
Radiology AI Faces EU AI Act Compliance Hurdles



Artificial intelligence is now embedded across radiology workflows, supporting tasks that affect clinical decision-making and patient outcomes. Within the European Union, new regulatory obligations apply as the EU AI Act enters force for general-purpose and foundation models, with a transition period of up to two years for high-risk systems such as medical devices. Most radiology applications will fall into the high-risk category because they process personal health data, influence expert judgements and can shape quality of life. Compliance therefore extends beyond existing Medical Device Regulation processes and must align with rights in the Charter, including data protection (Article 8), non-discrimination (Article 21), equality between men and women (Article 23) and access to health care (Article 35). The Act's flexibility aims to fit diverse technologies yet leaves practical gaps that risk inconsistent implementation by providers and deployers.

Risk Management Systems Under Article 9

High-risk AI systems must operate under a risk management system (RMS) as required by Article 9, building on familiar device risk practices but extending them to development and deployment. The Act offers limited detail on how to design an RMS or identify known, reasonably foreseeable and less likely risks, opening space for divergent approaches even among systems with similar purposes and technical designs. Some providers may map closely to ISO 14971, while others may rely on simplified internal checklists. Variability in metrics, thresholds and role assignments can lead to different risk identifications and mitigation choices for the same hazard, with uneven effectiveness.

Must Read: [Hidden Environmental Footprint of Radiography and Fluoroscopy](#)

For radiology AI, multidisciplinary input is essential to surface the breadth of risks and to reduce blind spots across clinical, technical and operational dimensions. Although the RMS requirements are intentionally broad to accommodate many use cases, that breadth can complicate consistent, rights-preserving implementation. Without clearer expectations, two systems addressing the same task may arrive at distinct analyses and controls, creating potential inconsistency in how fundamental rights and safety are protected in practice.

Data Governance and Dataset Fitness Under Article 10

Article 10 requires training, validation and test datasets for high-risk systems to be sufficiently representative and complete for the intended purpose. Much radiology AI research has used limited or non-representative datasets and has not routinely shared raw anonymised data, rich dataset descriptions or quantitative methods for assessing representativeness and completeness. Commercial offerings also often lack dataset transparency. Under Article 13, greater transparency and detailed information for deployers will be needed so results can be interpreted appropriately.

Determining whether data is representative and complete remains challenging in the absence of standardised quantitative methods. A contextual example illustrates the point: developing a chest radiograph tool to detect lung cancer for a busy public hospital in Berlin would require attention to demographic diversity, inclusion of immigrant communities, socioeconomic variation and environmental exposures such as urban air pollution. Intra-variability within features matters, for instance, differences in smoking intensity should be captured, with adequate representation of each relevant subgroup. The dataset should reflect the local disease spectrum, including cancer subtypes, stages and comorbidities and mirror realistic prevalence.

Completeness also extends beyond size. Expert annotations, ideally with consensus from multiple board-certified radiologists, should detail lesion location, size and type. Clinical metadata such as smoking history, symptoms and family history strengthen clinical relevance. Temporal and longitudinal data can capture progression, treatment response or remission. Operator-to-operator variability, image quality differences, X-ray machine variation and common artifacts—including hair, jewellery, motion blur, overlapping structures and skin fold—should be represented so

the model learns to handle real-world variability. Absent standardised measures for representativeness and completeness across contexts, there is a grey area that can affect equity of access and risk underdiagnosis in certain groups. Beyond data, a broader need for standardisation spans algorithm transparency, risk management, data security and regulatory harmonisation, as leading jurisdictions pursue differing pathways.

Post-Market Monitoring and Real-World Performance

Article 72 requires continuous post-market monitoring (PMM) throughout the lifecycle of high-risk AI systems. RMS and data governance are integral to PMM and should be tailored to the system's design and purpose. However, what constitutes a comprehensive, effective PMM—its architecture, user interface and communication cadence with the AI system—remains unsettled. Real-world inspections have already highlighted shortcomings. In the Netherlands, the Health and Youth Care Inspectorate visited 13 medical device providers in 2023 and 2024, finding insufficient and inconsistent PMM, including poorly developed plans, partial implementation, limited PMM skills and scopes not adapted to device types, nearly half had no PMM plan, and some plans were grossly incomplete. In Switzerland, Swissmedic inspected 27 manufacturers and reported that 19 (70%) lacked adequate PMM documentation. These inspections focused on Class I devices, at the time of writing, regulatory evidence of compliance for Class III devices in the EU is unavailable.

Radiology AI raises additional PMM challenges because many public datasets are geographically and demographically limited, leading to potential underperformance in subgroups. Monitoring must detect data drift when clinical data differ from training conditions, with statistical flags at the feature level and must surface concept drift when input–target relationships evolve. Deployers may need to generate new labels during routine use to enable realistic evaluation and potential retraining. Reported testing accuracy on hold-out sets may not translate to clinical performance, so disparity testing can help reveal subgroup differences. Regulatory bodies have not yet reached definitive conclusions on PMM operations, and no notified body has been appointed to audit against the AI Act.

Radiology AI now sits within a higher bar of EU compliance, where safeguarding fundamental rights requires consistent RMS design, robust data governance and effective PMM. Yet broad regulatory provisions and uneven guidance have created practical ambiguities that can translate into variable risk controls, opaque datasets and incomplete lifecycle oversight. Addressing representativeness and completeness in context, improving transparency for deployers and operationalising drift and disparity monitoring are central to reliable performance. Greater clarity and harmonised expectations will support providers and deployers in aligning high-risk radiology AI with the AI Act while maintaining safety, fairness and access to care.

Source: [npj digital medicine](#)

Image Credit: [iStock](#)

Published on : Mon, 10 Nov 2025