AI: A Double Edged Sword

This editorial is based on "Many elections, AI's dark dimension" which was published in The Hindu on 18/03/2024. The article discusses the potential impact of AI on democracies as several countries gear up for elections in 2024. It suggests that AI has the capacity to disrupt democratic processes and highlights the implications this may have.

For Prelims: Artificial Intelligence (AI), Ethical AI, Machine Learning, Large Language Models, Global Partnership on Artificial Intelligence, Artificial Intelligence Mission Generative AI, Deepfakes, ChatGPT, INDIAai.

For Mains: Generative AI – Benefits, Threats and Way Forward.

With the seven-phase general election in India having been announced, and to be held from April 19th to June 1st, 2024, political parties and the electorate cannot, however, afford to ignore the **Artificial Intelligence (AI)** dimension. This year, elections are also scheduled to be held in as many as 50 other countries across the globe, apart from, and including, India, Mexico, the UK and the US.

The potential of AI is already clear. Many such as Sam Altman of OpenAI in the US, believe that it is the most important technology in history. AI protagonists further believe that AI is set to turbocharge, and dramatically improve, the standard of living of millions of human beings. It is, however, unclear as of now, whether, as many suggest, AI would undermine human values and that **advanced AI could pose 'existential risks'.**

What are the Implications of AI for Elections Across the Globe?

The shadow of large language models looms over elections around the world, and stakeholders are aware that even one relatively successful deployment of an AI - generated disinformation tool could impact both campaign narratives and election results very significantly.

- **Emergence of AGI:**
  - Rapid technological breakthroughs in AI (especially its latest manifestation, such as **Generative AI**) carry their own burdens. It may be too early to fully contemplate the possible impact of **Artificial General Intelligence (AGI)** - AI systems that simulate the capability of human beings - but all this is indicative of yet another dimension to electoral dynamics that cannot be ignored.
    - The rapid development of AI models suggests that the world is at an inflection point in the history of human progress. The speed with which the development of newer skills is taking place suggests that the day is not far off when **Generative AI** will transform into AGI, which can mimic the capabilities of human beings.

Note

AGI vs. AI: What's the Difference?

- AGI is a subcategory of AI, and the former can be seen as an upgraded version of the latter:
  - Artificial intelligence is often trained on data to perform specific tasks or a range of tasks limited to a single context. Many forms of AI rely on algorithms or pre-programmed rules to guide their actions and learn how to operate in a certain environment.
  - Artificial general intelligence, on the other hand, is able to reason and adapt to new environments and different types of data. So instead of depending on predetermined rules to function, AGI embraces a problem-solving and learning approach — similar to humans. Because of its flexibility, AGI is capable of handling more tasks in different industries and sectors.

- **AIs, Gamechangers in Manipulating Electoral Behaviour:**
  - The global community is increasingly familiar with the use of AI models like ChatGPT, Gemini, and Copilot across various industries. However, 2024 is poised to demonstrate how newer AI models could significantly influence electoral behaviors and outcomes.
    - Underestimating the potential impact of AI on the electoral landscape would be a mistake. What may not materialise in 2024 could very well occur in the subsequent round of elections, both in India and around the world.
- **Promoting 'Deep Fake Elections':**
  - Employability of AI could well have a substantial impact to further confuse the electorate. As it is, many people are already referring to the elections in 2024 worldwide as the 'Deep Fake Elections', created by AI software.
    - Whether this is wholly true or not, the Deep Fake syndrome appears inevitable, given that each new election lends itself to newer and newer techniques of propaganda, all with the aim of confusing and confounding the democratic processes.

//

Fakes Around the World

ARGENTINA

Before the November 19, 2023 Argentina presidential election runoff, candidate Javier Milei posted a doctored image *(above)* of his rival Sergio Massa in Chinese communist military overalls. The allegedly AI-generated image got 3 mn views on X, a large number in a country of 46 mn. Milei is now Argentina's President.

INDIA

On January 21, 2024, the late M Karunanidhi 'told' a DMK youth wing meeting in Salem about the Centre's suppression of states' rights in an AI-generated video. Two days later, Karunanidhi, who died in 2018, appeared in another fake video *(screen-grab above)* and praised Chief Minister M K Stalin and DMK leader T R Baalu.

SLOVAKIA

In September 2023, before a key election, an audio surfaced online in which a top candidate was heard saying that he had bought the votes of a minority group, and that he would tax beer if voted to power. *AFP* fact checkers concluded that the audio was AI-generated. The candidate was eventually defeated.

- **Spreading Disinformation:**
  - The **World Economic Forum's (WEF's)** Global Risks Perception Survey, ranks misinformation and disinformation among the top 10 risks, with easy-to-use interfaces of large-scale AI models enabling a boom in false information and "synthetic" content - from sophisticated voice cloning to fake websites.
    - AI can be used to inundate voters with highly personalised propaganda on a scale that could make the Cambridge Analytica scandal appear microscopic, as the persuasive ability of AI models would be far superior to the bots and automated

social media accounts that are now baseline tools for spreading disinformation.
        - The risks are compounded by social media companies such as Facebook and Twitter significantly cutting their fact-checking and election integrity teams.
- **Inherent Inaccuracies in Such Models:**
    - The wide publicity given to a spate of recent inaccuracies associated with Google is a timely reminder that AI and AGI cannot be trusted in each and every circumstance. There has been public wrath worldwide over Google AI models, including in India, for their portrayal of persons and personalities in a malefic manner, mistakenly or otherwise. These reflect well the dangers of 'runaway' AI.
        - Inconsistencies and undependability stalk many AI models and pose inherent dangers to society. As its potential and usage increases in geometric proportion, threat levels are bound to go up.
- **Over Reliance on AIs:**
    - As nations increasingly rely on AI solutions for addressing their challenges, it becomes crucial to acknowledge what many AI experts refer to as AI's "hallucinations."
    - Specifically, concerning AGI, experts imply that it sometimes fabricates information to address novel issues. Such fabrications are often probabilistic and cannot be automatically deemed accurate. The implication of these factors is that excessive dependence on AI systems at this developmental stage could pose challenges.
- **Inherent Adversarial Capabilities:**
    - Various existential threats associated with AI can not be ignored. The dangers on this account pose an even greater threat than harm arising from bias in design and development.
        - There are real concerns that AI systems, oftentimes, tend to develop certain inherent adversarial capabilities. Suitable concepts and ideas have not yet been developed to mitigate them, as of now.
    - The main types of adversarial capabilities, overshadowing other inbuilt weaknesses are:
        - 'Poisoning' that typically degrades an AI model's ability to make relevant predictions;
        - 'Back Dooring' that causes the model to produce inaccurate or harmful results; and
        - 'Evasion' that entails resulting in a model misclassifying malicious or harmful inputs thus detracting from an AI model's ability to perform its appointed role.
- **Lack of Effective Regulation:**
    - India faces a dilemma in AI regulation. The Indian government itself has oscillated between a non-regulatory approach and a more cautious one, with an emphasis on mitigating user harm, which provides a fertile ground for misuse.
        - The argument against AI regulation is rooted in the pro-innovation stance, emphasising the need to promote and adapt to the rapid advancement of AI technologies rather than restrain their development and integration into society through regulatory measures.
- **Lack of Effective Checks by the Largest AI Platforms:**
    - Generative AI companies with the most popular visual tools prohibit users from creating "misleading" images. However, researchers with the British nonprofit Centre for Countering Digital Hate (CCDH), who tested four of the largest AI platforms - Midjourney, OpenAI's ChatGPT Plus, Stability.ai's DreamStudio, and Microsoft's Image Creator - succeeded in making deceptive election-related images more than 40% of the time.
        - According to a public database, users of Midjourney have created fake photos of Joe Biden handing wads of cash to Israeli Prime Minister Benjamin Netanyahu, and Trump playing golf with Russian President Vladimir Putin.

What are the Steps Required to Tackle the AI's Impact on Elections?

- **A Tech Accord to Combat Deceptive Use of AI in 2024 Elections:**
    - In February, at the Munich Security Conference, 22 of the companies, including tech giants Amazon, Google, Microsoft and Meta as well as AI developers and social platforms signed on to this Tech Accord, 2024 which pledged to address risks to democracy during this year of elections. **This was signed as a voluntary framework of principles and actions to advance seven principal goals:**
        - **Prevention:** Researching, investing in, and/or deploying reasonable precautions to

limit risks of deliberately Deceptive AI Election Content being generated.

- **Provenance:** Attaching provenance signals to identify the origin of content where appropriate and technically feasible.
- **Detection:** Attempting to detect Deceptive AI Election Content or authenticated content, including with methods such as reading provenance signals across platforms.
- **Responsive Protection:** Providing swift and proportionate responses to incidents involving the creation and dissemination of Deceptive AI Election Content.
- **Evaluation:** Undertaking collective efforts to evaluate and learn from the experiences and outcomes of dealing with Deceptive AI Election Content.
- **Public Awareness:** Engaging in shared efforts to educate the public about media literacy best practices, in particular regarding Deceptive AI Election Content, and ways citizens can protect themselves from being manipulated or deceived by this content.
- **Resilience:** Supporting efforts to develop and make available defensive tools and resources, such as AI literacy and other public programs, AI-based solutions (including open-source tools where appropriate), or contextual features, to help protect public debate, defend the integrity of the democratic process, and build whole-of-society resilience against the use of Deceptive AI Election Content.

- **Developing and Implementing Technology to Mitigate Risks:**
  - Supporting the development of technological innovations to mitigate risks arising from Deceptive AI Election Content by identifying realistic AI-generated images and/or certifying the authenticity of content and its origin, with the understanding that all such solutions have limitations.
  - Continuing to invest in advancing new provenance technology innovations for audio video, and images.
  - Working toward attaching machine-readable information, as appropriate, to realistic AI-generated audio, video, and image content that is generated by users with models in scope of this accord.
- **Appropriately Address Deceptive AI Election Content:**
  - Seeking to Appropriately Address Deceptive AI Election Content detected that is hosted on online distribution platforms and intended for public distribution, in a manner consistent with principles of free expression and safety.
  - This may include—but is not limited to—adopting and publishing policies and working to provide contextual information on realistic AI-generated audio, video, or image content .
- **Engaging with Global Civil Society:**
  - Continuing to engage with a diverse set of global civil society organisations, academics, and other relevant subject matter experts through established channels or events, in order to inform the companies' understanding of the global risk landscape as part of the independent development of their technologies, tools, and initiatives described.
- **Foster Public Awareness:**
  - Supporting efforts to foster public awareness and all-of-society resilience regarding Deceptive AI Election Content - for instance by means of education campaigns regarding the risks created for the public and ways citizens can learn about these risks to better protect themselves from being manipulated or deceived by this content.
    - Manipulations can happen via tools, interfaces, or procedures that can provide users with more useful context about the content than seen online; by developing and releasing open source tools to support others who try to mitigate these risks; or by otherwise supporting the work of organisations and communities engaging in responding to these risks.

What are India's Initiatives Related to Artificial Intelligence?

- **INDIAai.**
- **Global Partnership on Artificial Intelligence (GPAI).**
- **US India Artificial Intelligence Initiative.**
- **Responsible Artificial Intelligence (AI) for Youth.**
- **Artificial Intelligence Research, Analytics and Knowledge Assimilation Platform.**
- **Artificial Intelligence Mission.**

Conclusion

The rapid advancement of AI marks a significant milestone in human progress, potentially leading to the transformation of Generative AI into Artificial AGI, capable of emulating human capabilities. As the world prepares for a series of elections in 2024, including in India and numerous other countries, the implications of AI on electoral dynamics cannot be overlooked.

The use of AI, particularly its latest forms such as Generative AI, poses both opportunities and challenges for shaping electoral behaviours and outcomes. As AI's influence grows, it becomes imperative to address its disruptive potential, especially in the realm of elections, to safeguard democratic processes and uphold the integrity of electoral systems.

Drishti Mains Questions:

Discuss the implications of Artificial Intelligence (AI) on electoral processes, including its potential to influence voter behaviour and the challenges it poses to democratic principles.

UPSC Civil Services Examination, Previous Year Question (PYQ)

Q1. 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)
