

NEXT IAS

**DAILY EDITORIAL
ANALYSIS**

TOPIC

Dangers of AI

www.nextias.com

DANGERS OF AI

Context

- 2023 was perceived by both industry leadership and the populace as one where **artificial intelligence had a significant impact on social and economic relations.**

About

- This impact was visible due to the **apparent success of large language models, a family of generative models, in solving complex tasks.**
- The year started with **Microsoft deciding to invest \$10 billion in the OpenAI project**, as its **ChatGPT** became **the fastest-growing application.**
- **Google introduced its chatbot, Bard**, Amazon introduced **Bedrock**, giving its customers access to large language models of its own called **Titan.**

Dangers of AI

The benefits of AI are so large, from healthcare, defence and defence that the **downsides are often ignored.**

Do you know?

- **The AI safety letter**, which was signed by more than 2,900 industry experts and academics, called for a **six-month halt on training AI systems more powerful than GPT4** as AI could prove to be an **existential threat.**
- **Privacy:** The **data hunger of AI** might have implications on both **diluting privacy and on labour conditions** of platform workers.
- **Surveillance:** AI's opaque workings might have impacts on democratic processes when AI systems are used in public-use cases like **surveillance and policing.**
 - ♦ AI, with its **ability to drive face-recognition** and compare huge data streams, **can give governments 360-degree 24x7 profiles of all citizens**, making dissent against authoritarian regimes more difficult.
- **Deep fakes:** By cloning voices and faces, it can deliver authentic-seeming fake news, or scam people, or bypass security.
 - ♦ Eg. **A lady in Gurugram was recently scammed for Rs 11 Lakh using deep fakes on Skype**, scammers impersonating as Senior officials of CBI.
- **Perpetuates biases:** Algorithms trained on current data may recommend that **only males get STEM scholarships, and only upper-caste people get bank loans.**
- **Nuclear threat:** If AI is used to control nuclear missile systems, it could cause extinction, as some experts have warned.

Mitigating the dangers:

- **Ethical development and governance:** Establishing ethical frameworks and principles for AI development and deployment is crucial to ensure responsible and beneficial advancement.
- **Transparency and accountability:** AI systems should be transparent and accountable for their decisions, allowing for human oversight and intervention.
- **Education and awareness:** Raising public awareness about the potential dangers of AI and educating citizens about their rights and responsibilities in the digital age is essential.
- **Collaboration and international cooperation:** Addressing AI risks effectively requires global collaboration and coordinated efforts among governments, researchers, and the private sector.
 - ♦ Some have proposed the setting up of an **"International Agency for Artificial Intelligence" (IAAI)**, much like the International Atomic Energy Agency (IAEA) that was set up to regulate the uses of nuclear energy.

Global efforts to regulate AI:

- **Europe's AI Act:** It was passed in December 2023 and has concrete red lines like **prohibiting arbitrary and real-time remote biometric identification** in public spaces for law enforcement, **bans emotion detection**, etc.
- In November 2023, the '**Bletchley Declaration**' by **AI Safety Summit** called for,
 - ♦ to work together in an inclusive manner to ensure **human centric, trustworthy and responsible AI**,
 - ♦ **AI that is safe, and supports the good of all** through existing international fora and other relevant initiatives, and
 - ♦ **to promote cooperation to address the broad range of risks** posed by AI.
- In July 2023, the **US government** announced that it had persuaded the companies OpenAI, Microsoft, Amazon, Anthropic, Google, Meta, etc **to abide by "voluntary rules" to "ensure their products are safe"**.

Way Ahead

- **It is unlikely that the advancements in AI and research into mitigation and responsible usage** will proceed at the same pace.
 - ♦ Hence, countries should acknowledge this and should **develop safeguards fast enough to prevent catastrophic harm**.
- **By acknowledging the potential dangers of AI and proactively taking steps** to mitigate them, we can ensure that this transformative technology serves humanity and contributes to a safer, more equitable future.

Source: **IE**

