keypoint

AIFl@sh

August 2026

GCC

Kingdom of Saudi Arabia

NCA - Consultation on AI cybersecurity guidelines closes

On **5 August 2026**, the consultation period for the National Cybersecurity Authority's (NCA's) AI cybersecurity guidelines closed. The draft guidelines aim to ensure that cybersecurity requirements are applied to AI systems and address AI-related cyber risks across four areas: governance, defence, resilience and third-parties. The scope specifically includes emerging technologies such as generative AI and agentic AI.

AI governance and cybersecurity in Saudi Arabia are becoming increasingly interconnected. Organisations using or planning to use generative or agentic AI should integrate AI-specific risks into existing cybersecurity governance, resilience and third-party risk management arrangements rather than treating AI governance and cybersecurity as separate programmes.

Kingdom of Bahrain

No new Bahrain-specific AI law or policy update was identified for August 2026.

For organisations, Bahrain's direction remains implementation-led. AI governance should therefore be embedded into procurement, vendor due diligence, privacy reviews, workforce training and business-as-usual controls, rather than treated as a separate innovation activity.

The United Arab Emirates

AI programme expands focus on generative and agentic AI governance

On **23 August 2026,** the office of the Minister of State for AI opened registrations for the seventh cohort of the UAE AI Program, with a particular focus on generative AI and agentic AI. The programme includes a professional track covering strategic AI applications, governance, technology vendor assessment and digital adoption planning, alongside a technical track covering architecture, programming and model development.

The UAE is pairing rapid AI deployment with structured capability-building around governance and implementation. Organisations adopting generative and agentic AI should ensure that leaders understand both governance, vendor assessment and operating-model implications as well as the technical capabilities of the systems being deployed.

Qatar

Foreign Ministry launches *Sard*

On **24 August 2026,** Qatar's Ministry of Foreign Affairs launched its "Sard" platform to monitor news, events and social-media content across multiple languages and support the ministry and its staff with relevant information and data. The launch forms part of its plan to leverage AI-powered technologies in its daily operations, both within departments and across diplomatic missions abroad.

The development demonstrates how Qatar is embedding AI into core government workflows rather than limiting its use to standalone pilots. As AI-powered monitoring and analytical tools become more operational, organisations should consider controls around source quality, data provenance, access, validation, human review and auditability.

Kuwait

Ministry of Health issues AI and LLM guide for healthcare

On **11 August 2026,** Kuwait's Ministry of Health approved and published a practical guide regulating the use of AI language tools and large language models (LLMs) in healthcare practice. The guide classifies AI uses across three risk levels, prohibits sole reliance on AI for key clinical decisions, requires human review of AI-generated outputs and restricts identifiable or confidential patient information from being entered into unapproved AI or cloud tools. Patient-facing AI tools used in the ministry's name are also subject to formal assessment and accreditation requirements.

Kuwait is translating responsible AI principles into sector-specific operational controls. Healthcare organisations and technology providers working with the ministry should focus on AI use-case classification, approved-tool processes, data protection, human validation, model assessment and incident reporting as core components of AI governance.

Sultanate of Oman

Agentic AI challenge moves into development phase

Oman's **"Engineer it with agentic AI"** challenge, delivered through the Ministry of Transport, Communications and Information Technology's *makeen* programme, moved from preliminary evaluation (4-5 August) into its development phase (6-28 August). Participating teams are developing agentic AI solutions capable of analysing information, making decisions and executing tasks autonomously, supported by expert mentoring and practical development activities.

Oman's AI capability-building agenda is moving beyond general AI skills towards autonomous and agent-based systems. As these capabilities mature, organisations should complement technical experimentation with clear autonomy limits, human oversight, secure system and tool access, monitoring and accountability for actions performed by AI agents.

International

European Union

AI Act transparency requirements take effect

On **2 August 2026**, the European Commission's AI Office and national authorities began exercising enforcement powers under the EU AI Act, while Article 50's transparency requirements also became applicable. Providers must, where applicable, inform individuals when they are directly interacting with AI and apply machine-readable marking to AI-generated or manipulated content, while deployers have disclosure requirements covering areas including deepfakes, emotion recognition, biometric categorisation and certain AI-generated public-interest content. A limited transition until 2 December 2026 applies only to Article 50(2) marking and detection requirements for systems placed on the market before 2 August.

AI transparency is now a live compliance obligation in the EU, rather than a future readiness exercise. Organisations should ensure that applicable user notices, content marking, labelling and provenance controls are operating in practice and should not confuse the delayed high-risk AI deadlines with the Article 50 requirements that are already applicable.

Anthropic implements AI-generated content marking under Article 50

On **14 August 2026**, Anthropic published details of how it is implementing its commitments under the EU AI Act's Article 50(2) code of practice on transparency of AI-generated content. Claude models launched in the EU on or after 2 August 2026 support machine-readable marking at launch: generated text carries an imperceptible watermark embedded through patterns in the text, while supported generated files use digitally signed provenance metadata based on the C2PA standard. Anthropic states that marking for supported models applies across Claude products and wherever Claude is offered globally, while marking support for models released before 2 August is still being rolled out.

This demonstrates how EU transparency requirements are already influencing AI products beyond Europe. Organisations using generative AI should understand how provenance marks flow into their own content and records, particularly because Anthropic notes that a detected watermark only indicates that Claude was likely involved in processing the content – it cannot distinguish between content originally generated by Claude and human-authored content subsequently processed by Claude.

United States

NIST releases draft TEVV-Athlon AI evaluation framework

On **7 August 2026**, the US National Institute of Standards and Technology (NIST) released the initial public draft of NIST AI 200-2, the TEVV-Athlon framework for evaluating AI systems, with comments open until 6 October 2026. The framework provides a structured approach to testing, evaluation, verification and validation (TEVV) of the real-world impacts and outcomes of AI systems and is designed to apply across technologies including machine-learning models, large language models, multimodal models and agentic systems.

AI governance in the US is increasingly focusing on evidence that AI systems have been tested and evaluated in the context in which they will actually operate. Organisations can use the draft framework to strengthen AI evaluation programmes by linking testing to intended use, risk and real-world outcomes and retaining evidence that systems continue to perform as intended.

India

BHASHINI and NITI Aayog expand multilingual AI for public services

On **12 August 2026**, the Digital India BHASHINI Division under MeitY and NITI Aayog signed a statement of intent to advance multilingual, voice-first AI for inclusive governance and public-service delivery. The collaboration will integrate AI-powered language technologies into NITI Aayog platforms and services, including translation models, voice bots, document and communication workflows and mechanisms for collecting and curating linguistic data.

India is increasingly treating language AI as part of its digital public infrastructure rather than as a standalone technology experiment. As multilingual AI is scaled across public services, governance should address translation accuracy, linguistic and regional bias, data provenance, accessibility, model monitoring and appropriate human review.

Singapore

Government clarifies governance expectations for agentic AI

On **5 August 2026**, Singapore's Ministry of Digital Development and Information responded to a parliamentary question on whether the country's AI governance framework and AI Verify would be extended to autonomous agentic AI and whether mandatory requirements were being considered for high-consequence uses. The ministry pointed to its existing model AI governance framework for agentic AI and emphasised clear governance structures, meaningful human accountability and risk management controls proportionate to an AI system's risks and level of autonomy. The August response did not announce new mandatory requirements.

Singapore's position reinforces that greater AI autonomy requires stronger organisational accountability even where governance remains guidance-led. Organisations deploying agents should define accountable owners, autonomy boundaries, oversight and escalation mechanisms, with controls that increase in strength as the consequences and autonomy of the system increase.

Australia

AI Safety Institute publishes multi-agent systems risk framework

On **10 August 2026**, Australia's AI Safety Institute published a report on risks and controls for AI agents interacting across organisational boundaries, describing it as a world-first systematic technical framework for mapping selected risks, controls and responsibility. The framework considers agent interactions under singular, federated and open governance environments and identifies risks including cascading errors; propagation of malicious instructions or sensitive information; collusive behaviour; and failures affecting shared infrastructure.

The report highlights an important shift in agentic AI governance: individually safe agents do not necessarily create a safe multi-agent system. Organisations deploying agents that interact with suppliers, customers, partners or other external agents should assess interaction-level risks, shared governance arrangements, agent identity, hand-offs, monitoring and controls for cascading failures.

United Kingdom

Legal services AI growth lab opens

On **3 August 2026,** applications opened for the UK's legal services advisory AI growth lab, a regulatory sandbox designed to help AI innovators and adopters navigate existing regulatory frameworks. The lab provides coordinated access to legal services and information regulators to clarify how current requirements apply to emerging AI products and services, while making clear that participation does not provide regulatory approval, endorsement or exemption from existing legal obligations.

The growth lab illustrates the UK's continuing sector-led approach to AI governance, using coordinated regulatory engagement rather than waiting for a single horizontal AI law. Organisations developing regulated AI products should consider early engagement with relevant regulators and identify cross-cutting issues such as data protection, professional duties, accountability and customer outcomes before deployment.

China

Draft road traffic law introduces autonomous-driving chapter

On **25 August 2026,** a draft amendment to China's road traffic safety law was submitted to the standing committee of the National People's Congress for its first review, introducing a dedicated chapter on autonomous vehicles. The draft defines autonomous-driving and driver-assistance functions; sets conditions for autonomous vehicles to operate on public roads; and addresses traffic violations and insurance arrangements. Where a road-safety violation occurs while autonomous-driving functionality is activated, the draft provides for the vehicle manufacturer or importer to handle the violation.

China is moving towards embedding autonomous AI systems within mainstream legal frameworks and allocating responsibility for their operation. As the proposal remains under legislative review, stakeholders – including automotive organisations - should monitor the final text closely, particularly the rules governing operating conditions, insurance, system classification and manufacturer or importer responsibility.

Contact Keypoint's technology consulting team:

Srikant Ranganathan
Senior Director
srikant.ranganathan@keypoint.com

Darrshan Manukulasooriya
Executive Manager
d.manukulasooriya@keypoint.com

Sagar Rao
Executive Manager
sagar.rao@keypoint.com

Mahesh Girish
Assistant Manager
mahesh.girish@keypoint.com

Talal AlQahtani
Consultant
talal.alqahtani@keypoint.com