
    - **Security risks due to data breaches and unauthorised access** (vast amount of facial data - financially valuable target for hackers)
    - **Accountability and legal liability issues** (due to involvement of various entities in developing, testing, training and deploying the FRT)
- **Rights-Based Challenges:**
    - **Privacy related risks** (individual may not be aware/in control of the extent of their biometric facial data being processed)
    - **Issues of informational autonomy** (biometric facial images collected for one purpose and subsequently used for another purpose to which the concerned person remains oblivious)
    - **Threat to anonymity - a facet of privacy** (FRT systems being used to suppress dissent and protests across the world)

What is the Status of Regulating FRT?

- **European Union (EU):** Apart from the **General Data Protection Regulations (GDPR) and Data Protection Directive**, the **EU has now proposed an AI Act to establish a risk-based compliance framework** where **FRT systems have been categorised as "high risk"** with the highest level of compliance requirements.
- **UK, US, Canada and Australia:** Regulation of FRT **primarily comes under their data protection/ privacy laws.**

What is India's Digi Yatra Programme?

- **About:**
  - **Digi Yatra** is a **proposed biometric boarding system for use at Indian airports,** intended to create a seamless, paperless, and contactless check-in and boarding experience for passengers.
- **Aim:**
  - It envisages **an identity management ecosystem** for Indian airports which can enhance the **capabilities of Indian civil aviation** infrastructure, digitise manual processes at airports, improve security standards and lower the cost of operations of airports.
- **Technology:**
  - It **proposes use of FRT to authenticate a passenger's travel credentials**, which allows other checkpoints in an airport to be operated in an automated form with minimal human involvement.
- **Legal and Institutional Backing:**
  - A **Digi Yatra policy was released in 2018**, which sets out the passenger processes and technical features of Digi Yatra.
  - The **Digi Yatra Foundation ('DYF'), a not-for-profit company** under Section 8 of the **Companies Act, 2013** was established in 2019 for the implementation of the **Digi Yatra Central Ecosystem.**
- **Mandate:**
  - The Digi Yatra programme is **conceptualised as a purely voluntary mechanism**, and therefore, at various stages**, it sets out the alternative means i**n which the boarding process will operate for passengers who have not opted-in to the programme.
- **Benefits:**
  - Lower congestion at airports
  - Seamless, paperless and contactless passenger experience
  - Lower operational costs and enhanced civil aviation capabilities
- **Areas of Concern:**
  - Data privacy
  - Aadhar based authentication
  - Information Security
  - To address these concerns, the Digi Yatra Programme should be in compliance with the Responsible AI (RAI) Principles.

What are the Principles of Responsible AI?

- **Safety and Reliability**: AI systems must ensure **reliability regarding their intended functions** and must have **built-in safeguards** to ensure the **safety of stakeholders.**
- **Equality:** AI systems must be built keeping in mind that **similar people in similar circumstances are treated equally.**
- **Inclusivity and Non-Discrimination:** AI systems must be developed to be inclusive of all stakeholders, and must not discriminate through bias between stakeholders on **religion, race, caste, sex, descent, place of birth or residence in matters of education, employment, access to public spaces etc.**
- **Privacy and Security:** AI systems must ensure that the **personal data of data subjects must be safe** and secure, such that **only authorised persons must access personal data** for specified and necessary purposes, within a framework of sufficient safeguards to ensure this process.
- **Principle of Transparency:** The design and training of AI systems is key for its functioning. The **system must be audited and be capable of external scrutiny** to ensure that the **deployment of the AI system is impartial,** accountable and free from bias or inaccuracies.
- **Principle of Accountability:** Since there are various actors in the process of developing, deploying and operationalizing an AI system, the **accountability structures for any effects, harms or damages** by the AI system **must be clearly set out in a publicly accessible and understandable manner.**

- **Protection and Reinforcement of Positive Human Values:** This principle focuses on the **possible deleterious effects of AI systems** through collection of personal data for profiling, the **use of AI systems in manners contrary to fundamental rights** guaranteed by the Constitution of India.

How can Responsible AI Principles be Implemented?

Principles	Measures
Principles of Safety & Reliability and Accountability	▪ A responsible/**accountable agency** ▪ Creating standardised, annotated, high quality images ▪ **Periodically evaluating FRTs** ▪ Provision for monitoring of the performance of the entire system
Principle of Equality	▪ Comprehensive data protection law **clarifying the requirement of explicit consent** of the individual as well as of spouse/ guardian
Principle of Inclusivity and Non- Discrimination	▪ **Standards to avoid bias** in the FRT model must be developed ▪ **Alternatives to FRT** should be available so that **people on the other side of the** digital divide don't remain excluded
Principles of Privacy & Security and Transparency	▪ **Internal SOPs** for handling personal and sensitive personal data ▪ **Security-based exceptions** should be clearly identified and set out within the SOPs

What does the Report Recommend for Responsible Use of FRT?

- **Legal Reforms:**
  - It is **imperative to have a codified data protection regime** in the country at the earliest possible.
  - The data protection regime must not be limited to regulating data processing by private entities but must **adequately codify protections for fundamental right to privacy** against state agencies.
  - Any ongoing or future application of FRT systems by governments in India, must be compliant with the **three-pronged test of legality, reasonability, and proportionality,** set out by the **SC in the** Puttaswamy judgement.
- **Policy Reforms:**
  - An overly opaque FRT system may prevent independent scrutiny. **Transparency around the deployment of FRT systems must be a norm**; it is **necessary for securing public trust** in the development and deployment of such systems.
  - Organisations deploying an AI system can **constitute an ethical committee (with adequate autonomy)** to assess the ethical implications and oversee mitigation measures.
- **Recommendations for Developers of FRT Systems:**
  - Developers must **build FRT systems that are explainable,** i.e., the decision-making process of the system regarding a particular case output can be accurately explained to an auditor or judge.
  - Developers must co**nsider the realities of the Indian population** in training the AI model and **ensure accurate and inclusive identification based on gender, skin-tone etc.**

UPSC Civil Services Examination, Previous Year Questions (PYQs)

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

Q1. "The emergence of the Fourth Industrial Revolution (Digital Revolution) has initiated e-Governance as an integral part of government". Discuss. **(2020)**

Q2. What are the areas of prohibitive labour that can be sustainably managed by robots? Discuss the initiatives that can propel the research in premier research institutes for substantive and gainful innovation. **(2015)**

Q3. "Human beings should always be treated as 'ends' in themselves and never as merely 'means'." Explain the meaning and significance of this statement, giving it's implications in the modern techno-economic society. **(2014)**

Source