Watermarking on AI Generated Content

Source: ToI

Why in News?

Recently, ChatGPT-maker OpenAI **developed a tool to detect whether its AI (Artificial Intelligence) chatbot has been used** to write essays, research papers, or generate photos.

- To address issues concerning the authenticity and ownership of such content, leading technology companies like OpenAI, Meta, Microsoft, Google, and Adobe are **developing and implementing watermarking techniques**.

//

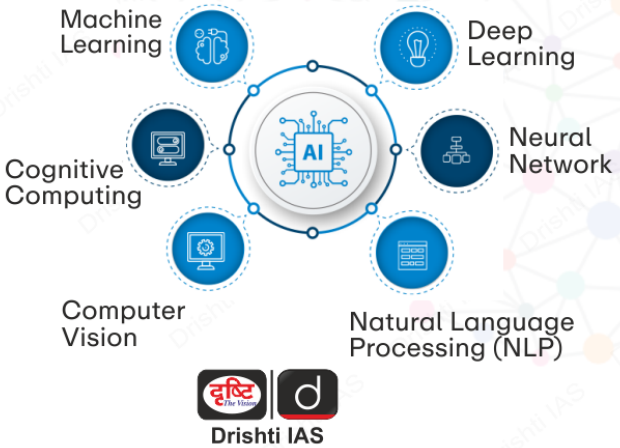
Artificial Intelligence(AI)

AI is the simulation of human intelligence in machines programmed to think and learn like humans, capable of problem-solving, reasoning, and adapting to new information.

AI Timeline - Major Milestones

- **1950s** – Turing Test Proposed; First AI Programs Developed
- **1956** – Dartmouth Conference Coins "Artificial Intelligence"
- **1960s** – Eliza Chatbot Created; Early Neural Networks Emerge
- **1996** – Deep Blue - a Chess-Playing Program
- **2012** – Deep Learning Breakthrough in Image Recognition
- **2014** – Generative Adversarial Networks (GANs) Introduced
- **2020** – GPT-3 Demonstrates Advanced Language Generation
- **2022** – Chatgpt Launches, Bringing Conversational AI to Masses
- **2023** – Generative AI Boom; Major Tech Companies Release AI Models

KEY COMPONENTS OF AI



Drishti IAS

Applications of AI

- **Healthcare:** Personalised medicine
- **Finance:** Algorithmic trading
- **Transportation:** Autonomous vehicles
- **Marketing & Customer Service:** Targeted advertising, chatbots
- **Education:** Adaptive learning systems, personalised tutoring
- **Agriculture:** Crop monitoring
- **Cybersecurity:** Threat detection
- **Energy:** Smart grid management, consumption forecasting

Concerns

- Deepfakes & misinformation
- Algorithmic bias
- Automation & job displacement
- Privacy issues
- Data ownership & liability issue
- Ethical decision-making complexes

Regulating AI

- **Global Partnership on AI (GPAI)** launched in 2020
- **Bletchley Declaration (2023):** Enhance Global Collaboration on AI
- **G20 New Delhi Leaders' Declaration (2023):** Harnessing AI responsibly for good and for all
- **Hiroshima AI Process (2023)** by G7

India and AI

- **National Strategy For AI 2018**
- **AI For All:** Self-learning online program
- **GPAI Summit 2023** hosted by India
- **IndiaAI Mission 2024**
- **US India Artificial Intelligence (USIAI) Initiative:** AI cooperation in critical areas
- **AIRAWAT** (AI Research, Analytics and Knowledge Assimilation Platform): Supercomputer

What is Watermarking on AI Generated Content?

- **About:**

- The AI watermarking technique is **used to identify what's AI-generated and what is an original work**.
    - It is a **digital signature that serves as a unique identifier**, akin to a fingerprint, allowing for the tracing of the AI model used to create the content back to its origin.

- **Need of Watermarking:**
    - **Authentication and Validation:** It provides a reliable method for verifying the authenticity of digital files, crucial in combating **deepfake** videos, manipulated images, and deceptive media in a digitally-driven world.
    - **Tamper-Evident Records:** Integrating AI watermarking with **blockchain** and **public key infrastructure** ensures that any attempts to alter or manipulate content are detectable, preserving the integrity of digital records.
    - **Trust and Confidence:** By assuring the authenticity of media, AI watermarking helps content creators, distributors, and consumers combat **misinformation** and counterfeiting, bolstering trust and understanding of content provenance.
- **Steps to Ensure AI Authenticity:**

    - **Coalition for Content Provenance and Authenticity (C2PA)**: A collaboration between Adobe, Intel, Microsoft, Sony, and other leading firms to establish standards for verifying the authenticity of audio-visual content.

    - **Use of Blockchain Technology**: **Blockchain technology** ensures transparency by providing an immutable, publicly accessible record of digital asset ownership and provenance.
    - **Ethereum Improvement Proposal:** This project proposes adding C2PA consent data, including content consent for AI and ML (Machine Learning) data mining, to on-chain metadata, leveraging blockchain technology **to enhance security and transparency**.

UPSC Civil Services Examination, Previous Year Questions (PYQs)

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)