‘Just Right’ Robot Reg

Amitabh Kant & Ajita Agarwala

Recently, computer scientist Geoffrey Hinton, who played a key role in developing AI tools, warned that AI systems could become more intelligent than humans and outsmart them in five years. Earlier this year, OpenAI CEO Sam Altman, admitted his worst fear that advanced AI technology could cause ‘significant harm’ to the world if guard rails are not put in place. He warned that the company’s latest AI language model, GPT-4, could be used for launching cyber-attacks and spreading disinformation. He suggested:

▶ Licensing and testing requirements for AI above a threshold of capabilities.

▶ Standardised safety requirements and global collaboration on guidelines.

▶ Regulating AI should be ‘something between the traditional European approach and the traditional US approach’. He, however, cautioned that overregulation could stifle the emerging technology.

As AI advances at a scorching pace, it is clear it could disrupt society. Imagine a future where AI systems influence voting behaviours, corrupt autonomy and create more biases than ever. Due to such fears, ChatGPT is banned in China, Iran, Russia and Italy. Countries like the US, India and Australia and regions like the EU are evaluating regulatory options to ensure ethical usage without stifling innovation.

On June 14, the EU passed the AI Act, the world’s first comprehensive AI law. It aims to ensure a human-centric and ethical development of AI in Europe. The regulation will follow a risk-based approach, prohibiting AI practices that pose an unacceptable level of risk such as subliminal manipulation, exploitation of vulnerabilities and social scoring. To ensure high-risk AI systems are held in check, the EU has banned biometric surveillance, and recognition and predictive policing AI systems. The member-states are ensuring that AI technology complies with the EU General Data Protection Regulation (GDPR).

The US National Institute of Standards and Technology’s (NIST) AI Risk Management Framework (AI RMF) provides a road map for organisations to manage responsible AI while identifying risks. The framework suggests key characteristics of trustworthy AI, including accuracy, explainability, reliability, privacy, security, safety and resilience. In October 2022, the US also released its Blueprint for an AI Bill of Rights, which aims to protect individuals’ rights, opportunities and access to critical resources and services in various contexts, including civil rights, equal opportunities and essential services.

Amid these regulatory efforts, achieving a consensus and shared understanding of the global challenge posed by General Purpose AI models is essential. Such regulation should go beyond chatbots and language models and encompass a range of AI technologies, including image and speech recognition, pattern detection and translation. Addressing the risks of General Purpose AI models requires regulations throughout the product cycle, from data collection to model development, testing and evaluation.

Dismissing legal disclaimers that absolve developers of responsibility is crucial, as it places an undue burden on downstream actors who may lack the means to mitigate risks. Furthermore, broad consultation with stakeholders, including civil society, researchers and non-industry participants, is necessary to ensure comprehensive evaluation practices and standardised documentation.

Drive AI carefully

FILE PHOTO

The G20 New Delhi Leaders’ Declaration is committed to the G20 AI Principles of 2019, emphasising their dedication to using AI for the greater good while upholding human rights, transparency, fairness, accountability and safety. Under India’s presidency, G20’s aim has been to promote a pro-innovation regulatory and governance framework that maximises the benefits of AI while managing its risks. By doing so, G20 is paving the way for responsible AI.

India showcased its innovative government-AI platform Bhashini at the summit. Bhashini aims to break the barrier between Indian languages by using technology, enabling underserved communities to access state services. This initiative reflects India’s commitment to leveraging AI for the betterment of society and promoting inclusivity in accessing essential services.

International initiatives like the Hiroshima AI Process and the Global Partnership on Artificial Intelligence (GPAI) are also discussing governance, responsible utilisation and safeguarding of AI. Japan’s G7 presidency established the Hiroshima AI Process, addressing generative AI and IPR while countering foreign information manipulation. India, also the current chair of GPAI, is leading efforts to promote responsible and human-centric development and use of AI.

As society grapples with AI regulation, reconciling innovation with the need to mitigate potential risks looms large. AI’s success will depend on collaborative and international efforts that prioritise human-centric values and establish ethical and accountable AI practices. Achieving this balance will require thoughtful deliberation, robust regulation and a dialogue to unlock AI’s full potential, and preserve the values and integrity we hold dear.

Kant is sherpa, and Agarwala is under-secretary, G20, GoI