

WEEKLY AI INTELLIGENCE REPORT

Radiology and Medical Imaging

Week 43 | 18 to 24 October 2026

Dr. Sergey Morozov | drsergeymorozov.com

Compiled from publicly available sources (indexed peer-reviewed literature, official press releases, regulatory databases). Does not constitute medical, regulatory, or financial advice.

8

Peer-reviewed papers

6

Industry / FDA items

3

Media highlights

2

Key opinion leaders

GOVERNANCE TAKEAWAY OF THE WEEK

Measure the pair and the place, not the model. This week a randomised crossover experiment in The Lancet Digital Health showed that ignoring correct AI advice damaged physicians' diagnostic performance more than following incorrect advice, with explanations helping mostly on hard positive cases; a prospective community screening study showed a high-AUROC triage tool delivering a 17.6% positive predictive value because disease was rare in the field; and a Turkish reporting-LLM study showed perfectly parsed output hiding incomplete clinical content. The action for a radiology leader this week: add two lines to your acceptance-test template, a reliance measure (how readers respond to deliberately mixed correct and incorrect AI output) and a prevalence-matched PPV with expected alert volume, and keep both as monthly monitoring KPIs after go-live.

SECTION 0

EXECUTIVE SUMMARY

Key trends, week of 18 to 24 October 2026

[GOVERNANCE] UNDER-RELIANCE COSTS MORE THAN AUTOMATION BIAS

In a Lancet Digital Health randomised crossover experiment, 125 physicians read head CTs with and without AI advice: dismissing correct advice damaged diagnostic performance more in absolute terms than following incorrect advice, and explainable advice lifted performance (OR 1.84) mostly on difficult haemorrhage-positive cases. Acceptance tests should measure both reliance directions, not only automation bias.

[EVIDENCE] FIELD PREVALENCE RESETS PPV

A parameter-locked mobile triage AI for oral cancer screening, run prospectively by nonspecialists on 602 community participants, kept specificity at 98.6% but delivered a positive predictive value of 17.6% on its referral alerts. A high AUROC (0.957) said little about the alert stream a low-prevalence field setting produces.

[IMPLEMENTATION] SCHEMA VALIDITY IS NOT CONTENT FIDELITY

A compact, locally deployable LLM structured Turkish spine MRI reports with 100% valid JSON and 100% grouped reader acceptance, yet one reader found negative-finding lists incomplete in 57.8% of reports, and the training targets came from another model's silver labels. Validate content against clinicians, not against the teacher model.

[EVIDENCE] ANNOTATION BUDGETS CAN SHRINK

Across 6000 radiology reports and 1260 fine-tuned models, expert-labeled transformers stayed best for negation detection (F1 up to 0.985), but LLM-distilled labels came within 0.02 to 0.04 F1 at about US \$0.018 per report, and the choice of teacher model barely mattered. A small human-labeled core set plus distilled scale is a defensible costing rule.

[OVERSIGHT] TEST THE PAIR AND THE PLACE

The week's thread: performance lives in the reader-AI pair and the deployment setting, not in the model file. Reliance patterns moved physicians' accuracy in both directions, field prevalence collapsed a screening tool's PPV to 17.6%, and perfectly parsed LLM output hid incomplete clinical content. Put a reliance measure and a prevalence-matched PPV into every acceptance test, and keep both as monthly KPIs.

SECTION 1

PEER-REVIEWED PAPERS

Source: PubMed | Verified indexing | No preprints | 18 to 24 October 2026

Development of a Novel Real-Time Computer-Assisted Colonoscopy Diagnostic Tool for Colorectal Neoplasia and Cancer to Enable Personalized Patient Management: Protocol for a Multicenter Proof-of-Concept Study.

Olabintan O, et al. | *JMIR research protocols* | 2026-10-21

A multicenter international consortium has registered the development and proof-of-concept validation of a real-time colonoscopy CADx system that classifies colorectal lesions into five categories: nonneoplastic, low-grade adenoma, high-grade adenoma, sessile serrated lesion and invasive cancer, moving beyond the binary calls of current systems. More than 17,000 deidentified images and video frames will be collected prospectively from eight European centers and one Japanese center, with semisupervised contrastive learning (SimCLR), an integrated detection module and Grad-CAM interpretability.

Key metrics: Protocol; no performance results yet. Predefined targets: multiclass diagnostic accuracy threshold of 85% as the primary outcome, plus sensitivity and specificity for invasive cancer detection; proof-of-concept validation on 250 previously unseen lesion images assessed independently by expert endoscopists. Secondary outcomes: interobserver agreement, performance across demographic and lesion subgroups, qualitative misclassification analysis. Ethics approval obtained at all participating centres.

Clinical relevance: Binary leave-in versus resect calls cap the clinical value of optical diagnosis; a five-class output maps directly onto resection technique and surveillance intervals, which is where personalised management actually happens. Results are pending, and the protocol's predefined 85% accuracy threshold and prespecified subgroup analyses set a clear, checkable bar for the proof-of-concept read-out.

PMID 42905971 DOI 10.2196/77277

Mobile AI-Assisted Oral Lesion Triage for Community Oral Cancer Screening: Prospective Field Evaluation Study.

Liu HY, et al. | *JMIR mHealth and uHealth* | 2026-10-20

Before you size a screening programme around an AI's curated-dataset accuracy, model what low prevalence does to its alert stream. In this prospective field evaluation in eastern Taiwan, trained nonspecialist staff used a parameter-locked, mobile AI triage system during routine community oral cancer screening. Of 68 lesions the system flagged red for specialist referral, 12 were confirmed high risk, and expert review found 4 further high-risk lesions the system had not flagged red.

Key metrics: 602 screened participants, 4283 interpretable intraoral images, 16 high-risk events. Red triage: sensitivity 75% (12/16, 95% CI 50.5 to 89.8), positive predictive value 17.6% (12/68, 95% CI 10.4 to 28.4), specificity 98.6% (3928/3984), NPV 99.9%, accuracy 98.5%. AUROC 0.957 (red), 0.931 (yellow), 0.964 (green); AUPRC 0.8235, 0.9052, 0.9476. Specialist pre-consensus agreement: Fleiss kappa 0.921, Gwet AC1 0.981.

Clinical relevance: High AUROC coexisted with a PPV of 17.6% because high-risk disease was rare in the field population, and the same arithmetic applies to any screening or triage AI deployed outside the enriched dataset it was reported on. So what for your AI committee: both validation and monitoring change. Require a prevalence-matched PPV estimate and the expected referral volume in the acceptance file (Article 15 accuracy), and track confirmed-positive rate per alert as a standing post-market KPI (Article 72).

PMID 42902899 DOI 10.2196/101721

TurkRadLLM: A Compact MedGemma-Based Model for Structuring Turkish Lumbar and Cervical Spine MRI Reports-A Multi-reader Validation Study.

Turgan AH, et al. | *Journal of imaging informatics in medicine* | 2026-10-20

If a report-structuring LLM was validated against another model's outputs, read that as a fidelity analysis, not a clinical validation; this study keeps the two honestly apart. The authors fine-tuned a compact MedGemma 4B model with QLoRA on 5210 lumbar and 5017 cervical Turkish spine MRI reports to produce six-field structured output on a single consumer GPU, then had radiologist readers score both the model's output and the proprietary-LLM silver labels it was trained toward.

Key metrics: Multi-reader validation on 180 authentic plus 20 adversarial reports: 100% valid, schema-conformant JSON; grouped acceptance 100% (95% CI 97.9 to 100); one reader judged negative-finding lists incomplete in 57.8% of reports. Held-out set of 1025 reports: agreement with the silver-label impression rose from 17.7% to 67.1% with fine-tuning, urgent-flag sensitivity against the silver reference from 18.8% to 62.5% (both $p < 0.001$); a much larger non-fine-tuned model reached 27.5% and 50.0%; a rule-based baseline hit 99.0% on body region. Deployment: 14.75 GB GPU memory.

Clinical relevance: Every output parsed, yet one reader found incomplete negative-finding lists in more than half of the reports: schema validity and content fidelity are different claims, and the readers' own scoring of the silver labels showed those labels are a training device, not a reference standard. So what for your AI committee: acceptance testing of report-structuring and report-mining LLMs changes. Score content against clinician adjudication on your own reports, include an urgent-flag sensitivity test, and record silver-label-only validation as a data-governance risk under Article 10.

PMID 42902038 DOI 10.1007/s10278-026-02427-8

Reducing Annotation Costs Using Knowledge Distillation for Label-Efficient Negation Detection in Radiology Reports: Multisite Internal Validation Study.

Weng KH, et al. | *JMIR medical informatics* | 2026-10-22

Before you budget expert annotation to adapt or validate report-mining NLP, test how far distilled labels get you. Across 6000 deidentified radiology reports from three hospital branches and 1260 fine-tuned transformer models, human-labeled supervision stayed best for negation and speculation detection, but pseudo-labels generated by large language models came close at a fraction of the cost, and the choice of teacher model barely mattered.

Key metrics: Human-labeled transformers: keyword-level F1 0.982 to 0.985 at $n=4200$. Distillation gap: delta F1 around 0.042 at small training scales and 0.022 at large scales (adjusted $p < 0.001$). 10-shot GPT-5.3 Chat: F1 0.974 (delta 0.009 vs human-labeled transformers, adjusted $p=0.047$), ahead of Gemma 4 (0.957, adjusted $p=0.01$) and Qwen3.6-Plus (0.950, adjusted $p < 0.001$). Labeling cost about US \$0.018 per report, cheaper than manual annotation once expert wages exceed US \$0.64 to 1.50 per hour. Reference annotations: 1000 reports, interrater kappa 0.957. Internal multisite validation only.

Clinical relevance: The practical output is a costing rule for text-pipeline governance work: keep a small expert-labeled core set as the reference and use distilled labels for scale, since dataset size mattered more than teacher quality or sampling strategy. So what for your AI committee: this changes how you resource acceptance testing and monitoring of NLP tools; the reference standard stays human (Article 10 data governance), and the framework counts as internal validation until externally replicated, as the authors state.

PMID 42914107 DOI 10.2196/103912

An interpretable contrast-enhanced CT-based cross-attention transformer model for predicting progression-free survival in muscle-invasive bladder cancer: a multicenter study.

Liu Y, et al. | *Abdominal radiology (New York)* | 2026-10-22

For muscle-invasive bladder cancer patients heading to radical cystectomy, this four-center study asks whether preoperative contrast-enhanced CT can stratify progression risk before surgery. A dual-channel cross-attention transformer produced a deep learning score that predicted progression-free survival and split patients into high- and low-risk groups with clearly different cumulative incidence of progression.

Key metrics: 427 patients with MIBC across four institutions, retrospective, split into training and validation cohorts. Validation cohort: C-index 0.774 and AUC 0.785 for the DL score, generally above the combined DL-plus-clinical signature, the clinical model and a Ki67 model as reported in the abstract.

Clinical relevance: A preoperative imaging score that stratifies progression risk could inform surveillance intensity and adjuvant treatment discussions after cystectomy. The design is retrospective with a modest validation cohort, so the finding needs prospective, external confirmation before it should steer individual patient care.

PMID 42911393 DOI 10.1007/s00261-026-05837-y

Automated carotid artery atherosclerotic inflammation segmentation in PET using MSR-Net.

Li R, et al. | *Physics in medicine and biology* | 2026-10-21

Quantifying carotid plaque inflammation on PET is limited less by tracers than by partial-volume physics at the thin vessel wall; this study attacks that bottleneck with simulation-grounded deep learning. A two-stage network localizes the carotid wall on T1-weighted MR, then estimates voxel-wise inflammation fraction from ^{64}Cu -CANF-comb PET, trained on Monte Carlo simulated images realistic enough that nuclear-medicine physicians could not reliably tell them from real scans.

Key metrics: Realism check: five physicians identified real images with 54% accuracy (54/100, Wilson 95% CI 0.44 to 0.63, $p=0.48$). On 184 simulated test slices from five held-out patients: weighted DSC 0.79 vs 0.73 (U-Net), 0.69 (Transformer U-Net) and 0.75 (DUCK-Net); HD 1.85 vs 2.29, 2.32 and 2.23 voxels; ENRMSE 0.24 vs 0.59, 0.68 and 0.59 for the DL baselines and 0.57 for iterative Yang partial-volume correction (Holm-corrected $p \leq 0.031$). In seven patients, corrected SUV associated positively with MR-defined lipid-rich necrotic core with hemorrhage.

Clinical relevance: Ground truth here is synthetic by necessity, which is both the strength and the limit: the authors state plainly that external validation against biological reference standards is required before clinical deployment. On merit, the gain over standard partial-volume correction is large, and the MR-guided two-stage design is reusable for other small-structure quantification problems in PET.

PMID 42906615 DOI 10.1088/1361-6560/aeb798

Frozen Echocardiography Foundation Model Embeddings Contain Linearly Recoverable LVOT VTI and TAPSE Report-Label Signal in MIMIC-IV-ECHO.

Karim P, et al. | *Journal of imaging informatics in medicine* | 2026-10-19

Before anyone in your network proposes a foundation model as a shortcut to echocardiography measurements, note how carefully this study bounds its own claim. Using public MIMIC-IV-ECHO data, the authors tested whether frozen EchoPrime study-level embeddings, with no fine-tuning, carry linearly recoverable signal for two structured report values: LVOT velocity-time integral and TAPSE. They do, beating a median null model and non-image baselines built on age, sex and acquisition metadata, but with moderate fit.

Key metrics: LVOT VTI: 3782 studies (2654 train, 573 validation, 555 test); ridge regression test MAE 3.64 cm vs 4.54 cm for the null model (95% CI 3.41 to 3.88), RMSE 4.66 cm, R^2 0.372, Pearson r 0.613 (0.556 to 0.664). TAPSE: 1131 studies (787, 184, 160); MAE 3.17 mm vs 3.79 mm null (95% CI 2.80 to 3.56), RMSE 3.94 mm, R^2 0.284, Pearson r 0.539 (0.424 to 0.632). Deterministic subject-level splits; non-image baselines weaker than imaging embeddings.

Clinical relevance: An R^2 of 0.28 to 0.37 shows that the embeddings encode measurement-related signal, not that they can replace a measurement, and the authors say so explicitly: the findings support representation-level experiments, not measurement extraction, replacement, clinical deployment or decision support. The design is a clean template for probing foundation models, with subject-level splits, validation-only hyperparameter selection and leakage-safe non-image baselines.

PMID 42897427 DOI 10.1007/s10278-026-02411-2

Examining reliance patterns on AI advice in medical imaging: a mixed-methods, randomised, crossover experiment.

Gaube S, et al. | *The Lancet. Digital health* | 2026-10-23

Do not assume explainable AI has solved over-reliance, and do not assume over-reliance is your main exposure. In this preregistered randomised crossover experiment, 50 radiologists and 75 non-radiologist physicians and trainees each read 50 head CT cases for intracranial haemorrhage under three conditions: no AI advice, basic AI advice, and explainable AI advice with scan annotations, with 80% of the AI advice correct. Both failure modes occurred, and ignoring correct advice cost more diagnostic performance in absolute terms than following incorrect advice.

Key metrics: 125 participants (50 task experts, 75 novices), 50 head CT cases per session, three conditions in randomised order; 80% of AI advice correct. Correct explainable advice vs no advice: diagnostic performance OR 1.84 (95% CI 1.49 to 2.28, $p<0.0001$); confidence $b=0.15$ (0.09 to 0.22); reading time not significantly reduced ($b=-1.81$, $p=0.071$); perceived usefulness above basic advice $b=0.12$ ($p=0.0029$), especially among novices.

Clinical relevance: Under-reliance clustered where uncertainty was high, and explainable advice reduced it most on difficult haemorrhage-positive cases, which is the opposite of the usual committee worry about automation bias. So what for your AI committee: this changes acceptance testing and human-oversight design under EU AI Act Article 14. Test local readers with deliberately mixed correct and incorrect AI outputs, measure reliance in both directions by experience level, and monitor override rates after go-live rather than assuming explanations alone calibrate trust.

PMID 42919730 DOI 10.1016/j.landig.2026.101050

SECTION 2

INDUSTRY & REGULATION

Sources: Official registers | Press releases | FDA databases

[FDA] Elucid announces FDA clearance of FFR-IQ, BioIntegrated FFR-CT software

Elucid | 21 October 2026

Cleared for what: 510(k) clearance (K262015) for noninvasive estimation of fractional flow reserve from coronary CT angiography; the software folds Elucid's histology-validated plaque quantification into the physiology estimate and reports alongside Plaque-IQ from a single CCTA. On what evidence: the vendor cites its ELUCIDATE FFR-CT validation study, sensitivity 92.2% and specificity 93.1%, and calls it the highest reported among cleared FFR-CT products; the claim is footnoted to the FDA 510(k) summary dated 29 September 2026, and the announcement is a press release, so ask for the peer-reviewed publication. Questions for the vendor: reference standard and per-vessel versus per-patient analysis; sites, scanners and CCTA quality-rejection rate; calibration and handling of the 0.75 to 0.80 gray zone; head-to-head data against the FFR-CT incumbent you already use. Monitoring regardless of clearance: track your local rejection rate and concordance with invasive FFR for revascularized cases.

Source: [Business Wire](#)**[FDA] 3DR Labs and Deep Reasoning AI clear NeuroDR for autonomous neurovascular 3D post-processing**

3DR Labs | 20 October 2026

Cleared for what: a 510(k) for fully automated 3D segmentation, anatomical modeling and vessel centerlining of head and neck CTA, returning MIPs, volume renderings and curved planar reformats to PACS with no technologist clicks. This is a post-processing and visualization clearance, not a detection or triage claim: the stroke decision-making language in the release describes workflow speed, not diagnostic performance. On what evidence: vendor materials; the release cites no published accuracy study. Questions for the vendor: the cleared indications and predicate; segmentation and centerline accuracy on artifacted or poorly opacified scans; how failures are flagged and routed to a human; turnaround under load. Monitoring: compare a monthly sample of automated reconstructions against technologist-built ones and log implausible centerlines before anyone treats the output as diagnostic-ready.

Source: [3DR Labs](#)**[FDA] RIVANNA BoneEnhance cleared as first ML module on the Accuro XV ultrasound platform**

RIVANNA | 22 October 2026

Cleared for what: a machine learning module (K262093) that segments bone and surrounding tissue on Accuro XV musculoskeletal ultrasound volumes and supports 3D skeletal visualization during image review. Anatomy labeling only: the injury-detection module on the roadmap is investigational and not cleared. On what evidence: an eight-site study with a genuinely independent split, three sites for training and five for testing, 233 volumetric datasets, 69,900 images, 173 subjects, per-volume label accuracy 95.1 to 100% against three board-certified radiologists; unusually clean site separation for a 510(k) summary and worth holding other vendors to. Questions for the vendor: subgroup performance by body habitus and age, reference-reader variability, and behaviour on atypical or postsurgical anatomy. Monitoring: confirm in your own acceptance test that overlays never obscure the underlying B-mode image in review workflows.

Source: [PR Newswire](#)**[FDA] Coreline Soft clears AVIEW Aorta Metrics for quantitative aortic assessment on chest CT**

Coreline Soft | 23 October 2026

Cleared for what: automated aortic segmentation and zone-wise diameter measurement on chest CT, with straightened CPR, cross-sectional and 3D views and analysis reports; the clearance was received on 16 October and is the vendor's thirteenth 510(k). A quantification tool, not an aneurysm detection or triage claim. On what evidence: the announcement quotes no clinical performance figures; vendor materials position it within an opportunistic-screening portfolio. Questions for the vendor: measurement agreement against expert manual measurement on your scanners and protocols, contrast and non-contrast performance, zone-classification failure rate, and whether reported diameters map onto the measurement conventions your surveillance thresholds assume. Monitoring: re-measure a fixed reference set after every scanner or protocol change; diameter drift of a millimetre moves surveillance decisions.

Source: [PR Newswire](#)

[FDA] NeuroXT NEXT-Cue cleared for prognostic assessment support in mild cognitive impairment*NeuroXT | 20 October 2026*

Cleared for what: a Class II 510(k) for AI software that integrates brain MRI, demographics and cognitive assessment data to provide supplementary information supporting assessment of progression from mild cognitive impairment to Alzheimer's dementia within five years. On what evidence: the announcement cites no performance figures and names no published validation; evidence level is vendor materials until the 510(k) summary is read. Questions for the vendor: development and validation cohorts and whether they resemble a real memory-clinic casemix rather than research registries; discrimination and calibration by age, sex and education; how the five-year risk output is worded in the UI to avoid anchoring clinicians. A five-year prognostic claim cannot be locally verified quickly, so the full validation report, not a demo, is the acceptance document here.

Source: [PR Newswire](#)**[PARTNERSHIP] Optellum lung nodule suite joins RapidAI's enterprise platform***Optellum | 20 October 2026*

Optellum's Virtual Nodule Clinic with Lung Cancer Prediction AI is now deployable through the Rapid Enterprise Platform, used in more than 2,500 hospitals; Optellum describes it as the first FDA-cleared and Medicare-reimbursed image-based lung cancer risk stratification tool (CPT 0721T and 0722T). A distribution and integration deal, not new clinical evidence: clearance and reimbursement status are unchanged, what changes is the deployment route and the platform's shared governance tooling. Questions before signing: who holds and who can export the monitoring data the platform collects, how model versions are controlled when two vendors update independently, and whether per-study pricing survives platform bundling.

Source: [Optellum](#)

SECTION 3

POST-MARKET SIGNAL*Drift | Recalls and MAUDE | Real-world monitoring | EU AI Act Article 72***FDA-cleared PE AI runs below clearance benchmarks across 17 real-world facilities***Health AI Central / npj Digital Medicine | 22 October 2026*

An Emory-led evaluation published in npj Digital Medicine analyzed 67,869 CT scans across a 17-facility academic health system running two FDA-cleared Aidoc BriefCase algorithms. The PE triage model reached 86.8% sensitivity on 30,678 CTPAs and the incidental-PE model 73.5% on 37,191 routine contrast CTs, both below their clearance benchmarks, while specificity exceeded cleared levels at 99.1% and 99.8%. Sensitivity fell from 99.3% for saddle emboli to 72.9% for subsegmental and from 89.7% for acute to 65.3% for non-acute PE. Reference labels came from radiology reports via a validated LLM pipeline (97% accuracy), which readers had written with the AI output visible, so the true gap may be larger. The practical reading: a negative flag cannot end the search for small or chronic clots, and clearance-summary sensitivity is not a deployment estimate.

Link: <https://healthaicentral.com/news/fda-cleared-ai-for-pulmonary-embolism-detection...>**RSNA sessions report: real-world ICH triage sensitivity 82% versus 96% in clearance data***Diagnostic Imaging Europe | 21 October 2026*

A DI Europe report on three RSNA sessions assembles the post-deployment evidence in one place. At Emory, validated LLM-generated labels audited commercial triage algorithms across more than 101,000 intracranial hemorrhage cases: real-world sensitivity was about 82% against 96% in the FDA clearance data presented, with additional subgroup weaknesses. A 227-radiologist chest X-ray study found the average effect of AI assistance on reader performance was zero, hiding extreme winners and losers that experience could not predict, with automation neglect more frequent than automation bias. The framing worth stealing for your committee: AI outputs as an always-on quality sensor, and the post-deployment question shifting from will it work to is it still working for us.

Link: <https://dieurope.com/it-ai/from-prompts-to-performance-what-it-takes-to-make-ai-...>

Lancet commentary: cardiovascular imaging AI faces endless retraining unless monitoring is designed in

The Lancet Medical Imaging and Theranostics | 20 October 2026

A comment in The Lancet Medical Imaging and Theranostics names the paradox now visible across cardiovascular imaging AI: models keep getting more accurate while durable performance gets harder to maintain, because scanner generations such as photon-counting CT, evolving disease definitions and new therapies continuously shift the data away from what the model was validated on. The authors' point for deployers is that the implicit assumption behind every locked model, that tomorrow's data resemble the development data, rarely holds for long in imaging, which makes planned monitoring and revalidation a design requirement rather than an afterthought.

Link: [https://www.thelancet.com/journals/lanmit/article/PIIS3117-4523\(26\)00035-3/fullt...](https://www.thelancet.com/journals/lanmit/article/PIIS3117-4523(26)00035-3/fullt...)

SECTION 4

PLAYBOOK SNIPPET

One concrete step your AI committee can take this month

EU AI Act Article 10 (data governance)

Step: Add one clause to your next imaging-AI RFP: the supplier discloses the demographic and scanner composition of training and test data, and reports performance by subgroup. Record non-disclosure as a risk item under Article 9 rather than treating silence as acceptable.

Why: Diversity, not volume, is the specification worth writing in: a published bias audit found the model trained on the most homogeneous cohort was both the worst and the most biased, while a moderately sized but diverse cohort beat a much larger one (Lee RS et al., AJNR, PMID 41663205, DOI 10.3174/ajnr.A9217).

Worked example: This week's multicenter colonoscopy CADx protocol (Olabintan O et al., JMIR Research Protocols, PMID 42905971, DOI 10.2196/77277) shows what good composition looks like when it is designed in rather than disclosed under pressure: more than 17,000 images from eight European centers and one Japanese center, a broad demographic spectrum, and subgroup performance predefined as a secondary outcome. Demand that level of composition reporting from vendors in your own procurement and audits.

SECTION 5

MEDIA HIGHLIGHTS

AuntMinnie | Radiology Business | The Imaging Wire | Diagnostic Imaging | ITN

Generative AI policy is a White House priority, FDA digital health chief says

MedTech Dive | 21 October 2026

At AdvaMed's MedTech Conference in Boston, Rick Abramson, the radiologist directing the FDA's Digital Health Center of Excellence, said draft guidance on generative AI enabled devices is coming within months and that the policy push is a priority at the level of the HHS Secretary and the White House. The comment period on the FDA's generative AI discussion paper closed on 20 October with more than 500 submissions. A draft on risk assessment and evidence generation for generative AI devices is under construction, and the UK's MHRA is moving in parallel after adopting 44 recommendations on AI regulation including lifecycle performance monitoring.

Link: <https://www.medtechdive.com/news/gen-ai-policy-is-a-white-house-priority-cdrh-di...>

Providers to FDA: developers, not doctors, must be accountable for generative AI devices

Healthcare Dive | 22 October 2026

In comments to the FDA on generative AI regulation, the American Hospital Association, American Medical Association and American Telemedicine Association argued that monitoring generative AI devices would overload providers, especially small and rural practices, and that developers should carry the post-market burden, with FDA-run channels for providers to report problems. The AMA added a point radiology committees should note: a human in the loop does not automatically mitigate risk, because reviewers make errors and over-rely, so the humans overseeing AI need evaluating too.

Link: <https://www.healthcaredive.com/news/developers-not-doctors-held-accountable-gene...>

ESER column: the unannounced guest, client-supplied AI in teleradiology

AuntMinnie Europe | 22 October 2026

The European Society of Emergency Radiology's Committee on Teleradiology published a column on a governance gap most frameworks ignore: AI tools selected and supplied by the client hospital that arrive in the teleradiologist's worklist unannounced. The committee discusses how AI should be implemented to support rather than surprise the reading radiologist, who may neither know the tool's validation status nor control its thresholds, and argues implementation standards must cover the teleradiology chain, not just the deploying department.

Link: <https://www.auntminnieeurope.com/imaging-informatics/artificial-intelligence>

SECTION 6

PROFESSIONAL SOCIETIES

SIIM | ACR | RCR | Official publications and event coverage

RCR publishes second edition of guidance on integrating AI into RIS and PACS reporting workflows

21 October 2026

The Royal College of Radiologists published the second edition of Integrating image analysis artificial intelligence with radiology reporting workflows (RIS and PACS), updated for developments since the 2021 first edition. The guidance targets radiologists, reporting radiographers, sonographers and procurement teams, and aims to integrate AI without increasing reporting burden; working group chair Amaka Offiah framed patient safety, clinical need and real-world workflows as the center of implementation.

Source: <https://www.rcr.ac.uk/news-policy/latest-updates/integrating-image-analysis-artificial-intelligence-with-radiology-reporting-workflows-ris-and-pacs-second-edition-published/>

RSNA and ACR file comments on FDA generative AI regulation

19 October 2026

RSNA submitted comments to the FDA on 19 October arguing that generative AI risks cannot be addressed by one-time premarket review and calling for continuous monitoring of model performance, human-AI interaction and workflow effects, with manufacturers rather than radiologists accountable for device safety. ACR's comments urge the FDA to regulate the complete clinical function of a tool, balance premarket flexibility with real-world performance monitoring, and measure clinically consequential failures such as false assertions and omissions rather than generic hallucination rates; both societies asked the agency to convene radiology-specific consultation before draft guidance.

Source: <https://radiologybusiness.com/topics/artificial-intelligence/radiology-societies-weigh-fdas-push-regulate-generative-ai>

SECTION 7

KEY OPINION LEADERS

LinkedIn | Public statements | Week of 18 to 24 October 2026

Nina Kottler

Chief Medical AI Officer, Mosaic Clinical Technologies (Radiology Partners)

Posted on 24 October an expert-insights piece with CARAVAN Wellness, Making Radiology More Understandable With AI, arguing for moving from efficiency AI to opportunity AI. Her governance point: patients now often read imaging reports before their physician does, and access to information is not the same as understanding it; as AI surfaces more findings, including incidental ones, radiology should help patients and referrers understand what a finding means and what should happen next, which in her words requires thoughtful implementation, appropriate clinical oversight and ongoing governance, with AI supporting rather than replacing physician judgment.

Source: [LinkedIn](#)

Curt Langlotz

Professor of Radiology, Medicine and Biomedical Data Science, Stanford; Director, Stanford AIMI

Shared on 24 October a Stanford AIMI post on a STAT article about radiology practices building or embedding AI in-house, in which he and Nina Kottler are featured. The thread's central claim: owning a practice gives developers the clinical setting to evaluate AI, and Kottler describes radiologists' corrections of AI-generated reports at Mosaic as reinforcement learning. Woojin Kim of ACR DSI, also quoted and resharing, supplied the necessary counterweight for any committee: the practice still has to show the benefit.

Source: [LinkedIn](#)

QUALITY ASSURANCE NOTE

Compiled by Dr. Sergey Morozov from publicly available sources: peer-reviewed papers indexed in PubMed (PMIDs and DOIs verified, no preprints or arXiv), official press releases, regulatory databases and specialist media. It does not constitute medical, regulatory or financial advice. All papers have publication dates verified within 18 to 24 October 2026. Paper selection is deterministic (pinned PubMed query, Q1/flagship journal allow-list, exclusion of pure reviews [meta-analyses kept], radiomics-only and interventional-radiology work, then a transparent ranking score). Industry items come from official press releases / regulatory notices; KOL items from public LinkedIn activity within the window.