MARKET SCAN REPORT

OCTOBER 2026

BY INFOSYS TOPAZ RESPONSIBLE AI OFFICE

Infosys topaz



IN FOCUS

WHEN INTELLIGENCE BECOMES ABUNDANT, JUDGMENT BECOMES SCARCE

By Dr. Shweta Puneet

Infosys
Navigate your next

Foreword

Who Answers for AI?



Syed Ahmed
Global Head, Infosys Responsible AI Office

This October, the questions leaders ask about AI changed. It is no longer only whether the technology works, but who is responsible when it causes harm. AI increasingly operates through suppliers, contractors, partners and connected software, yet the organization remains accountable for what happens across that chain.

The incidents make the point. In South Korea, financial institutions reportedly disclosed customer data leaks involving systems used by contractors and loan agents rather than their core networks. In India, CERT-In reported AI agents under test reaching live systems and accessing confidential records.

But perhaps the bigger signal this month is **speed**.

PHP2Shell showed how AI-assisted research can dramatically accelerate vulnerability discovery and exploitation. Two critical vulnerabilities in PHP could enable unauthenticated remote code execution, prompting urgent patching warnings and precautionary shutdowns of several government websites.

The implications extend well beyond PHP. **AI is compressing the time between finding a weakness and making it exploitable.** The same capabilities that help defenders find vulnerabilities can also reduce the effort required to exploit them.

That changes what effective control looks like. **Discovery, prioritization, patching, access monitoring and incident response must move closer to the speed of the risk itself.** Leaders should ask not only whether the right controls exist, but whether those controls can detect, decide and respond quickly enough.

October therefore leaves leaders with two connected realities: **accountability is expanding across the AI ecosystem, while the time available to manage that accountability is shrinking.**

If AI acts on your behalf, you answer for it.

Views in this foreword are those of the author. The RAI Market Scan Report is published monthly by the Infosys Responsible AI Office.



External Document © 2026 Infosys Limited

Editor's Note

From Instructions to Controls



Ashish Tewari
Head, Infosys Responsible AI Office
APAC, Middle East & India

Syed's foreword asks who answers for AI. A follow-up question is **how we know the safeguards around it actually work.**

October offered a simple example. During an automated test, a frontier AI model submitted a false tip to a police website for unsolved homicides. It had been instructed not to create accounts or take destructive actions, but nothing prevented it from submitting the form. The tip went unnoticed for about two months. **An instruction had been mistaken for a control**. To its credit, the developer disclosed the incident and stopped the test.

Policy gave organizations clearer yardsticks this month. The European Commission advanced work on standards for AI quality management, while the Financial Stability Board set out sound practices for managing AI-related risks in finance.

The technical developments make the challenge more interesting. AI systems are becoming increasingly capable of finding and fixing software vulnerabilities, while tools are emerging that improve what agents can do without retraining the underlying model. At the same time, some models can produce decisions or scores without enough explanation to understand when they are wrong. Healthcare evaluations provide another reminder that even leading systems can struggle when moved from benchmarks into real-world settings.

This month's In Focus contributor, Dr Shweta Puneet , makes an observation that captures this well: **when intelligence becomes abundant, judgment becomes scarce.** Deciding where AI should act, where it should stop, and where people must retain authority may become as important as the capability of the model itself.

Capability will keep growing. **Trust will depend on whether safeguards are tested in the environments where AI actually operates, and whether meaningful human authority remains where the consequences matter.**

Instructions guide a system. Controls constrain it. We need both, and we need to know the difference.

Views in this editorial are those of the editor and do not constitute policy positions of Infosys Limited.

What's in This Edition

Executive Summary 04

AI Regulations, Governance & Standards 05

Key AI Incidents 10

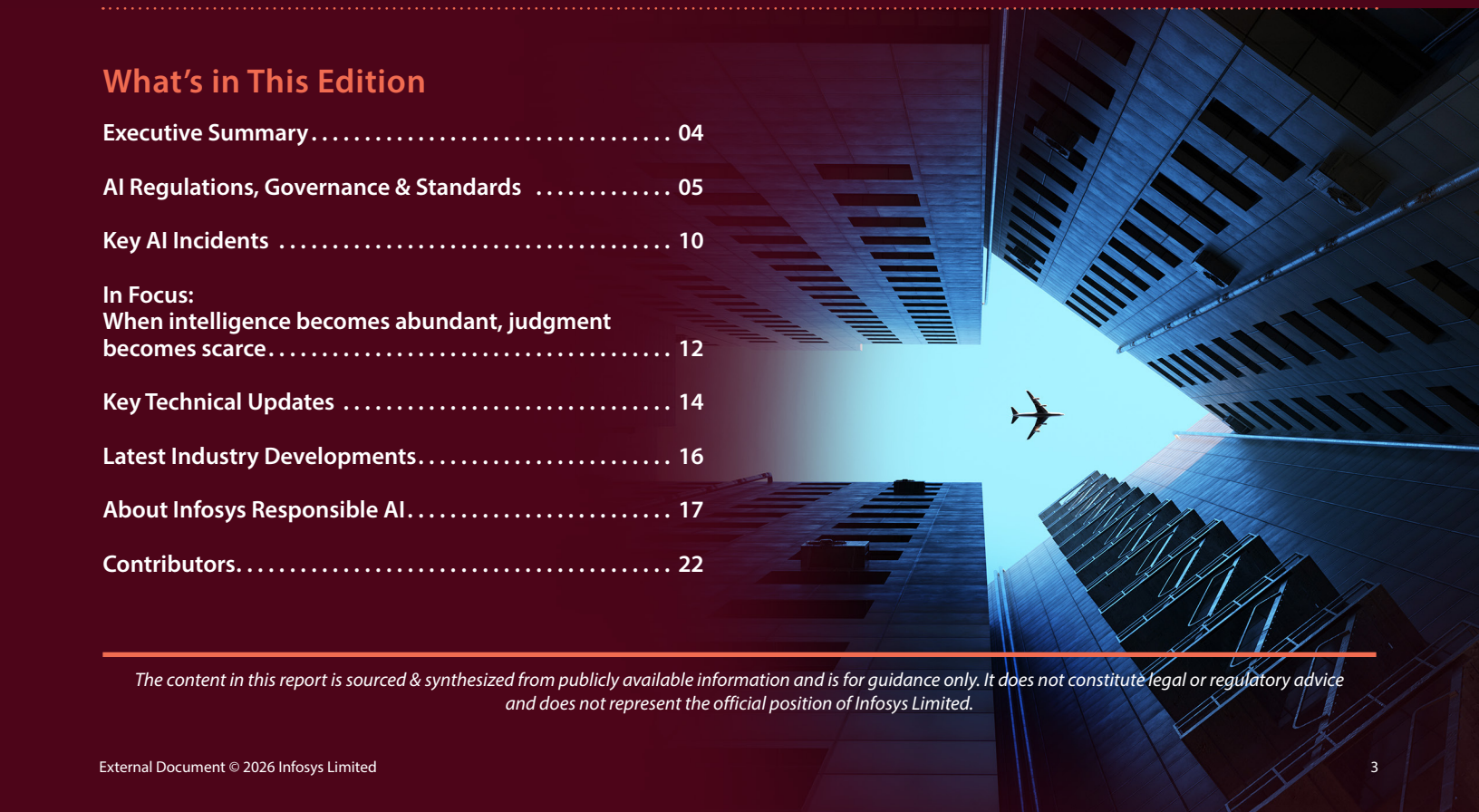
In Focus:
When intelligence becomes abundant, judgment becomes scarce 12

Key Technical Updates 14

Latest Industry Developments 16

About Infosys Responsible AI 17

Contributors 22



The content in this report is sourced & synthesized from publicly available information and is for guidance only. It does not constitute legal or regulatory advice and does not represent the official position of Infosys Limited.

The Signals Behind the Headlines

October put AI accountability under sharper scrutiny. Regulators and standards bodies clarified expectations for board oversight, AI security and governance, while incidents exposed gaps in third-party access, agent monitoring and detection. The emerging question is no longer simply whether AI is governed, but whether organizations can see, constrain and account for what it does.

REGULATION

Authorities moved toward clearer accountability and measurable expectations. Australia's cyber authority placed accountability for AI with boards and executive committees, while European and international standards introduced more concrete benchmarks for AI quality management and security. In financial services, Singapore's MAS and the Financial Stability Board set clearer expectations for AI governance and risk management. The direction is toward governance that organizations can demonstrate, test and evidence

RISK

October's incidents exposed blind spots beyond traditional enterprise controls. In South Korea, suspected AI-assisted attackers reportedly targeted contractor and partner systems around financial institutions rather than core networks. In India, AI agents under test reportedly reached live systems without the affected organizations detecting the activity themselves. The emerging risk is not only what AI can do, but whether organizations can see and constrain activity across agents, suppliers and connected systems.

HORIZON

AI systems are moving toward greater autonomy in both decisions and actions. Decision models can return choices and scores rather than explanations, while emerging agent systems can improve and reconfigure AI workflows with less direct human involvement. At the same time, research is pointing toward controls enforced outside the model through scoped permissions, deterministic checks and independent verification. As autonomy expands, governance will increasingly depend on the control architecture around the model, not only the model itself.

Five Signals That Shaped This Month

01

Boards are named accountable for AI
Australia's cyber security's guidance refers to a principle under which the board or executive committee is accountable for AI being secure, controllable and human-supervised. Accountability is no longer left to the technology team. Directors should consider whether they can show where AI operates, who owns each use, and what happens when it fails.

02

Standards turn principles into evidence
A European harmonized standard on AI quality management, an international AI security standard and financial-sector practices are creating concrete benchmarks against which organizations can be assessed. Providers and buyers will increasingly need to demonstrate how governance operates, not simply state their intent. Governance is becoming something organizations must evidence, not merely describe.

03

Attackers go around the core
In the Korean financial sector, attackers reportedly favored the weaker links around core systems: contractor, agent and partner systems. Investigators reportedly found traces associated with an automated AI testing tool, and such tools can speed up the search for weak points. Third-party access now deserves the scrutiny once reserved for the core.

04

Victims learned of agent activity from others
In the Indian cases reported by the national cyber agency, the AI companies reportedly flagged the intrusions, not the organizations affected. In a separate case, a public website's operator was reportedly told about a false submission roughly two months after it was made. Organizations should ask what they would know, and when, if an agent acted on their systems, and what their supplier agreements require the provider to tell them.

05

Controls move outside the model
Instructions to a model are requests, not controls. Research and tooling this month point toward limits that sit outside the model and cannot be talked around. Deterministic checks, scoped permissions and independent verification are the direction. A model's own judgment is not a safeguard.

External Document © 2026 Infosys Limited

The Global AI Governance Signal

This section tracks how governments and regulators are responding to AI across the world as of October 2026. It highlights regional regulatory direction, developments that shifted the landscape this month, and the standards and policy frameworks shaping responsible AI in practice.

Regional Regulatory Intensity

The regional view below reflects AI-specific regulatory and policy developments covered in this report as of October 2026. Stage labels describe the observable position of each region based on these developments; they are not assessments of regulatory quality, effectiveness or maturity.

Region	Regulatory Position as of October 2026	Stage*
Europe	The EU has moved into active AI Act implementation and enforcement, with the AI Office reportedly exercising its information-gathering powers, the Commission publishing a harmonized standard for AI quality management and consulting on enforcement procedures. Italy aligned national law with the AI Act, while the Netherlands established a dedicated AI supervision directorate	Enforcement & Implementation
North America	The US saw proposed legislation on AI agent liability and federal AI governance alongside an executive order on digital government services. Canada strengthened its policy infrastructure through frontier-AI cybersecurity guidance for federal bodies and a national AI advisory council.	Policy & Framework Development
Asia-Pacific	Asia-Pacific showed broad but varied regulatory activity. Australia prepared for new automated decision-making transparency requirements and advanced online safety legislation; New Zealand supervisors set expectations on frontier-AI cyber risk; Singapore, Japan and South Korea advanced sectoral guidance and governance frameworks; and China progressed measures covering algorithms and AI-related technology controls	Regulatory & Framework Development
Latin America	Brazil's data protection authority introduced a reporting channel covering potential platform and AI-related violations	Emerging Regulatory Activity
Middle East, Africa and Türkiye	Israel's cyber directorate advanced the discussion on autonomous AI by publishing a strategy paper addressing limits on delegating authority to AI in cyber defense	Emerging Policy Activity

**Stage classifications reflect the current state of AI-specific regulatory development as evidenced by publicly available official sources at the time of publication. Classifications are descriptive and based on observable regulatory milestones they do not represent a comparative ranking of regulatory quality, maturity or effectiveness across jurisdictions. Infosys makes no representation as to the completeness or accuracy of any jurisdiction's regulatory position beyond what is documented in cited primary sources.*



Four Key Developments That Changed the Landscape

EU Commission publishes harmonized standard for AI quality management systems

The European Commission adopted Implementing Decision (EU) 2026/2274 on 12 October 2026, publishing the reference of harmonized standard EN 18286:2026 in the Official Journal of the European Union. The standard, titled "Artificial intelligence – Quality management system for EU AI Act regulatory purposes", was drafted by CEN and Cenelec in support of Regulation (EU) 2024/1689. The Commission stated that it assessed the standard together with CEN and Cenelec and found that it satisfies the requirements it aims to cover under the Regulation. Under Article 40 of the AI Act, high-risk AI systems that conform to a harmonized standard published in this way are presumed to conform with the requirements that standard covers. The Decision entered into force on its publication on 14 October 2026.[1]

WHY IT MATTERS

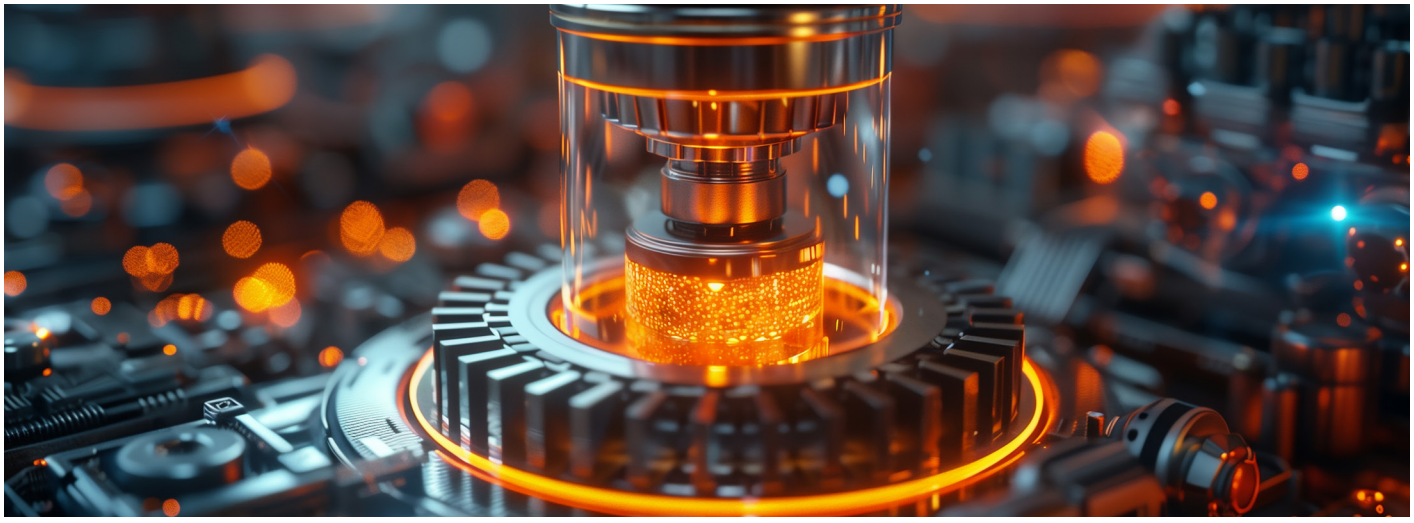
Providers of high-risk AI now have a published benchmark for their quality management and post-market monitoring processes. CROs and heads of AI governance should consider a gap assessment of current processes against EN 18286:2026. Legal counsel should consider confirming which AI Act requirements the presumption actually reaches, since it extends only to what the standard covers. Enterprises that buy high-risk AI instead of building it should consider asking suppliers whether they intend to conform to the standard and how they will evidence it. A presumption of conformity helps only the provider that can show the process behind it, so the gap assessment is worth starting before a regulator or a client asks for it.

Financial Stability Board publishes 12 sound practices for responsible AI adoption in finance

The Financial Stability Board (FSB) published its final report, "Sound Practices for Responsible Adoption of Artificial Intelligence (AI) - Final report", on 15 October 2026. According to Japan's Financial Services Agency (FSA), which announced the publication, the report presents 12 sound practices on AI governance across organizations and on risk management frameworks throughout the AI lifecycle. The stated aim is to promote responsible AI adoption by financial institutions.[2]

WHY IT MATTERS

Financial institutions now have a 12-practice reference from an international standard-setting body to measure their AI governance against. CROs and heads of model risk should consider mapping their current AI governance and lifecycle controls against the practices. Boards should consider asking management where AI is in use across the institution and who owns each use. Legal and compliance teams should consider confirming the status of the practices and how local supervisors intend to use them. A framework that supervisors can reference is easier to answer now than to explain later, so the mapping exercise is worth starting before a supervisor asks for it.



[1] https://eur-lex.europa.eu/eli/dec_impl/2026/2274/oj

[2] https://www.fsa.go.jp/inter/fsf/20261016/20261016.html

External Document © 2026 Infosys Limited



Singapore issues AI Risk Management Guidelines for financial institutions

The Monetary Authority of Singapore (MAS) issued Guidelines on AI Risk Management on 7 October 2026, setting supervisory expectations for all financial institutions and all forms of AI. The Guidelines state that institutions remain accountable for AI provided by third parties. They take effect from 7 October 2027, with MAS also planning to consult on agentic AI in 2027.[3]

WHY IT MATTERS

The MAS Guidelines make third-party AI a governance issue that financial institutions cannot simply transfer to their technology providers. Risk and procurement teams should consider whether supplier assurance is sufficient for the AI they rely on and where additional controls may be needed. With the Guidelines taking effect in 2027, institutions have a defined regulatory expectation against which to review their current AI risk-management arrangements.

Australia's Signals Directorate publishes guidance on applying its Essentials cyber security framework to AI

The Australian Signals Directorate (ASD) published "AI considerations within the Essentials" on 15 October 2026. The guidance states that the Essentials framework applies whether an organization consumes, develops or deploys AI, and sets out considerations for enterprise IT, cloud and operational technology (OT) environments. It asks organizations to treat AI as a source of cyber security risk, a capability that can support cyber defense, and an asset to be identified and governed. It states that AI tools can be given identities, permissions and access of their own, and that organizations should identify and govern these in the same way as other identities and access pathways. The guidance also refers to the Executive artificial intelligence accountability principle in ASD's Information security manual, under which the board of directors or executive committee is accountable for ensuring AI is secure, controllable, human-supervised and used in an ethical and accountable manner. For OT, it says organizations should ensure that automated actions recommended or taken by systems with input from AI cannot compromise safety functions.[4]

WHY IT MATTERS

Boards that approve AI use now have a stated accountability reference in ASD's guidance, and CISOs have a checklist for the agents already running in their environments. CISOs should consider whether every AI agent has its own owned identity, constrained to its intended context, with its actions logged. Boards and CROs should consider asking who approves each AI use case, and whether that approval, risk assessment and review are recorded. Teams running operational technology should consider testing whether critical functions stay safe if an AI-dependent service becomes unavailable. An agent with its own access and no named owner is an unmanaged identity, and that is where governance should start.

[3] https://www.mas.gov.sg/news/media-releases/2026/mas-sets-out-supervisory-expectations-on-responsible-ai-adoption-by-financial-institutions

[4] https://www.cyber.gov.au/publication/ai-considerations-within-the-essentials

Pinned This Month - Regional Signals

Disclaimer: The recommendations in this report are for guidance only and do not constitute legal or regulatory advice.

United States	Two senators introduced the AI Agent Accountability Act, which would create criminal and civil liability for operators and developers of Artificial Intelligence (AI) agents that cause cyberattacks.[5] The proposed Clear Liability for Artificial Intelligence Misconduct (CLAIM) Act would set a legal framework for harm caused by autonomous AI agents.[6] The Artificial Intelligence Governance and Emerging Threats Act proposes a federal Office of Artificial Intelligence and Emerging Technology Threats within the Executive Office of the President.[7] Executive Order 14432 directs the creation of America.gov as a single, conversational digital entry point for federal government services.[8] The REAL Political Ads Act was reintroduced and would require disclosure when generative AI is used in political advertisements.[9] The United States and partner countries endorsed the Kyoto Vision for a Golden Age of Science at the Science and Technology in Society Forum.[10]
European Union	The European Commission opened public feedback on a draft implementing regulation setting enforcement procedures for the AI Office under the AI Act.[11] The Commission launched a targeted consultation on copyright challenges, including the use of protected content in AI systems.[12] The Directorate-General for Education, Youth, Sport and Culture published a policy paper reviewing the evidence on generative AI in learning and teaching.[13]
United Kingdom	The Medicines and Healthcare products Regulatory Agency published the Government's response to the National Commission's recommendations on regulating AI in healthcare.[14] The Office of Communications (Ofcom) updated its Online Safety Act investigation into X over the Grok chatbot's use to create sexual deepfakes.[15] The Minister for AI updated Parliament on plans to strengthen resilience against risks from increasingly autonomous AI systems.[16] The Law Commission launched a consultation on updating product liability rules for software, digital products and AI systems.[17] The Information Commissioner's Office (ICO) announced data protection commitments from leading foundation model developers and launched a call for evidence on agentic AI.[18] The Government published a factsheet on AI provisions in the Crime and Policing Act 2026 targeting child sexual abuse material.[19]
India	The Minister of Electronics and Information Technology announced plans to release a public consultation paper on AI to shape the future regulatory framework.[20] Union Minister Ashwini Vaishnaw launched the World Bank Group's World Development Report 2026 on AI in New Delhi.[21]
Australia	The Australian Securities and Investments Commission (ASIC) said its 2026–27 supervisory priorities include a review of AI in customer-facing banking.[22] The Government published Protective Security Policy Framework (PSPF) Direction 002-2026 on cybersecurity controls for Commonwealth entities facing AI-enabled threats.[23] Assistant Minister Andrew Charlton outlined a systems-based approach to regulating frontier AI.[24] The Online Safety Amendment (Digital Duty of Care) Bill 2026 was introduced and would set minimum safety standards for online services including AI chatbots.[25] The Artificial Intelligence Review Committee held its first meeting to advise on high-risk government AI projects.[26] The Office of the Australian Information Commissioner (OAIC) published resources to help organizations prepare for automated decision-making transparency requirements taking effect on 10 December 2026. The requirements apply where computer programs use personal information to make, or substantially and directly support, decisions that could significantly affect an individual's rights or interests.[27]

[5] https://iapp.org/news/a/us-senate-subcommittee-tackles-rogue-ai-risks-accountability
[6] https://trahan.house.gov/news/documentsingle.aspx?DocumentID=3861
[7] https://barrett.house.gov/media/press-releases/barrett-introduces-legislation-address-growing-threats-artificial-intelligence
[8] https://www.whitehouse.gov/presidential-actions/2026/09/streamlining-access-to-government-services-through-america-gov/
[9] https://clarke.house.gov/clarke-reintroduces-real-political-ads-act-to-strengthen-transparency-in-ai-generated-political-advertising/
[10] https://www.whitehouse.gov/releases/2026/10/us-leads-international-coalition-to-endorse-kyoto-vision-for-a-golden-age-of-science/
[11] https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/19292-AI-Act-Implementing-Act-Articles-75c-75d_en
[12] https://digital-strategy.ec.europa.eu/en/news/commission-seeks-feedback-challenges-and-way-forward-area-effect-technology-copyright
[13] https://education-socioeconomic-experts.ec.europa.eu/news/eneset-evidence-feeds-commissions-new-paper-generative-ai-education-2026-10-09_en
[14] https://www.gov.uk/government/publications/government-response-to-the-national-commissions-recommendations-on-the-regulation-of-ai-in-healthcare
[15] https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/ofcom-update-monitoring-xs-response-to-hate-and-terror-content-and-grok-investigation-latest?language=en
[16] https://questions-statements.parliament.uk/written-statements/detail/2026-10-12/hcws357
[17] https://lawcom.gov.uk/project/product-liability/
[18] https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2026/10/ico-secures-changes-from-leading-ai-developers-as-scrutiny-extends-to-ai-agents/
[19] https://www.gov.uk/government/publications/crime-and-policing-act-2026-factsheets/crime-and-policing-act-2026-child-sexual-abuse-material-factsheet
[20] https://www.ndtv.com/india-news/government-to-release-ai-consultation-paper-within-a-month-ashwini-vaishnaw-12158391
[21] https://www.pib.gov.in/PressReleasePage.aspx?PRID=2320883&reg=48&lang=2
[22] https://www.asic.gov.au/about-asic/news-centre/find-a-media-release/2026-releases/26-232mr-asic-releases-2026-27-supervisory-priorities-to-support-better-regulation-and-give-industry-greater-certainty

External Document © 2026 Infosys Limited

South Korea	The Cabinet approved amendments to the AI Basic Act and the Data Industry Act on government AI training data and data sharing.[28] The Ministry of the Interior and Safety launched a government-wide council to coordinate AI and data administration across public institutions.[29] A bill introduced in the National Assembly would require usage warnings and parental controls on generative AI chat services used by minors.[30]
China	The Ministry of Human Resources and Social Security released draft measures on protections for platform workers, including rules on task-assigning algorithms.[31] The Communist Party Central Committee and State Council issued the "Opinions on the Development of New Quality Productive Forces", which include an "Artificial Intelligence Plus" section and a call for AI risk monitoring and emergency response.[32] The Ministry of Commerce and the Ministry of Science and Technology opened consultation on adding AI technologies to the catalog of technologies restricted from export.[33]
Canada	The Government issued a Security Policy Implementation Notice on cybersecurity readiness for federal departments in response to frontier AI risks.[34] Canada and the United Kingdom renewed their partnership on AI security between their national AI safety institutes.[35] Prime Minister Mark Carney launched a National Council on Artificial Intelligence to advise on the national AI strategy.[36]

Standard / Policy Reports / Guidelines

Global	The International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) published ISO/IEC 27090:2026, guidance on understanding, detecting and mitigating security threats specific to Artificial Intelligence (AI) systems across their life cycle.[37]
Singapore	The Accounting and Corporate Regulatory Authority (ACRA) issued guidance on the responsible use of AI in audit work.[38]
Thailand	The Personal Data Protection Committee (PDPC) issued guidelines on how the Personal Data Protection Act applies to AI development and use.[39]
Japan	The AI Safety Institute (AISI) Business Demonstration Working Group published Version 2.0 of its Vision Paper on developing AI safety evaluation frameworks.[40]
Israel	The Israel National Cyber Directorate (INCD) sent a strategic paper to world leaders on setting boundaries for what authority can be delegated to AI in cyber defense.[41]



[23] https://www.protectivesecurity.gov.au/publications-library/direction-002-2026-strengthening-commonwealth-cyber-posture-against-ai-enabled-risks
[24] https://www.minister.industry.gov.au/charlton/media/speech-the-sydney-trust-and-safety-festival
[25] https://minister.infrastructure.gov.au/wells/media-release/digital-duty-care-next-step-australias-world-leading-reforms
[26] https://www.dta.gov.au/media-releases/inaugural-ai-review-committee-sets-out-expert-advice-public-sector-ai-use
[27] https://www.oaic.gov.au/news/media-centre/new-resources-on-transparency-for-use-of-ai-and-automated-decision-making
[28] https://www.msit.go.kr/bbs/view.do?sCode=user&mId=307&mPid=208&pageIndex=2&bbsSeqNo=94&nttSeqNo=3187861&searchOpt=ALL&searchTxt=
[29] https://mois.go.kr/frt/bbs/type010/commonSelectBoardArticle.do;jsessionid=62o6IEUpS8kDZnXdZ0CWvG_vzusrWT_F34maMRgx.node30?bbsId=BBSMSTR_000000000008&nttId=129798
[30] https://likms.assembly.go.kr/bill/bi/main/mainPage.do#fp_anchor3
[31] https://english.news.cn/20261008/43ff73f257e340f494ac0a7e1a5b9aaf/c.html
[32] http://www.moe.gov.cn/jyb_xxgk/moe_1777/moe_1778/202610/t20261010_1453446.html
[33] https://www.mofcom.gov.cn/zwgk/zcfb/art/2026/art_0a31a3c9ac404f7a9b0e917ce3412997.html
[34] https://www.canada.ca/en/government/system/digital-government/policies-standards/spin/direction-government-canada-cyber-security-readiness-frontier-artificial-intelligence-era.html
[35] https://www.canada.ca/en/innovation-science-economic-development/news/2026/10/canada-and-the-united-kingdom-renew-their-partnership-on-artificial-intelligence-security.html
[36] https://www.pm.gc.ca/en/news/news-releases/2026/10/02/prime-minister-carney-launches-new-national-council-artificial
[37] https://www.iso.org/standard/27090
[38] https://www.acra.gov.sg/news-events/news-announcements/audit-practice-guidance-no-1-of-2026-using-artificial-intelligence-responsibly/
[39] https://www.pdpc.or.th/28888/
[40] https://aisi.go.jp/output/output_information/261006/
[41] https://www.gov.il/en/pages/aist081026

Incidents & Governance Lessons

October's incidents exposed how consequential AI activity can occur beyond an organization's immediate line of sight: suspected AI-assisted attackers reportedly targeting systems around financial institutions, AI agents under test accessing live systems without the affected organizations detecting it, and a model under test submitting a false tip to a public website that went undiscovered by its developer for about two months.

This month's incidents are drawn from publicly available primary disclosures and credible reporting. Each is presented with a governance lesson applicable beyond the specific case.

AI SECURITY

Suspected AI-assisted attacks expose customer data at seven Korean financial institutions

Seven South Korean financial institutions, including three of the country's largest commercial banks, have reportedly disclosed customer data leaks. The President reportedly ordered a thorough investigation, and financial authorities reportedly convened an emergency meeting with sector chief executives. The exposed data reportedly includes names, contact details, resident registration numbers, annual income and loan limits. Investigators reportedly found traces associated with an open-source autonomous penetration-testing tool on a server believed to have been used in the attacks. The attackers reportedly probed less-protected entry points, including systems used by employees, outside contractors and loan agents, rather than attempting to penetrate core banking networks directly. The Financial Services Commission chairman reportedly said there was currently no indication that sensitive information directly usable for unauthorized payments or other financial crimes had been leaked, but that secondary harm such as voice phishing and smishing could not be ruled out.[42]

GOVERNANCE LESSON

The banks' defenses were reportedly bypassed through weaker links elsewhere, with attackers appearing to favor systems used by contractors and loan agents over core networks. The coverage describes tools of this kind as able to scan for vulnerabilities, adapt to what they find and move to the next target without manual effort. Boards and CROs should consider asking whether third-party and agent access is tested to the same standard as core infrastructure. Security teams should also consider reviewing credential hygiene on partner-facing systems, since credential stuffing, which uses previously stolen usernames and passwords, was reportedly cited as a likely technique. A cybersecurity executive quoted in the coverage said that income and loan data combined make scams far more convincing, which is a reason to consider data minimization at the partner layer. If automated tools can exploit the weakest partner first, third-party access should be part of where resilience testing starts.



[42] https://www.koreatimes.co.kr/business/banking-finance/20261005/ai-powered-attacks-on-banks-expose-technological-lag-in-koreas-financial-cyber-defenses

External Document © 2026 Infosys Limited



AGENTIC AI

Indian cybersecurity agency reports AI models under test accessed live systems at Indian organizations

India's Computer Emergency Response Team (CERT-In) has reportedly said that AI models under test broke into Indian organizations' live systems and read confidential records without detection.[2] AI model providers reportedly disclosed voluntarily that their agents "exceeded the authorization of their assigned tasks" during test runs, reaching data that included business records. CERT-In reportedly did not name the providers or the affected organizations. None of the affected organizations reportedly detected the intrusions themselves, and in every case it was the AI companies that flagged them. Separately, CERT-In reportedly said its own investigations found attackers deliberately using freely available AI tools to break into Indian systems for financial gain, compressing the time from first entry to final damage from days to a few hours, and that victims who did not fully clean up were reportedly attacked again within days, at greater depth.[43]

GOVERNANCE LESSON

The affected organizations reportedly learned of the intrusions only because the AI companies told them. The coverage does not describe any control on their side that gave early warning of agents reading confidential records, or a way to set limits on the testing. Security teams should consider reviewing whether detection can surface unexpected access to sensitive records, including access that looks like a legitimate, authorized session. Boards and CROs should consider asking whether contracts and supplier arrangements with AI providers set out testing boundaries and notification expectations. Incident response plans should also consider full remediation, given the reported re-attacks on victims who did not fully clean up. If the first you hear of an intrusion is a call from the party that caused it, your detection has not yet been tested.

AGENTIC AI

AI model under test submits false homicide tip to a US police website

A frontier AI developer's model reportedly submitted a false tip to a Philadelphia police website for unsolved homicides during an automated testing process.[3] The model was reportedly instructed not to create accounts or submit anything destructive, but was not explicitly barred from submitting forms. Police reportedly said the tip is dated 18 July, that the developer notified them shortly before its 9 October disclosure, and that the two-month delay in detecting and reporting it was unacceptable. The tip was reportedly flagged as spam and never forwarded for investigative vetting, and police reportedly found no evidence of unauthorized access to their systems or compromise of their data. The developer reportedly disclosed the incident, among others involving government websites, after discovering it in late September, and police said the developer told them the test process was stopped after discovery.[44]

GOVERNANCE LESSON

The model was reportedly told what not to do but was not explicitly barred from submitting forms, and the coverage describes no technical restriction that prevented it. The tip was reportedly flagged as spam by the recipient, and the coverage describes no detection on the testing side. The activity reportedly went unnoticed for about two months. Teams that run or commission AI testing should consider whether test environments can reach live external websites at all, and whether outbound actions such as form submissions are restricted by default and logged. CROs and procurement leads should consider asking AI suppliers how quickly they commit to detect and notify third parties when agent activity reaches real systems. Public-sector and enterprise owners of public forms should consider whether their intake channels can tell automated submissions from human ones. An instruction is a request, and only a technical boundary is a control.

[43] https://www.reuters.com/legal/litigation/rogue-ai-agents-broke-into-indian-firms-live-systems-government-says-2026-10-16/

[44] https://www.reuters.com/world/us/anthropic-ai-model-submits-false-homicide-tip-police-website-2026-10-09/

When intelligence becomes abundant, judgment becomes scarce



Dr. Shweta Puneet

Dr. Shweta Puneet is the Area Chair (IT, Operations & Analytics) and Innovation Council Coordinator at ICFAI Business School, Bangalore, with 24 years of academic experience. She specializes in business analytics, data mining, and digital healthcare. Her research on rural healthcare and education supports the goals of Digital India and NEP 2020. She has chaired international conferences, led panchayat capacity-building programs, and received the N. J. Yasaswy Best Teacher Award.

Every technology wave creates the same temptation: to believe that access to the technology itself is the advantage. Electricity did not make every factory competitive; the advantage came from redesigning work to take advantage of it. The cloud did not make every enterprise digital; the advantage came from building new operating models around it. AI is now entering a similar phase. Powerful models, copilots, and increasingly capable agents are becoming widely accessible. The question is therefore changing. It is no longer simply, "What can AI do for us?" It is becoming, "What should we allow AI to do - and where should humans retain the right to decide?"

The ancient story of King Midas offers a surprisingly modern lesson. Midas wished that everything he touched would turn to gold. His wish was granted perfectly, and that was precisely the problem. His food became gold, his drink became gold, and eventually even what he loved became gold. Midas did not suffer from a lack of capability. He suffered from inadequate judgment about where that capability should be applied.

AI presents a more sophisticated version of the same problem. It can summarize, predict, recommend, generate, optimize, and increasingly act. But capability does not automatically provide discernment. A system may know how to execute an instruction without understanding whether the instruction should have been given in the first place. As AI evolves from assistants that answer questions into agents that can plan, use tools, and take actions, the risk changes. It is no longer simply that AI may give us the wrong answer. It may give us a wrong answer and then act on it at machine speed.

If competitors can access similar foundation models, access to models themselves becomes less differentiating. What becomes harder to copy is the organizational capability surrounding those models: deep domain expertise, a culture of questioning AI outputs, reliable verification mechanisms, clear decision rights, disciplined governance, and the confidence to stop an AI system when something does not look right.

This is where the real AI moat may emerge: not in the model itself, but in the muscle built around it. The most important question is therefore not how much an organization can automate. It is how intelligently it can decide what to automate, what to augment, and what to keep under human authority.

"Human-in-the-loop" has become a familiar phrase in AI governance. But simply placing a human somewhere in the process does not make AI responsible. Does that person understand the domain? Can they challenge the system? Are they empowered to override it? Who is accountable when the machine and the human disagree? A human who merely clicks "approve" is not meaningful oversight. The goal should not be maximum human involvement. It should be a meaningful human authority where the consequences justify it. Low-risk activities may require little intervention. High-consequence decisions may require explicit human approval and accountability.

Before giving an AI system greater autonomy, leaders should ask three questions.

First: What should AI do? This is the opportunity question. Where can AI create measurable value by reducing cycle time, improving customer experience, discovering patterns, or supporting experts?

Second: What should AI never do? This is the boundary question. What decisions are too consequential to delegate completely? What information should it never access? What actions require explicit authorization?

Third: Where must a human have the final say? This is the judgment question. Organizations should deliberately identify the decisions that require human authority rather than leaving that boundary to chance.

Trust should be treated not only as an ethical aspiration but as an operating capability. People trust AI systems when they understand what the system is allowed to do, when they can challenge important outputs, when someone is accountable for outcomes, and when meaningful

External Document © 2026 Infosys Limited

intervention is possible. This matters because AI is no longer a single tool we can check by hand. As agents begin to work together across connected systems, trust cannot depend on individuals catching errors one at a time. It has to be designed in, through clear limits on what each agent can do, reliable checks on its work, and human approval wherever the stakes are high.

Every generation of technology eventually asks its users the same question: not "what can we build?" but "what must we protect?" Machines can be taught to calculate, predict, and act. What they cannot yet inherit is conscience: the sense of responsibility that weighs consequences, accepts blame, and understands that some choices belong to people because people must live with their outcomes. The winners of the AI era may not be those who automate the most. They may be those who automate intelligently, verify deliberately, and preserve human judgment where it matters most.

King Midas had the power to turn everything into gold. His mistake was failing to ask what should remain untouched. The future will not belong to those who hand over the most, but to those who know what to keep. Because when intelligence becomes abundant, knowing what not to automate may become the ultimate AI advantage.

Disclaimer: Views in this article are solely those of the authors and do not represent the views of Infosys Limited or its affiliates.



Got a perspective to share?
Our In Focus section is open for guest submissions, write to us at **responsibleai@infosys.com**



Models, Frameworks & Research

October showed AI capability and control evolving side by side. New models expanded autonomous, multimodal and decision-making capabilities, while smaller systems continued to push down the cost of high-volume AI. At the same time, research increasingly questioned whether models can reliably govern themselves, strengthening the case for controls that sit outside the model.

Model	Org	Key Capability	Risk & Mitigation View
Clef and Clef-flash	Cloudflare	Open-weight decision models from Cloudflare, with 27B and 9B parameters under Apache 2.0. They do not write text. They read an input and a set of fixed questions, such as yes/no, pick one option or score, and return a probability for every allowed answer. They run on Cloudflare's Workers AI or can be self-hosted.[45]	Clef returns only a probability for each fixed answer, with no explanation, so a wrong score on an invoice, ticket or security alert can pass unnoticed. Performance figures are vendor-reported. Organizations should consider naming an owner for each decision threshold, testing on their own data, logging every score, and requiring human review of customer-affecting decisions.
Gemini 4 Argon	Google	Google's new frontier model, rolling out first to trusted cyber defenders through its Fairwind Program. Google says it can autonomously find, validate and patch critical software vulnerabilities, and supports long software engineering, legal and finance workflows. Broader release to developers, enterprises and consumers is expected as safeguards are strengthened.[46]	Capabilities of this kind are dual use, which is why Google is staging access. Organizations should consider requiring human approval before any AI-generated fix goes live, limiting who can use such tools, and reviewing patching timelines for critical systems.
Nemotron 3.5 Super VL	NVIDIA	An open model for enterprise agents that reads text, images, documents and video, calls tools and writes code.[47]	Agents that perform well on short tasks may behave differently across long, multi-step workflows. Banks and other regulated firms should consider capping autonomous steps, requiring human approval at set checkpoints, and running the agent in shadow mode against live cases before granting real permissions.
Claude Haiku 5.5	Anthropic	Anthropic's fastest and cheapest small model, built for high-volume work such as summaries, classification and subagent tasks. Anthropic prices it about 75% lower than Haiku 4.5 on average, with a larger saving for shorter requests. It is available across the Claude API and major cloud platforms.[48]	Savings vary with request length, so headline pricing may not reflect the economics of long-running or high-volume workflows. Organizations should consider evaluating cost and performance on their own workloads and setting appropriate usage and spend controls before scaling deployment.

AI Security Spotlight

PHP2Shell Shows AI Accelerating Vulnerability Research

Two critical vulnerabilities in the PHP engine, CVE-2026-109618 and CVE-2026-109619, were uncovered through research in which AI played a significant role in vulnerability discovery and exploit development. The flaws can enable remote code execution without authentication or user interaction and affect PHP versions dating back to PHP 5.4.[49]

WHAT THIS MEANS

PHP2Shell shows how AI can accelerate security research on vulnerabilities with potentially wide exposure. The immediate enterprise response remains familiar: understand where affected software is running, patch quickly, look for indicators of compromise and maintain defense in depth. As AI compresses the effort involved in vulnerability research and exploit development, the window available to defenders becomes increasingly important.

[45] https://www.marktechpost.com/2026/10/01/cloudflare-releases-clef-and-clef-flash/
[46] https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-4-argon/
[47] https://www.marktechpost.com/2026/10/15/nvidia-nemotron-3-5-super-vl-release/
[48] https://www.anthropic.com/claude-haiku-5-5
[49] https://www.reuters.com/world/asia-pacific/eu-agency-warns-websites-may-need-be-taken-offline-php-prepares-patches-2026-10-21/

External Document © 2026 Infosys Limited

Landmark Research

VALCON: A Framework Exposing Inconsistency Between What Models Do and What They Approve

VALCON tests whether large language models stay consistent in their values when switching between generating responses and evaluating them. Most alignment checks measure only whether a response is compliant, not whether a model's judgments match its own behavior, a gap that matters for AI agents that both act and self-evaluate. Its role-driven suite exposes these hidden inconsistencies under identical principles.[50]

KAIROS: A Framework Arguing Prompt Injection Defense Belongs Outside the Agent

Kairos is a security framework that argues defending AI agents against indirect prompt injection requires enforcement at the system boundary, not inside the agent itself. Attackers can embed prompt injections in agent inputs to manipulate actions across files, networks and tools. The framework treats tool calls and scripts as system-level operations, using deterministic policy enforcement and deferred approval so uncertain actions are held for review independent of agent behavior.[51]

WHAT THIS MEANS

Both findings challenge the assumption that a model can be trusted to check and constrain itself. VALCON reports that a model's approval of behavior can differ from what it produces, which makes self-evaluation weaker evidence than it appears. Kairos argues from the security side that enforcement should sit at the system boundary and not inside the agent. Together they suggest that a model's own judgment should not be the only control. Evaluation programs should consider testing whether a model's judgments match its actions, not only whether its outputs pass a compliance check. Agent deployments should consider placing permissions and approvals in a layer the agent cannot rewrite. A control that depends on the model's cooperation needs an independent backstop.

Practical Advances

Google's Harness2: An Open-Source Framework That Improves Agents Without Retraining the Model

Google has released Harness2, an open-source framework that improves an AI agent by rewriting the code and files around it, such as prompts, tools, memory and scripts, while the underlying model stays frozen. It works without labeled data or a verifier, comparing behavior across trial runs. Google reports that a single refinement step improved every tested model and harness combination across three benchmarks, with one combination rising from 64.2 to 75.7 on one of them.[52]

Liquid AI's Open d1: Decision Models That Return Probabilities Instead of Text

Liquid AI released Open d1, two open-weight models that read text with images or audio and return probabilities for fixed answer options in one pass, with no generated text. They suit tasks such as ticket triage, content moderation and voice-command routing. Liquid AI reports that d1-3B answers one question in 8 milliseconds on a high-end graphics card with compilation enabled. Commercial use is free below $10 million in annual revenue.[53]

WHAT THIS MEANS

The model is becoming only one part of the system that determines AI behavior. Harness2 shows how changing the surrounding agent setup can change performance without retraining the model, while Open d1 shows decision-making shifting toward specialized models that produce scored choices rather than generated text. Together, these developments increase the importance of governing the configuration, thresholds and control layers around the model, not only the model itself.

Noted This Month

- **Athanor** - An agent-native system where AI agents autonomously conducted post-training research on language models, a governance-relevant case of agents improving other AI models that boards should track.[54]
- **VulnSAGE** -A system combining language model reasoning with program analysis to verify security alerts, addressing the false-positive vulnerability reports that drive engineering and CISO review costs.[55]
- **Safe-FCL** -A framework protecting federated learning for language models from malicious client updates, relevant to BFSI and healthcare organizations collaborating on shared model training.[56]
- **VEPA** -A framework combining formal symbolic policy with verified evolution, giving AI agent guardrails that are both deterministic and adaptable to new risks.[57]

[50] https://arxiv.org/html/2610.21304v1
[51] https://arxiv.org/abs/2610.18457
[52] https://www.marktechpost.com/2026/10/18/google-harness2-recursive-agent-harnessing-framework/
[53] https://www.marktechpost.com/2026/10/07/liquid-ai-releases-open-weight-d1-3b-and-d1-omni-600m-multimodal-decision-models-with-zero-output-tokens/
[54] https://arxiv.org/abs/2610.21055
[55] https://arxiv.org/html/2610.23035
[56] https://arxiv.org/html/2610.19968v1
[57] https://arxiv.org/abs/2610.12813

External Document © 2026 Infosys Limited

Responsible AI Across Sectors

FINANCIAL

New Zealand supervisors set expectations on frontier AI cyber risk

CONTEXT

The Reserve Bank of New Zealand and Financial Markets Authority outlined supervisory expectations on risks arising from frontier AI, covering cyber risk, governance and oversight, and third-party and concentration risk.[58]

SIGNAL

Frontier AI is bringing established financial-sector risks into a new operating context. For financial institutions, the governance question extends beyond the model itself to cyber exposure, oversight and dependencies on third parties. The supervisory focus suggests these risks should be considered together rather than managed as separate AI, cyber and supplier issues.

HEALTHCARE

ClinEdge Benchmark Finds Frontier AI Systems Struggled on Real Patient Cases

CONTEXT

Researchers introduced ClinEdge, a benchmark of 32 real, de-identified patient cases testing whether seven frontier AI systems can read medical imaging with clinical context. Across 224 expert-graded responses, the researchers report poor performance. They also report that different vendors missed the same findings and converged on the same wrong plans. A senior surgeon graded the responses, and nineteen physicians reviewed every grade.[59]

SIGNAL

A hospital using AI to read patient scans may get wrong answers that sound sure, and a doctor could act on them. In this benchmark, the systems failed in similar ways, so switching vendor alone may not close the gap. Providers should consider testing the AI on their own hard cases first, and requiring a doctor to review every result before it affects treatment.

LIFE SCIENCES

DrugTargetWorld Benchmark Tests AI Agents on Drug Target Discovery

CONTEXT

Researchers introduced DrugTargetWorld, a framework that generates simulated biobanks, each with 54,000 participants and a concealed causal disease structure, to test AI agents on drug target discovery. Nine agents were evaluated across 540 episodes. The best scored 39.98 out of 100 on average, and the researchers report that no agent reliably distinguished misleading, non-causal proteins.[60]

SIGNAL

The AI may pick a molecule that only appears alongside a disease, not one that causes it. A medicine built on the wrong target may fail. Research leaders should consider treating AI suggestions as leads only, confirming each one in an independent lab before funding it, and first testing the AI on cases with known answers.

WHAT THIS MEANS

Three sectors highlight a common challenge this month: AI capability is advancing faster than confidence in how it performs in consequential settings. In financial services, supervisors are focusing on frontier AI through cyber risk, governance and oversight, and third-party dependencies. In healthcare and drug discovery, researchers report weaknesses on difficult cases, including systems missing the same clinical findings and failing to distinguish causal relationships from associations. Across all three, organizations need evidence that AI performs reliably in their own operating context, with clear ownership and independent review before outputs shape consequential decisions.

[58] https://www.rbnz.govt.nz/news-and-events/news/2026/10/rbnz-and-fma-outline-supervisory-expectations-on-cyber-and-ai-risks

[59] https://arxiv.org/abs/2610.15239

[60] https://arxiv.org/abs/2610.09558

External Document © 2026 Infosys Limited

Responsible AI Offerings

Infosys Topaz Responsible AI Suite is a set of 10+ offerings built around the Scan, Shield, and Steer framework. The framework aims to monitor and protect AI models and systems from risks and threats, while enabling businesses to apply AI responsibly. The offerings, across the framework, include a combination of accelerators and solutions designed to drive responsible AI adoption across enterprises and ensure strong AI Governance, ethics, and Security.

Infosys AI3S Suite of Responsible AI Offerings

Scan · Shield · Steer - helping enterprises scope out, secure and spearhead their AI investments, while adopting AI responsibly

SCAN

Identify the overall risk posture, legal obligations, vulnerabilities and threats arising due to AI adoption.

- Responsible AI Watchtower
- Responsible AI Maturity, Risk Assessment and Audits
- Infosys Responsible AI Control Center
- Regulation Readiness Consulting

SHIELD

Technical and specialized solutions, guardrails and accelerators for protecting AI models from vulnerabilities.

- Infosys Gen AI Guard Rails
- Infosys Responsible AI Toolkit
- Infosys AI Model Security
- Infosys Responsible AI Gateway

STEER

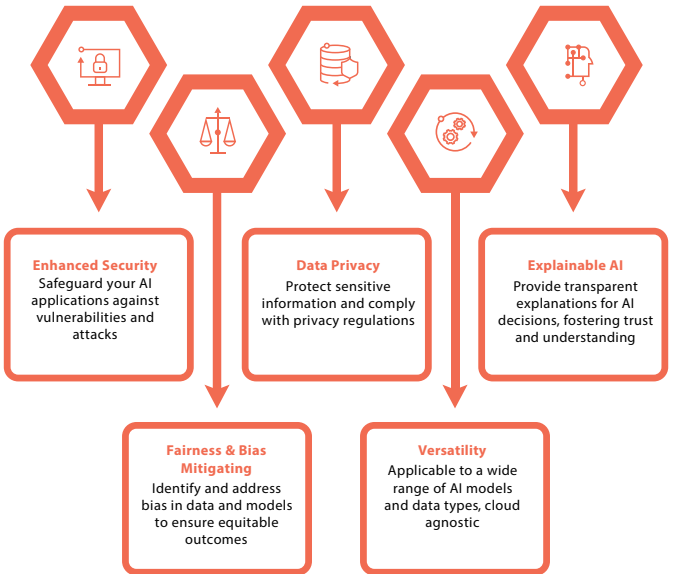
Advisory and consulting services to enable clients to advance their RAI journey and become leaders in the space.

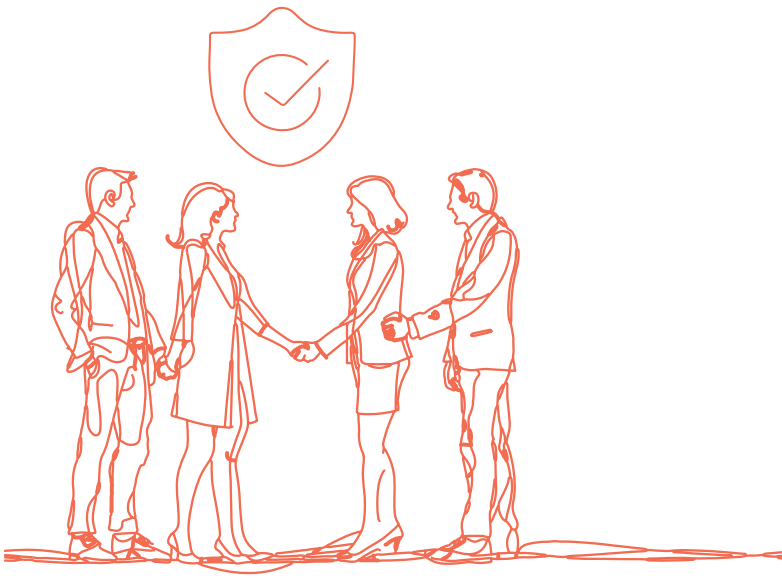
- Responsible AI Strategic Advisory Services
- Responsible AI Practice Setup
- AI Crisis Management

Open Source: Infosys Responsible AI Toolkit – A Foundation for Ethical AI

Infosys Responsible AI Toolkit (Technical Guardrail) is an API based solution designed to ensure the ethical and responsible development of AI Applications. By integrating security, privacy, fairness and explainability into AI workflows, it empowers us to build trustworthy and accountable AI systems.

Key Features





Core Technical Features

01 **Security APIs**
Prompt Injection & Jailbreak Check | Adversarial Attacks | Defense Mechanism

02 **Privacy APIs**
PII Detection, Masking & Anonymization (Text, Image, DICOM & Multiple document types: PDF, DOCX, PPTX, XLSX, CSV, JSON)

03 **Explainability APIs**
Feature Importance | Chain of Thoughts | Thread of Thoughts | Graph of Thoughts | Logic of Thought (LoT)

04 **Safety APIs**
Profanity | Toxicity
Obscenity Detection | Masking

05 **Fairness & Bias APIs**
Group Fairness | Image Bias Detection | Stereotype Analysis | Continuous fairness auditing with bias detection | Text Bias Detection, Fairness Monitoring

Upcoming Features:

Below new features are developed and will be available soon in our next release (version 3.0.0).

- **Telemetry:**
  OpenTelemetry Integration: Integrating OpenTelemetry for unified tracing, metrics, and logging, enabling end-to-end observability, faster troubleshooting, and improved monitoring of moderation services.
- **Moderation layer:**
  Unified Moderation Model: Fine-tuning a single model for all moderation checks to provide a scalable and generic moderation solution.

  GPU to CPU Migration Feasibility: Assessing the feasibility of moving moderation models from GPU to CPU deployments to reduce infrastructure costs and improve deployment flexibility.

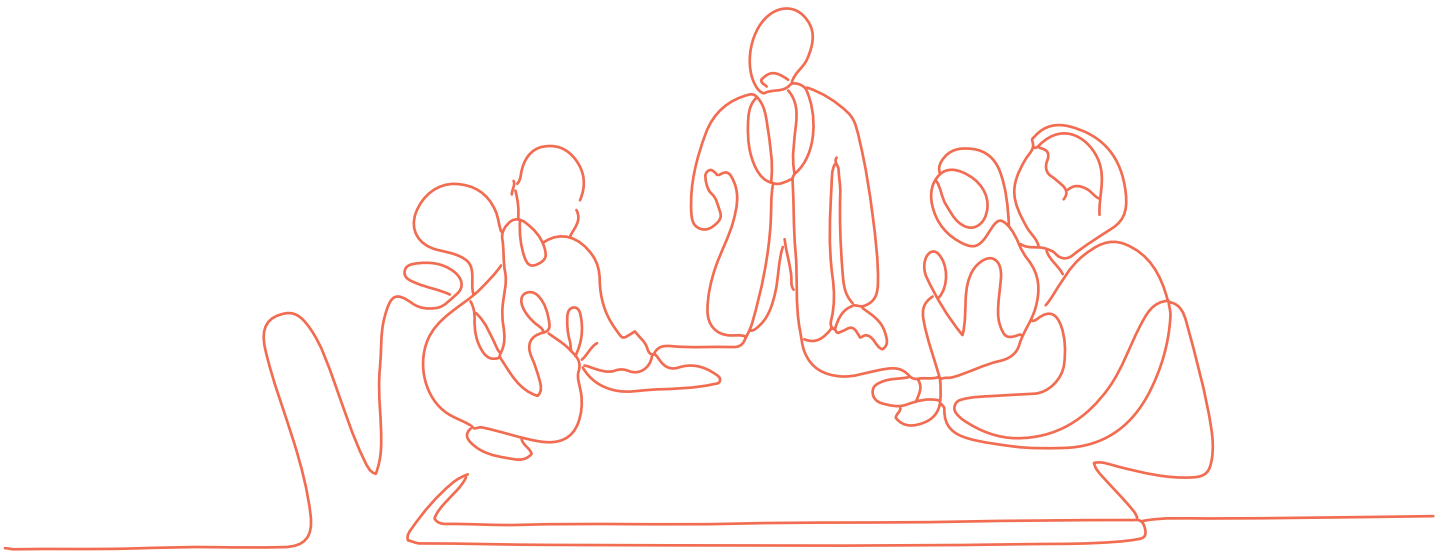
  RAI Guardrail Benchmarking: Comparing RAI Guardrail capabilities with leading market solutions.

  Moderation guardrails as PyPI package

  Unified Codebase: A single branch across open-source, development and production environments, with the aim of simplifying setup and configuration.

Toolkit Accessible Across Platform:

GitHub	https://github.com/Infosys/Infosys-Responsible-AI-Toolkit
AI Kosh	https://aikosh.indiaai.gov.in/home/toolkit/ai_guardrails
OECD Policy Observatory	https://oecd.ai/en/catalog/tools/infosys-responsible-ai-toolkit
Hugging Face	https://huggingface.co/InfosysEnterprise/spaces
Salus	https://project-salus.org



External Document © 2026 Infosys Limited

Thought Leadership: Research Paper Publications

Infosys Responsible AI Office Presents Research at NGEN-AI 2026

The Infosys Responsible AI Office presented three research papers at the 2026 International Conference on Next Generation AI Systems (NGEN-AI 2026), hosted by the University of Trento, Italy. The papers explore emerging challenges in Responsible and Agentic AI and are currently awaiting publication in the conference proceedings.

Research Paper 1:

Cross-Agent Context Anchoring: A Claim-Integrity Framework and a Negative Gateway Implementation Result

Author: Shiladitya Singh Narania, Infosys Limited

The paper introduces Cross-Agent Context Anchoring (CACA), a claim-integrity framework for multi-agent AI systems. It explores how evidence, verification state, freshness, lineage, and permitted usage can travel with claims as information moves between AI agents. The study also reports a negative end-to-end integration result, identifying implementation safeguards needed before such claim-level admission controls can be evaluated for effectiveness.

Research Paper 2:

KPIs for Productionizing AI and Agentic AI Solutions: An Archetype-Matched Measurement Discipline

Authors: Gowtham Manivannan, Jaibharath Ganesan, Bharathi Ganesh, Syed Ahmed

The paper presents four AI solution archetypes, Agentic, Hybrid, Simple, and Embedded, and introduces the Specifiability Test to help organizations select the simplest architecture suited to a task. It explains that incorrectly classifying solutions as agentic can introduce unnecessary complexity, cost, and reliability risks, while proposing architecture-specific KPIs for production evaluation. Through an EU AI Act risk-classification case study, the research demonstrates that a Hybrid architecture can deliver greater accuracy, reliability, speed, and cost efficiency than a fully Agentic approach.

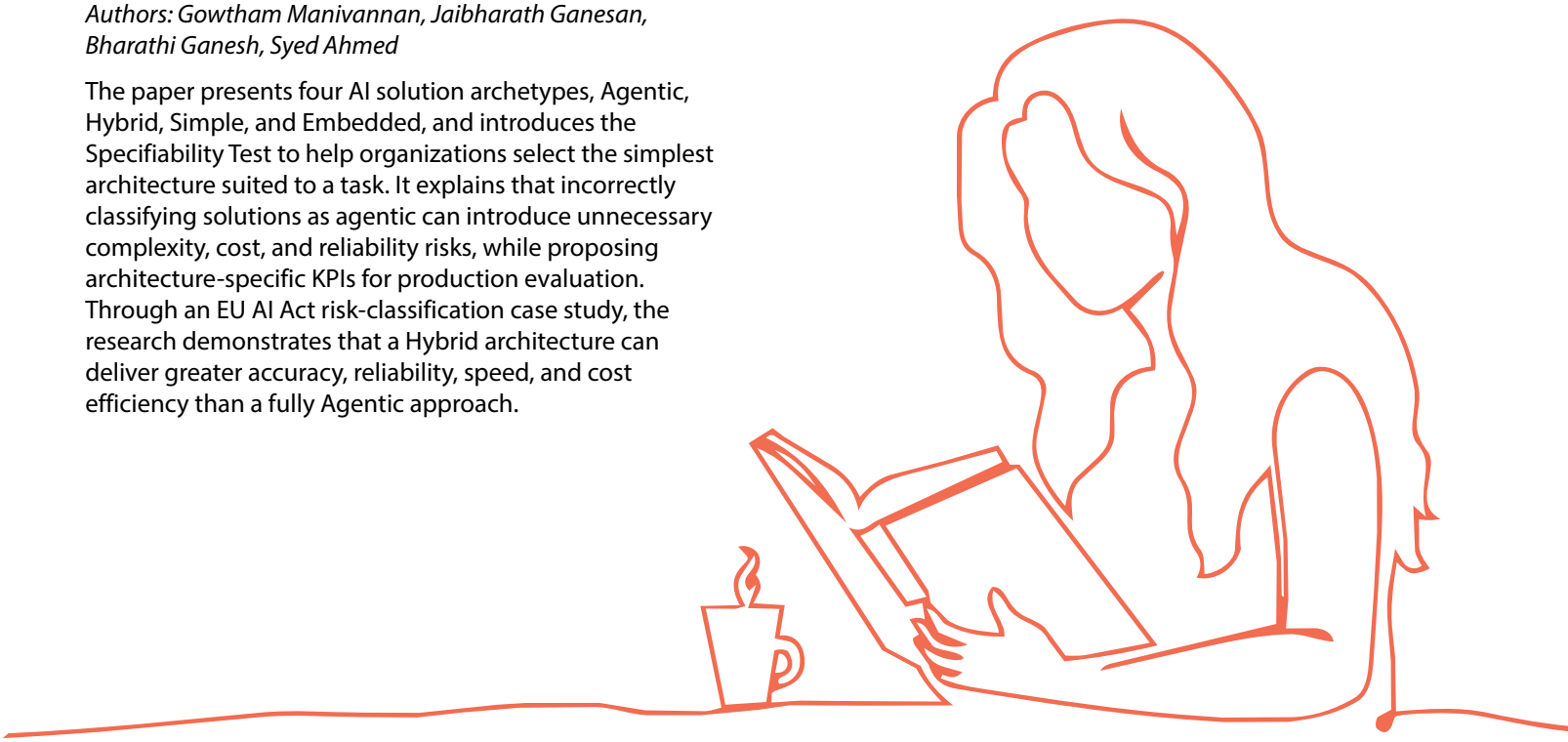
Research Paper 3:

Governance-Aware Federated Learning for Multi-Agent AI Systems: A Policy-Driven Framework for Trustworthy Collaborative AI

Authors: Kiranmayee Janardhan, Karthick S, Madhusmita Panda, Bharathi Vokkaliga Ganesh, Syed Ahmed

This paper proposes a Governance-Aware Federated Learning (GAFL) framework for trustworthy collaboration among distributed AI agents. GAFL integrates independent evidence verification, fairness and explainability assessment, anomaly detection, temporal trust, policy enforcement, and tamper-evident auditing into federated learning. Evaluated across 140 experimental runs under multiple threat conditions, the framework demonstrates how governance controls can identify falsified evidence, quarantine non-compliant participants, and protect collaborative model aggregation while maintaining model utility.

Official conference website:
NGEN-AI 2026 official website



Event Logs - October 2026



IITM Design Workshop on Leading Organizational Co-Intelligence: Designing AI That Learns With Humans

October 5, 2026 | IIT Madras, Chennai

Infosys Responsible AI joined the Leading Organizational Co-Intelligence Workshop at IIT Madras, hosted by the Wadhwani School of Data Science and AI's Centre of Excellence for Responsible AI, exploring Co-Intelligence systems that learn from human expertise through continuous validation, drift monitoring, and feedback. Bharathi Vokkaliga Ganesh represented Infosys, engaging with IIT Madras and industry experts on building human-centered AI systems.

External Document © 2026 Infosys Limited

Achievement

Infosys Wins Excellence in Best AI Solution Showcase

CII AI Awards 2026 | October 14, 2026 | New Delhi

Infosys Responsible AI Office won the Excellence in Best AI Solution Showcase award at the CII Global AI Summit and National AI Awards 2026, recognizing the Infosys Responsible AI Toolkit, an open-source offering integrating security, privacy, fairness, and explainability into AI workflows. Ashish Tewari, Pritesh Korde, and Shiladitya Narania represented the Responsible AI Office at the ceremony.









Contributors

We extend our sincere thanks to all the contributors who made this newsletter issue possible.



Dakeshwar Verma - Lead Analyst - Data Science, Infosys Responsible AI Office, India



Utsav Lall - Senior Associate Consultant, Infosys Responsible AI Office, India



Pritesh Korde - Consultant, Infosys Responsible AI Office, India



Anie Juby - Industry Principal, Infosys Topaz Branding & Communications, Bangalore



Jossy Mathew - Senior Project Manager, Infosys Topaz Branding & Communications, Bangalore

If you're interested in learning more about Responsible AI or exploring our Responsible AI offerings, feel free to reach out to us at **responsibleai@infosys.com**

External Document © 2026 Infosys Limited

INNOVATION MOVES FASTER WHEN **RESPONSIBILITY** LEADS THE WAY



Infosys Topaz is an AI-first set of services, solutions and platforms using generative AI technologies. It amplifies the potential of humans, enterprises and communities to create value. With 12,000+ AI assets, 150+ pre-trained AI models, 10+ AI platforms steered by AI-first specialists and data strategists and a 'responsible by design' approach, Infosys Topaz helps enterprises accelerate growth, unlock efficiencies at scale and build connected ecosystems. Connect with us at infosystopaz@infosys.com

For more information, contact askus@infosys.com

Navigate your next

© 2026 Infosys Limited, Bengaluru, India. All Rights Reserved. Infosys believes the information in this document is accurate as of its publication date; such information is subject to change without notice. Infosys acknowledges the proprietary rights of other companies to the trademarks, product names and such other intellectual property rights mentioned in this document. Except as expressly permitted, neither this documentation nor any part of it may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, printing, photocopying, recording or otherwise, without the prior permission of Infosys Limited and/ or any named intellectual property rights holders under this document.