HealthManagement
Promoting Management and Leadership

AI Security Risks Demand Built-In Defences in Healthcare

Artificial intelligence is reshaping healthcare by accelerating diagnostics and streamlining operations, yet its rapid adoption has expanded exposure to cyber threats. Patient data has long been a target for attackers and artificial intelligence depends on large volumes of such data to function and improve. Small manipulations of inputs can trigger harmful or misleading outputs, and poisoning or corruption of training data can compromise entire systems. With artificial intelligence now embedded across clinical and administrative tools, the attack surface has widened and vulnerabilities have multiplied. Ensuring safe, reliable and trustworthy use requires security to be embedded from the outset and sustained throughout the lifecycle so that performance gains do not come at the cost of safety, confidentiality or service continuity.

Trust Hinges on Robust Protection

Trust in artificial intelligence is inseparable from trust in its security. Technology now touches many points in the interaction between patients, clinicians and systems, from diagnostic support to workflow automation. If any part of this ecosystem is compromised, confidence can erode quickly and adoption can stall. Adversarial attacks exploit the sensitivity of models to small perturbations, while data poisoning and model theft can distort or degrade outputs in ways that may be difficult to detect. Evidence cited in the source shows that altering a very small fraction of training tokens with medical misinformation increased the likelihood of clinical errors, underscoring how limited interference during training can later surface as unsafe recommendations. Because artificial intelligence is intertwined with processes across departments, weaknesses in one component can propagate through connected systems, creating operational disruption and patient safety risk. Protecting data, models and outputs end to end is therefore foundational to sustaining clinical confidence and organisational resilience.

Must Read: ***AI Scribes Promise Relief but Raise Safety and Trust Risks***

Continuous Risk Management Across Deployment

Security for artificial intelligence must be treated as a strategic discipline that spans procurement, development, validation, deployment and routine use. Embedding protections early helps ensure that controls persist as models and datasets evolve. Data validation, continuous monitoring and clear governance are central to maintaining assurance over time. Regular risk assessments should track how changes in data, configuration or integration points alter the threat profile, while real-time monitoring supports timely detection of emerging issues.

Testing needs to reflect scenarios unique to artificial intelligence. Targeted exercises that probe how models respond to manipulated inputs, corrupted training data or unexpected operational conditions can expose weaknesses before they affect care. As adoption accelerates, collaboration across organisations becomes increasingly important to establish consistent expectations for secure development and deployment and to inform governance that keeps pace with technological change. Shared benchmarks and aligned assurance practices reduce variability, clarify requirements for implementers and support safer innovation across the health system. In combination, these measures help ensure that the benefits of artificial intelligence are not undermined by preventable vulnerabilities introduced during rapid rollout.

Clinician Readiness as a Safety Layer

Technical controls are necessary but, on their own, insufficient. Clinicians interact with artificial intelligence outputs at the point of care and can recognise when results conflict with clinical judgement or patient context. Education therefore becomes a critical layer of defence. Training that explains how tools function in practice, highlights signs of manipulation or drift and reinforces critical appraisal equips users to spot anomalies early. Scenario-based learning can demonstrate how subtle changes in inputs produce incorrect conclusions, turning abstract risks into concrete lessons that shape day-to-day vigilance.

Ongoing learning is essential as systems evolve and new threat patterns appear. When clinicians understand where artificial intelligence supports decision-making and where it may fall short, they are better positioned to escalate concerns, request review and contribute to safer deployment. Informed users effectively create a human firewall that complements technical safeguards and governance. Their role helps ensure that artificial intelligence augments expertise, maintains patient safety and supports consistent, reliable delivery of care.

Artificial intelligence is driving meaningful improvements in speed, accuracy and efficiency across healthcare, yet the same mechanisms that enable these gains introduce significant cybersecurity risk. The path to safe adoption rests on recognising that security is not a bolt-on feature but a prerequisite for trust. Protecting data, models and outputs across the lifecycle, aligning assurance through realistic testing and shared standards and equipping clinicians to interrogate results provide a balanced approach that addresses risks without slowing progress. By embedding security from the outset and sustaining it through continuous risk management and education, healthcare organisations can realise the benefits of artificial intelligence while safeguarding patients, services and confidence in digital care.

Source: **HIT Consultant**

Image Credit: **iStock**

Published on : Tue, 18 Nov 2025