Business Standard MUMBAI | M

GUEST COLUMN

LALITAGAURI KULKARNI

Time to evaluate AI's risks to institutions

The Reserve Bank of India's "Financial Stability Report" has identified artificial intelligence (AI) as a new form of operational risk. It comes at a time when frontier AI is rapidly evolving.

In April, Anthropic unveiled Claude Mythos Preview. Interestingly, instead of making the model immediately available to the general public, Anthropic introduced it as a gated research preview, with access initially restricted to select organisations and defensive cybersecurity applications. The decision's significance became clearer in June, when the United States imposed export controls on Claude Fable 5 and Claude Mythos 5, requiring Anthropic to restrict access by foreign nationals amid national security concerns. Anthropic subsequently suspended access to the models. Less than two weeks later, access was restored to more than 100 trusted US companies and institutions after Anthropic strengthened its safeguards, improved monitoring, and agreed to closer engagement with US authorities.

The episode is a signpost of how regulators may need to monitor, supervise and, where necessary, regulate AI in the future.

Frontier AI refers to the most advanced and capable AI models, including the latest variants of Claude, GPT and Gemini. These models can not only make work more efficient but can also make it easier, faster and cheaper for attackers to discover and exploit vulnerabilities. It is a key reason why central banks are increasingly flagging this emerging threat.

TO PROTECT CRITICAL DIGITAL INFRASTRUCTURE, INDIA MUST BUILD ROBUST INSTITUTIONS TO EVALUATE RISKS FROM FRONTIER AI MODELS

What makes the development particularly concerning is the concentration of advanced cyber capabilities in a single frontier AI provider. It raises an important regulatory question: Should access to such powerful systems be left entirely to market forces, or should it be subject to government oversight and, where necessary, licensing requirements?

In Europe, there is growing emphasis on building sovereign and independent AI infrastructure to reduce dependence on foreign technology providers. France has increased support for domestic AI system developers, while the UK has strengthened its AI Security Institute to facilitate independent evaluation of advanced models before deployment.

These developments could have serious repercussions for India. Models become a matter of economic and strategic power, not merely technology. The Mythos episode has shown how quickly the emergence of a single high-risk model can trigger global concern. The lesson for India is to build the institutional capacity to assess them and determine when they are safe enough for its conditions.

One has to note that when institutions operating Unified Payments Interface, telecom networks, health care systems, and public digital platforms become dependent on foreign frontier AI models, they start becoming infrastructure rather than merely software. In such a scenario, India could consider an independent frontier AI evaluation institution, along the lines of the UK's AI Security Institute.

AI systems are vulnerable to errors, bias and unexpected behaviour. What changes with frontier AI is the scale and speed at which these risks can materialise, particularly when highly capable models are connected to critical systems. Frontier AI can therefore transform what was once an individual institution's technology risk into a broader question of operational resilience, financial stability, and national technological resilience.

However, there are two additional risks that India must not overlook.

The first is provider and concentration risk. As banks, hospitals, telecom companies and other critical institutions increasingly depend on a small number of global AI providers, a disruption in one provider can potentially affect a large number of institutions simultaneously. The second is sovereignty and geopolitical risk. When critical functions depend on AI models developed and controlled outside the country, access to those models can ultimately be influenced by decisions made elsewhere.

India has recognised these emerging risks while focusing on developing indigenous AI models. It is also evaluating and planning to deploy global models such as DeepSeek on domestic infrastructure. India's AI journey will, therefore, depend not only on its ability to develop and deploy powerful models, but equally on its ability to build and sustain trust through institutions and regulatory frameworks.

The writer is associate dean of research & publications, Gokhale Institute of Politics and Economics. Arjun Goyal, assistant professor, contributed to the column