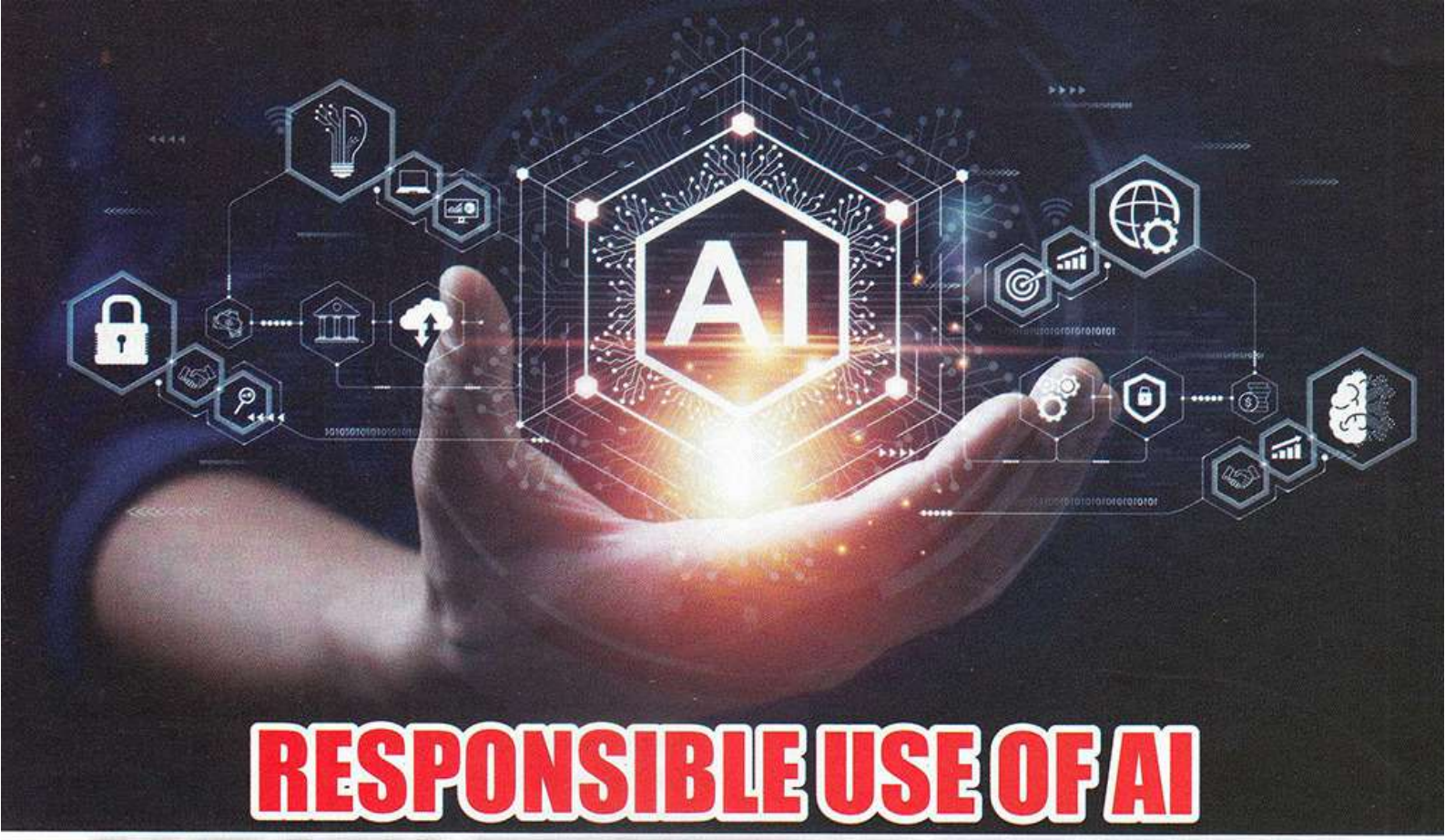


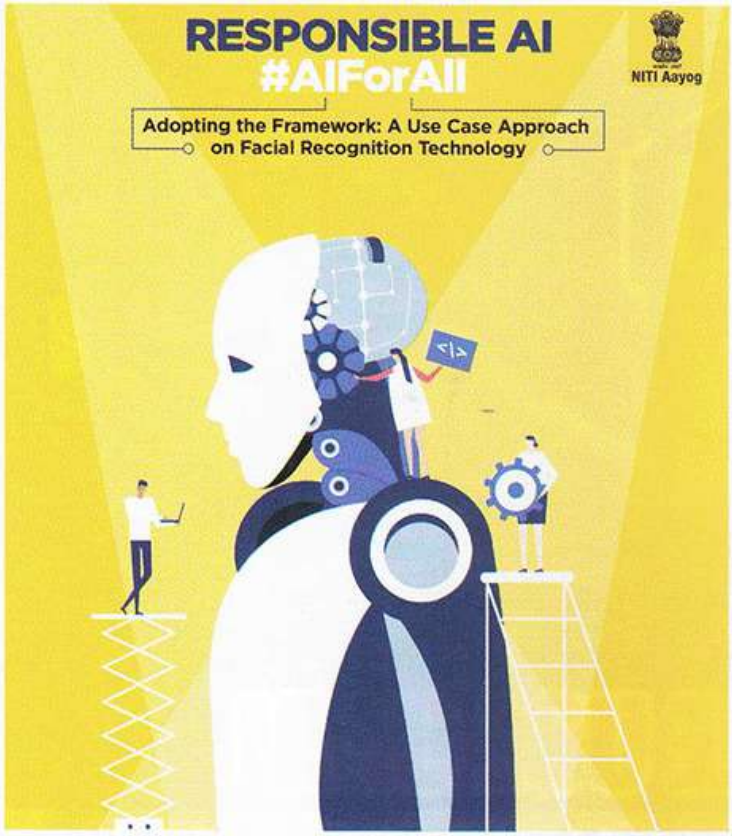
RESPONSIBLE USE OF AI
BRIDGING INNOVATION AND ETHICS

The New Delhi Leaders' Declaration highlights the significance of harnessing 'AI responsibly for good and for all'. It states that the G20 leaders are committed to leveraging AI for the public good by solving challenges in a responsible, inclusive, and human-centric manner while protecting people's rights and safety. Groupings like these are in an opportune position to take the lead in this regard, thereby bridging the gap between innovation and the ethics of the use of AI.

DR SAMEER PATIL	The author is a Senior Fellow and Deputy Director at the Observer Research Foundation (ORF). He works at the intersection of technology and national security. Email: sameer.patil@orfonline.org.
SHIMONA MOHAN	The co-author is a Junior Fellow at the Centre for Security, Strategy and Technology, ORF. She works at the intersection of security, technology (especially AI and cybersecurity), gender and disarmament. Email: shimona.mohan@orfonline.org.

Artificial Intelligence (AI) is transforming the way humans interact, industries function, and societies are structured. The seemingly limitless potential of AI across multiple domains, countries, and human imaginations has spawned numerous applications. Current applications include image and text analysis for data analysis purposes, logistics, assistance in decision-making, autonomous vehicles, and aerial systems, cybersecurity, etc. Additionally, it is being used for security, surveillance, and inventory management. It is also being applied extensively to areas like agriculture, fintech, healthcare, manufacturing, and climate change, yielding sizeable dividends in all of them.

It has become abundantly clear in the recent past that AI can augment human capabilities and aid us in tackling some of the most pressing challenges of our time. AI is a force that has the



capacity to create a more sustainable, equitable, and interconnected world. However, it also raises critical ethical and societal concerns, which require adequate policy consideration and responses. This highlights the need for the responsible development and deployment of AI to ensure that its transformative power benefits everyone and leaves no one behind.

G20 New Delhi Leaders' Declaration and Responsible AI

States are increasingly being compelled to practise responsible behaviour in their engagements with AI for civilian, security and defence purposes. In this context, the recently concluded G20 Summit in New Delhi (9-10 September 2023) has tackled multiple aspects related to Responsible AI (RAI). Most of the G20 members have been working towards establishing regulations for the responsible use of AI, especially since the advent of GenAI applications. The European Union's proposed AI Act is the most comprehensive attempt to establish a regulatory framework for the responsible development of AI that focuses primarily on strengthening rules around data quality, transparency, human oversight, and accountability.[1]

The New Delhi Leaders' Declaration highlights the significance of harnessing 'AI responsibly for good and for all'.[2] It states that the G20 leaders are committed to leveraging AI for the public good by solving challenges in a responsible, inclusive, and human-centric manner, while protecting people's rights and safety. It adds that to ensure responsible AI development, deployment and use, the protection of human rights, transparency and explainability, fairness, accountability, regulation, safety, appropriate human oversight, ethics, biases, privacy, and data protection must be addressed. In addition, the declaration mentions that the G20 members will pursue a pro-innovation regulatory/governance approach that maximises the benefits and takes into account the risks associated with the use of AI.

The declaration also reaffirms the leaders' commitment to G20 AI Principles of 2019. These principles had been adopted at the 2019 Osaka Summit and underline the human-centred approach of AI.[3] They take a cue from the Organisation for Economic Cooperation and Development principles on AI, also adopted in 2019, that support the technology to become innovative and trustworthy, and respect human rights and democratic values.[4] Besides this, the declaration also underlines the importance of investment in supporting human capital development. Towards this, G20 leaders agreed to extend support to educational institutions and teachers to enable them to keep pace with emerging trends and technological advances including AI. This will play an important role in imparting skills for the youth entering the job market and will offset the concerns around the adverse economic impacts of AI.

How does AI pose Ethical Risks?

According to the AIAAIC (AI, Algorithmic, and Automation Incidents and Controversies) database, which tracks incidents related to the ethical misuse of AI, the number of AI incidents and controversies has increased 26 times since 2012.[5]

Several critics of AI have also raised concerns about gender and racial bias when it comes to the application of AI to services like healthcare and finance. Although it may appear to be so, AI is not neutral; it can internalise and then catastrophically enhance biases that societies possess, programme them into the code, and/or ignore them in outputs

in the absence of sensitivities to those biases, to begin with.[6] If the datasets used in developing any AI system are incomplete or skewed towards or against a sub-group, they will produce results that marginalise those sub-groups or make them invisible in some way. Yet, even if a dataset is precise and representative of the intended population, biased Machine Learning (ML) algorithms applied to the data may still result in biased outputs.

In most supervised ML models, training datasets are given labels by a human developer or coder to enable the ML model to classify the information it already has. The model then characterises new information given to it based on this classification syntax, after which it generates an output. There are two possible modes of bias introduction in this process: first, if the human developers have their own biases, which they either introduce into the system or retain due to ignorant oversight; and second, if biases are incorporated in the processing of the data within the 'black box' of the AI/ML system, that is not explainable to or understandable by human operators.[7] The black box, as the name suggests, makes the learning process of the system opaque, and its algorithms can thus only be fixed once an output is generated and the human developer affirms that there was a problem with processing the input data.

Besides this, there are also ethical concerns that have arisen over issues like copyright infringement and privacy violations due to apps that create realistic images and art from a description in natural language.[8,9] Several artists have accused apps of training their algorithms based on images and illustrations scraped from the web without the original artists' consent.[10]

Then there are concerns regarding the misuse of AI in the defence domain to enhance targeting and surveillance capabilities of drones on the battlefield. This is a use-case of AI in drone warfare with the potential of ensuing violence. In other cases, critics have also noted the misuse of AI for illegal surveillance. In the cybersecurity sphere, generative AI applications are increasingly posing legitimate security threats as they are being used to conduct malware attacks. For instance, cybercriminals, with the help of AI, mass generating phishing emails to spread malware and collect valuable information. These phishing emails have higher rates of success, than manually crafted phishing emails. However, an even more insidious threat has emerged through 'deepfakes,' which generate synthetic or artificial media using ML. Such realistic-looking content is difficult to verify and have become a powerful tool for disinformation, with grave national security implications. For instance, in March 2022, a deep fake video of Ukrainian President Volodymyr Zelenskyy asking his troops to surrender went viral among Ukrainian citizens, causing significant confusion, even as their military was fighting against the Russian forces.[11]

Beyond defence and security, AI has also evoked fears of adverse economic impact. An emerging apprehension is that AI automation could potentially alter the labour market in a fundamental manner, with grave implications for economies in the Global South that rely on their labour and human resources.[12,13]

What is Responsible AI?

These dynamics have created the necessity for the 'Responsible AI' (RAI) and the need to regulate it. There has been a gradual momentum around rallying for responsible innovation ecosystems. This is especially valid in the development and deployment of AI, where there is a chance for responsible innovation and use to be institutionalised right from the get-go and not as an afterthought or a checkbox to performatively satisfy policy and/or compliance-

related constraints. In this context, RAI is broadly understood as the practice of designing, developing, and deploying AI to empower employees and businesses and impact society in a fair manner. Given AI's dual-use character, this is a loose and flexible understanding, and it posits RAI as an umbrella term that usually encompasses considerations around fair, explainable, and trustworthy AI systems.

India has been working on RAI since 2018, and NITI Aayog also released a two-part report in 2021 on approaches towards[14] and operationalisation of[15] RAI principles for the deployment and use of civilian AI architectures. The seven principles that NITI Aayog highlights are: safety and reliability; equality; inclusivity and non-discrimination; privacy and security; transparency; accountability; and protection and reinforcement of positive human values. It also recommends measures for the government, industry bodies, and civil society to implement these principles in the AI products they develop or work with. Indian tech industry body NASSCOM embedded the principles of this framework into India's first RAI Hub and Toolkit[16] released in late 2022, which comprises sector-agnostic tools to enable entities to leverage AI by prioritising user trust and safety.

Pertinently, the focus on RAI in G20 New Delhi Leaders' Declaration also aligns with India holding the chair of the Global Partnership on Artificial Intelligence (GPAI), a multistakeholder initiative that brings together experts from science, industry, civil society, international organisations, and governments.[17] It contributes to the responsible development of AI via its Responsible AI working group.[18]India chairing the GPAI is important since the Global South is underrepresented in the forum: out of its 29 members, only four are from the Global South - Argentina, Brazil, India, and Senegal. Therefore, India is better positioned to play an active role in bridging this divide and ensuring that the less developed economies also get to reap the benefits of this technological shift towards AI. New Delhi will host the annual GPAI Summit on 12-14 December 2023. .At the last year's summit in Tokyo, India urged the members to work together on a common framework of rules and guidelines on data governance in order to prevent user harm and ensure the safety of both the internet and AI.

Conclusion

Though the rise of AI and its applications in the past few years has been meteoric and the scope for innovation in the field is endless, nations all around the world are waking up to the dangers of its potential misuse. While there are several initiatives attempting to address the issue, there is currently no global consensus or regulatory framework on the ethical and responsible use of AI. Hence, groupings like the G20 and GPAI are in an opportune position to take the lead in this regard, thereby bridging the gap between innovation and the ethics of AI use. The G20 New Delhi Leaders' Declaration demonstrates that leaders of the world's largest economies are aware of the potential benefits and risks of AI and are committed to working together to ensure that the technology is developed and used in a responsible and inclusive manner. The G20 members must follow this declaration by adopting the anticipatory regulation approach, doing over-the-horizon thinking, and building a coalition of diverse stakeholders. ❑

References

1. "EU AI Act: first regulation on artificial intelligence," June 14, 2023, https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.
2. G20 New Delhi Leaders' Declaration, September 9-10, 2023, https://www.g20.org/content/dam/gtwenty/gtwenty_new/document/G20-New-Delhi-Leaders-Declaration.pdf.
3. G20 AI Principles, https://www.mofa.go.jp/policy/economy/g20_summit/osaka19/pdf/documents/en/annex_08.pdf.
4. "OECD AI Principles overview," https://oecd.ai/en/ai-principles.
5. Artificial Intelligence Index Report 2023, Stanford University Human-Centred Artificial Intelligence, https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report-2023_CHAPTER_3.pdf.
6. Shimona Mohan, "Filling the Blanks: Putting Gender into Military A.I.," ORF Issue Brief No. 655, August 2023, Observer Research Foundation, https://www.orfonline.org/research/filling-the-blanks-putting-gender-into-military-ai/.
7. Shimona Mohan, "Gender-ative AI: An enduring gender bias in generative AI systems,"Observer Research Foundation, April 27, 2023, https://www.orfonline.org/expert-speak/gender-ative-ai/.
8. DALL.E2, https://openai.com/dall-e-2.
9. Midjourney, https://www.midjourney.com/home/.
10. James Vincent, "AI art tools Stable Diffusion and Midjourney targeted with copyright lawsuit," *The Verge*, January 16, 2023, https://www.theverge.com/2023/1/16/23557098/generative-ai-art-copyright-legal-lawsuit-stable-diffusion-midjourney-deviantart.
11. The Telegraph, "Deepfake video of Volodymyr Zelensky surrendering surfaces on social media," March 17, 2022, https://www.youtube.com/watch?v=X17yrEV5sl4.
12. Ian Shine and Kate Whiting, "These are the jobs most likely to be lost – and created – because of AI," *World Economic Forum*, May 4, 2023, https://www.weforum.org/agenda/2023/05/jobs-lost-created-ai-gpt/.
13. Accenture, "A new era of generative AI for everyone," https://www.accenture.com/content/dam/accenture/final/accenture-com/document/Accenture-A-New-Era-of-Generative-AI-for-Everyone.pdf.
14. NITI Aayog, "RESPONSIBLE AI #AIFORALL: Approach Document for India Part 1 – Principles for Responsible AI," February 2021, https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf.
15. NITI Aayog, "RESPONSIBLE AI #AIFORALL: Approach Document for India: Part 2 - Operationalizing Principles for Responsible AI," August 2021, https://www.niti.gov.in/sites/default/files/2021-08/Part2-Responsible-AI-12082021.pdf.
16. INDIAai, "NASSCOM launched the Responsible AI hub and resource kit," October 11, 2022, https://indiaai.gov.in/news/nasscom-launched-the-responsible-ai-hub-and-resource-kit.
17. Prateek Tripathi, "India's chairmanship of the Global Partnership on AI," Observer Research Foundation, August 8, 2023, https://www.orfonline.org/expert-speak/indias-chairmanship-of-the-global-partnership-on-ai/.
18. The Global Partnership on Artificial Intelligence, "Working Group on Responsible AI," https://gpai.ai/projects/responsible-ai/.