AI-RADS: A Framework for Case-Level Assessment of Radiology AI Output

Development and multireader evaluation of a structured system for AI reliability, clinical utility, and report communication

Maximilian F. Russe[1], Anna Fink[1], Carl P. Simon[1], Kai Kästingschäfer[1], Fabian Bamberg[1], Alexander Rau[2]

[1]Department of Diagnostic and Interventional Radiology / [2]Department of Neuroradiology, Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Germany

RCR Global AI Conference 2026

Why AI-RADS?

Radiology AI is entering routine workflows across detection, segmentation, quantification, classification, and generative tasks. However, most systems still lack a standardized case-level method to document whether an individual output was reliable, corrected, overridden, or suppressed. AI errors, image-quality limitations, workflow failures, and radiologist overrides may remain poorly traceable. AI-RADS addresses this need by linking each AI output to a structured reliability category and a corresponding clinical action.

Regulatory need

Real-world deployment of radiology AI requires traceable oversight beyond initial validation. AI-RADS creates auditable case-level documentation. This supports privacy- and regulation-conscious governance, including GDPR/HIPAA-aware documentation workflows, EU AI Act-aligned post-market monitoring, EU MDR surveillance concepts for medical-device software, and US trustworthy-AI / FDA-style real-world performance monitoring.

Objective

To develop and evaluate AI-RADS as a structured case-level framework for assessing radiology AI output reliability, clinical utility, and required reporting action. The study tested whether radiologists can reproducibly apply AI-RADS across representative image-based and generative AI tasks.

The AI-RADS framework

1 – Excellent / optimal utility
Integrate into workflow or report

2 – Highly accurate / high utility
Use with minor confirmation or edits

3 – Acceptable / modification needed
Verify and modify before use

4 – Unreliable / limited utility
Override critical sections

5 – Unreliable / no clinical utility
Suppress from report or display

Task labels

[DET] Detection
[SEG] Segmentation
[CLS] Classification
[QNT] Quantification
[GEN] Generative AI

Context modifiers

[IL] Image limitation
[SC] Superior confirmation
[WL] Workflow limitation

Example: AI-RADS 4 [SEG, IL]

Study design

- **7 AI applications**
- **350 total cases**
- **5 radiologist readers**
- **AI-RADS category + modifiers + correctness rating**
- **Agreement and usability analysis**

200 image-based AI outputs
150 generative AI outputs

Image-based tasks: fracture detection, chest radiograph pathology detection, lung nodule detection, and stroke MRI segmentation/classification

Generative tasks: CT protocol and contrast-phase selection, Spinal Instability Neoplastic Score assignment, and patient-friendly report generation

Assessment endpoints

- Interreader agreement
- Alignment with correctness
- Modifier use
- Usability and workload

Key results

0.89	0.87	0.93	94.8%	88.3%	87.5 / 2
Overall Krippendorff's alpha	Image-based tasks Krippendorff's alpha	Generative tasks Krippendorff's alpha	Reader-rated correct AI outputs mapped to AI-RADS 1–2	AI-RADS ratings unanimous or within one category	Median System Usability Scale / NASA-TLX

Conclusion and Clinical interpretation

AI-RADS showed high reproducibility across image-based and generative radiology AI tasks and standardizes documentation of radiologist oversight. Structured case-level ratings support safer workflow integration, quality assurance, audit trails, post-market monitoring, and regulation-conscious AI governance.

Take-home message

From AI output to accountable action: AI-RADS 1-5: integrate, edit, verify, override, or suppress.
Structured for evolving regulatory requirements.

Recently published: Russe MF, Fink A, Simon CP, et al. AI-RADS: A Framework for Assessment of Artificial Intelligence Output in Radiology - Development and Multireader Evaluation. Investigative Radiology. Online ahead of print 2026. doi:10.1097/RLI.0000000000001272