HealthManagement
Promoting Management and Leadership

Recognising Errors in Radiology AI Implementation

Artificial intelligence promises faster reporting, greater consistency and streamlined workflows in radiology, yet implementation often falls short. Underperformance on real-world data, workflow disruptions, added costs and clinician distrust recur across settings. Problems arise through the AI product lifecycle, within technical infrastructure and across human factors that shape everyday use. Reported issues include bias and generalisability gaps, limited external validation, model and data drift after deployment, bottlenecks in integration with clinical systems, and ergonomic or cultural barriers that impede adoption. Mapping where and why implementations fail helps radiology departments plan safer procurement, deployment and oversight, with steps ranging from clearer intended-use definitions and rigorous testing to continuous monitoring, better interfaces and targeted education that builds trust without overreliance.

Failures Across the AI Lifecycle

Errors can occur from inception to retirement. Early in development, late engagement with clinicians and patients undermines usability, while vague intended-use statements misalign tools with clinical pathways and complicate procurement. Even where clearance requires declared intended use, the information may be hard to find, increasing the risk of use in unintended contexts.

Must Read: ***Generalist Radiology AI: Advancing Finance, Operations and Care***

Bias and robustness gaps frequently reflect training datasets that under-represent subgroups by age, sex, race or socioeconomic status, with downstream risks of over- or underfitting and false predictions. Larger datasets do not guarantee better accuracy if data quality is weak. Balancing datasets and federated learning are cited as ways to improve generalisability when populations differ across sites.

Testing often focuses on internal accuracy rather than fairness, safety, usability, explainability or productivity. Performance can drop on external cohorts when tools are evaluated beyond training institutions. Multi-site benchmarking, careful ground truthing, avoidance of data leakage, appropriate sample sizes and clinically relevant metrics reduce later surprises.

Deployment exposes further vulnerabilities. Real-world variation, including anatomical differences, rare pathologies, post-operative changes, foreign bodies and image artefacts, can degrade performance. Accuracy varies by pathology type or lesion location, and operating thresholds influence sensitivity and false positives. Post-market, model drift and data drift arise from software updates, changing demographics, new diseases and user variability. Continuous surveillance aligns with lifecycle-based regulatory expectations for human oversight and real-world monitoring. Adaptive clinical decision support (CDS) requires particular vigilance as performance evolves with new data.

End-of-life decisions also matter. Persisting data within products, residual vendor access and local or national storage requirements must be addressed so that privacy and security risks do not persist after decommissioning. Governance should span inception through retirement to ensure an orderly, safe shutdown.

Infrastructure Gaps That Derail Adoption

Technical infrastructure is a frequent point of failure. Many organisations keep AI compute and storage segregated from clinical systems, limiting continuity of care and data integration. Whether on-premises or cloud, reliable data orchestration is essential so that the right Digital Imaging and Communications in Medicine (DICOM) and Health Level Seven (HL7) messages reach the correct algorithms. Standards-based interoperability supports safe information flow.

Integration with picture archiving and communication systems (PACS) and radiology information systems (RIS) remains challenging. One-way

transfers can expose referring clinicians to AI outputs without radiologist mediation. Radiologists need to edit or confirm results inside the reporting environment and store the final version in PACS. Convenience features such as push-to-PACS reduce manual steps. As models begin pre-populating report text, tight RIS integration is needed so radiologists can review and modify content before signing off. Unified platforms or widget-based interfaces that host multiple vendors can reduce error-prone context switching and simplify governance.

Hardware constraints influence reliability. Graphics processing units are commonly required for training and inference, including for MRI analysis, but they are costly and concentrated in a narrow supplier base. Suboptimal computational resources can slow or limit performance. Although infrastructure failures are underreported compared with model or human factors, rising computational demands make this a growing risk that needs proactive resourcing.

Human Factors, Bias and Workflow Consequences

Socio-cultural dynamics shape adoption. Resistance to change is common where workloads are high and time is scarce, so structured change management, coproduction and inclusion in decision making support uptake. Publication bias underreports failures and fosters a black-box culture that inflates expectations and encourages overreliance. Adherence to reporting standards remains inconsistent, with failure analysis among the least covered items in guideline assessments.

Ergonomics and interface design affect efficiency and safety. Fragmented, complex interfaces outside the reporting workflow increase errors and dampen acceptance. Embedding AI within a single, user-friendly platform reduces friction. Emotional and perceptual impacts are also noted. Some radiologists report higher burnout when tools add steps, training is limited and decision pressure rises. Automation bias can increase false positive and negative rates, especially among less experienced users exposed to incorrect outputs. Reading times can increase when normal studies are flagged as abnormal or when tools require additional interactions.

Cognitive factors shape both development and use. Annotation quality affects supervised learning, with fatigue and inconsistent labelling undermining reproducibility. Self-supervised strategies are described as promising for improving annotation efficiency and performance, while natural language methods combined with computer vision have been proposed to cross-check reports against images. Operational impact on throughput can be minimal or negative when tools add steps or noise, underscoring the need for long observation windows and continuous post-deployment evaluation. Regulation adds complexity but provides expectations for safety, testing, data protection, accountability and change control to manage drift.

Radiology AI fails for interconnected reasons across lifecycle stages, infrastructure and human factors. Mitigation relies on clear intended-use definitions, inclusive and high-quality data, multi-site benchmarking and continuous post-market surveillance to track drift and safeguard performance. Seamless integration with PACS and RIS, standards-based data orchestration and adequate compute are preconditions for reliable use. Adoption improves when interfaces are ergonomic, change is actively managed and education raises AI literacy without encouraging overreliance. Stronger reporting standards, multiprofessional collaboration, effective leadership and realistic business cases help stabilise implementation so AI can support clinicians and benefit patients.

Source: European Journal of Radiology

Image Credit: iStock

Published on : Sat, 11 Oct 2025