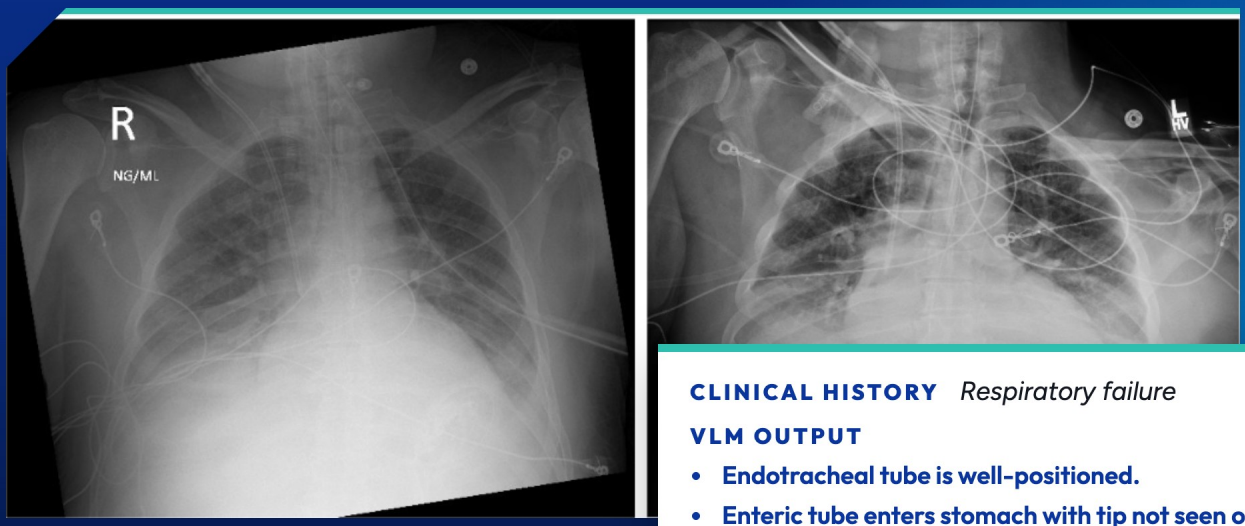
RESEARCH PAPER · X-RAY REPORT DRAFTING

Vision-Language Models for Radiograph Interpretation Can Improve Report Quality and Radiologist Efficiency

Mosaic Clinical Technologies, Inc. & Cognita Imaging, Inc.

VLM OUTPUT EXAMPLE · CHEST RADIOGRAPH



CLINICAL HISTORY *Respiratory failure*

VLM OUTPUT

- Endotracheal tube is well-positioned.
- Enteric tube enters stomach with tip not seen on this image.
- Right internal jugular central venous catheter terminates at the cavoatrial junction.
- Stable small bilateral pleural effusions.
- Mild pulmonary edema, similar.
- Stable hazy bibasilar lung opacities.

57.5%
highest rEDR: lung nodules/masses, Chest Radiograph

166,198
radiographs in performance testing

31%
radiograph drafting efficiency gain

RESEARCH PAPER SERIES

CT X-RAY

Vision-Language Models for Radiograph Interpretation Can Improve Report Quality and Radiologist Efficiency

Mosaic Clinical Technologies, Inc. & Cognita Imaging, Inc.

DRAFTED EXAMS (ALL TIME)

6.33M

Radiographs drafted through Oct. 22, 2026 (see Table 4)

DRAFTED REPORTS PER DAY

25,000+

Radiograph exams drafted per day, 30-day average through Oct. 22, 2026

EXAMS COVERED

Chest Radiograph · Abdomen Radiograph · Musculoskeletal Radiograph

KEY FINDINGS

Highest relative enhanced detection rate (rEDR) by radiograph model: additional expert-confirmed findings identified by the vision-language model (VLM) compared with original radiologist reports (retrospective analysis).

57.5%

Chest Radiograph
Lung Nodules/Masses

45.8%

Musculoskeletal and Abdomen Radiograph
Suspicious Bone Lesions

166,198
radiographs in performance testing; 713 manually reviewed

89.3%
of Chest Radiograph VLM drafts rated "good" (no/minor edits), vs. 73.6% of radiologist reports

31%
radiograph drafting efficiency gain (weighted, radiologists with >50 drafted exams)

01 PROBLEM-ORIENTED SUMMARY

INTRODUCTION

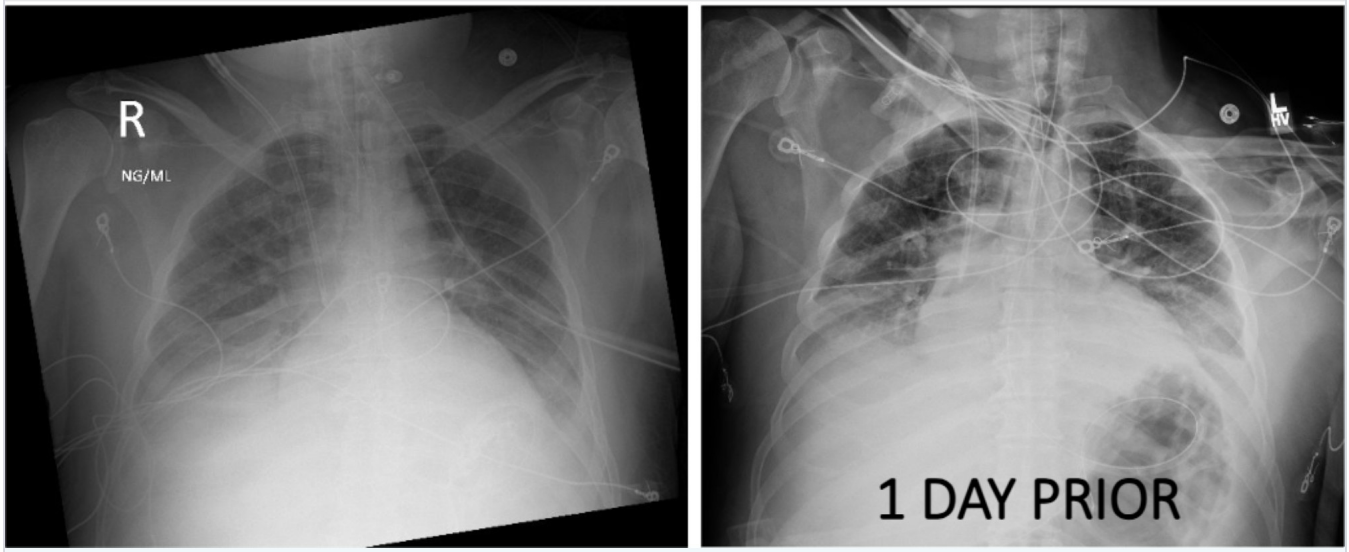
Radiology faces a growing mismatch between imaging demand and the available physician workforce, creating an urgent need for technologies that improve efficiency without compromising diagnostic quality.[1] AI-powered report drafting offers a promising solution by reducing the time radiologists spend generating reports while potentially enhancing the accuracy and completeness of the report.[2,3]

CAUTION: Investigational device. Not for sale in the United States.

Historically, radiology AI has focused on narrow detection tasks, identifying individual abnormalities such as pulmonary embolism (PE), intracranial hemorrhage, or fractures. Vision-language models (VLMs) offer the potential to move beyond this narrow paradigm.[4] Rather than generating a single binary prediction, VLMs can analyze medical images and produce multiple natural-language observations describing all relevant imaging findings.[5]

Cognita Imaging, Inc.™ (Cognita), the advanced AI lab of Mosaic Clinical Technologies, Inc.™ (Mosaic), has developed VLMs to assist radiologists with interpretations (Figures 1 and 2). Capabilities currently include drafting for Chest Radiographs, Abdomen Radiographs and Musculoskeletal Radiographs, and computed tomography (CT) examinations (detailed in the accompanying CT research paper). The following report describes results from our research program, in which the VLMs are used under a research protocol, including performance data, user acceptance testing, and efficiency gains for radiograph drafting. A comprehensive lifecycle clinical AI governance process including continuous post-deployment monitoring has been established for radiograph drafting to ensure diligent oversight, mitigate risks and better understand the real-world effects of this new radiologist-AI collaborative technology.

Figure 1. Chest Radiograph VLM Output Example



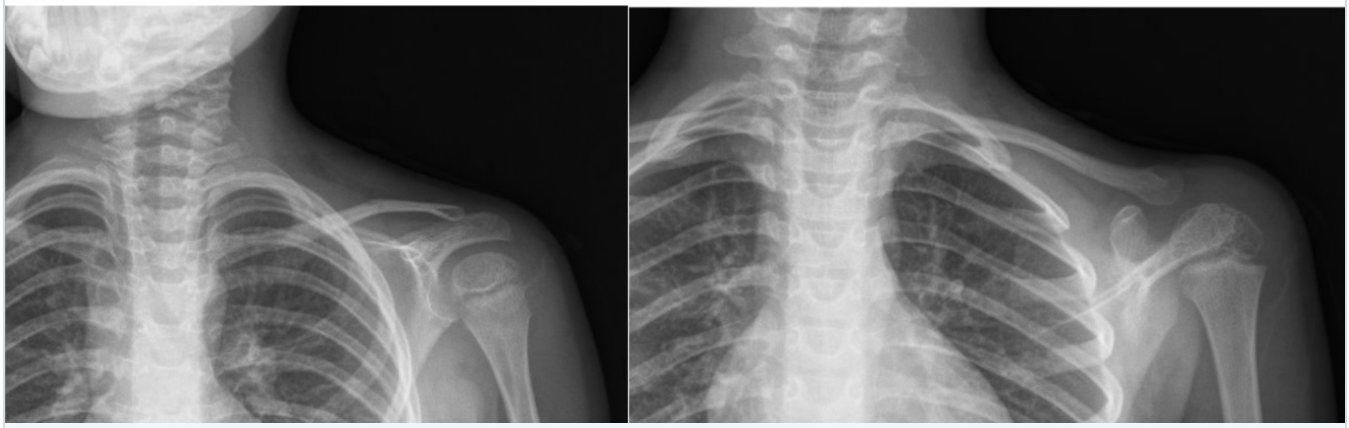
CLINICAL HISTORY

Respiratory failure

VLM OUTPUT

- **Endotracheal tube is well-positioned.**
- **Enteric tube enters stomach with tip not seen on this image.**
- **Right internal jugular central venous catheter terminates at the cavoatrial junction.**
- **Stable small bilateral pleural effusions.**
- **Mild pulmonary edema, similar.**
- **Stable hazy bibasilar lung opacities.**

Figure 2. **Musculoskeletal Radiograph VLM Output Example**



CLINICAL HISTORY

Acute left shoulder pain

VLM OUTPUT

- **Acute fracture of the midshaft of the left clavicle, without significant displacement or angulation.**

02 MODEL VALIDATION METHODS

Methods and Clinical Validation Framework for Responsible AI Implementation

To evaluate the potential clinical impact of report drafting, Mosaic developed a two-part pre-deployment clinical validation framework combining Performance Testing to measure diagnostic accuracy and opportunities for enhanced radiologist detection, and User Acceptance Testing to assess clinical usability and editing burden.

Performance Testing: Does the VLM Detect the Full Spectrum of Disease with High Accuracy and Augment the Radiologist?

Performance testing used large multicenter, geographically diverse retrospective datasets of radiograph examinations processed through the corresponding VLM on patients aged 1 year and older. AI-generated outputs and outputs from the findings sections of final radiologist reports from the retrospective dataset were classified for the presence or absence of pathology using internally validated large language model (LLM) classifiers.

Multiple experienced U.S. board-certified radiologists who included a mix of subspecialists and generalists independently and blindly reviewed a natural prevalence sample of radiograph examinations from a large retrospective dataset of examinations interpreted by radiologists in the Radiology Partners Clinical Services practice before AI drafting was in use. VLMs for Chest Radiograph, Abdomen Radiograph and Musculoskeletal Radiograph examinations were evaluated. The Musculoskeletal Radiograph VLM is inclusive of all extremity, pelvis and chest bone (rib, clavicle, sternum) radiographs, and excludes spine and above the neck (skull, sinus, orbit, mandible, etc.) radiographs. We identified instances of discordant cases, where the VLM and the radiologist report had opposite findings for specific high-priority pathologies chosen by Mosaic (specified for each exam type below). We conducted a human-in-the-loop manual review of a randomly selected subset of discordant cases from the full dataset to establish an expert reference standard, from which expert-adjudicated traditional model performance metrics, including prevalence, sensitivity and positive predictive value (PPV), were computed. For each examination review, the output generated by the pertinent VLM was presented alongside the output derived by an LLM from the findings section of the radiologist report, with the reviewer blinded to the identity of the two positive statement lists. We further computed the augmentation metric relative enhanced detection rate (rEDR),[6] which represents the number of enhanced detection cases due to the VLM divided by the number of true positive diagnoses made by the radiologist without access to AI.

A total of 166,198 radiographic examinations were included for performance analysis across Chest Radiograph, Abdomen Radiograph and Musculoskeletal Radiograph examinations. Among these radiographic examinations, 713 discordant exams were manually overread by the radiologist panel.

Performance Testing Cohort. Retrospective Radiograph Dataset, Patient Ages and Reviewer Panels

Exam Type	Cases Processed by VLM	Adult (Ages 18+)	Pediatric (Ages 1–17)	Cases Manually Reviewed	Reviewing Radiologists
Chest Radiograph	100,000	94.4%	5.6%	434	6
Musculoskeletal and Abdomen Radiograph	66,198	88.3%	11.7%	279	6
Total	**166,198**	—	—	**713**	—

Manually reviewed cases are discordant cases.

Reviewers are U.S. board-certified radiologists.

03 SUPPORTING CLINICAL EVIDENCE

CHEST RADIOGRAPH PERFORMANCE

A total of 100,000 cases were processed by the Chest Radiograph VLM (94.4% adult ages 18+; 5.6% pediatric ages 1-17), and there was a manual review of 434 discordant cases by a team of 6 U.S. board-certified radiologists. The Chest Radiograph VLM demonstrated the potential to enhance radiologist detection across all pathologies with relative enhanced detection rates ranging from 12.7% for lung opacity/consolidation to 57.5% for lung nodules/masses (Table 1). Of note there was significant enhanced detection of >50% despite the low prevalence of device malposition, non-lung masses, and pneumomediastinum. These potential improvements to disease detection are achieved while maintaining high positive predictive values (above 73% for 9 of the 10 conditions).

Table 1. Chest Radiograph Model Performance

Chest Radiograph Conditions	Prevalence	Sensitivity	PPV	rEDR
Device Malposition	1.0%	90.9%	92.3%	51.6%
Non-Lung Masses	0.6%	80.7%	97.8%	56.2%
Pneumoperitoneum	0.2%	97.2%	72.8%	35.8%
Bone Fracture/Malalignment	0.6%	85.7%	94.6%	46.9%
Pneumothorax	2.9%	96.8%	81.2%	46.5%
Pneumomediastinum	0.3%	96.4%	88.9%	50.2%
Pleural Effusion	21.1%	96.5%	89.5%	37.1%
Lung Opacity/Consolidation	26.1%	88.1%	98.4%	12.7%
Lung Nodules/Masses	4.5%	93.5%	100.0%	57.5%
Pulmonary Edema	13.7%	98.6%	73.6%	39.5%

MUSCULOSKELETAL AND ABDOMEN RADIOGRAPH PERFORMANCE

A total of 66,198 cases were processed by the Musculoskeletal and Abdomen Radiograph VLM (88.3% adult ages 18+; 11.7% pediatric ages 1-17) and there was a manual review of 279 discordant cases by a team of 6 U.S. board-certified radiologists. In Table 2, the Musculoskeletal and Abdomen Radiograph VLM demonstrated meaningful enhanced detection across all tested pathologies (rEDRs all ≥6.8%) with high PPV (all >75%) even for low prevalence conditions such as bone infection and suspicious bone lesions.

Table 2. Musculoskeletal and Abdomen Radiograph Model Performance

Musculoskeletal and Abdomen Radiograph Conditions	Prevalence	Sensitivity	PPV	rEDR
Bowel Obstruction (Abdomen)	13.6%	90.2%	90.2%	7.3%
Acute Fracture	16.1%	92.2%	96.3%	6.8%
Bone Infection	1.0%	78.7%	76.4%	27.1%
Bone Malalignment	4.5%	75.9%	95.9%	32.4%
Suspicious Bone Lesions	0.4%	87.4%	77.7%	45.8%

User Acceptance Testing: Is the Output Clinically Usable?

The same group of radiologist reviewers also assessed outputs along two clinically relevant dimensions at the case level: (1) clinical errors and (2) editing required. For Chest Radiographs, clinical errors were categorized as: absent; present without significant downstream clinical impact; or present with potentially significant downstream impact. For Musculoskeletal and Abdomen Radiographs, only clinical errors with potentially significant downstream impact were recorded. When assessing anticipated editing burden for Chest Radiographs, statements were categorized as "good" if they could be signed with no or minimal changes, "decent" if moderate editing was required, or "problematic" if starting from scratch would be preferable. For Musculoskeletal and Abdomen Radiographs, outputs were categorized by the number of edits required, either zero edits required or up to one edit required.

Table 3 demonstrates that the VLM output for all radiograph models has a clinically significant error rate comparable to or lower than that of the radiologist reports. For the Chest Radiograph VLM, a higher percentage of error-free reports was noted compared to radiologist reports. Moreover, the Chest Radiograph VLM resulted in a higher percentage of reports requiring no or minimal editing and a lower percentage of problematic reports compared to the human radiologist reports. For the Musculoskeletal and Abdomen Radiograph VLM outputs, a higher percentage of reports could be signed without change compared to radiologist reports.

Table 3. User Acceptance Testing Data: Report Quality and Clinically Significant Errors

Reader	"Good" Reports (No/Minor Edits)	"Decent" Reports (Moderate Edits)	"Problematic" Reports (Start from Scratch)	No Errors	Non-significant Errors	Clinically Significant Errors
CHEST RADIOGRAPH · N=159						
VLM	**89.3%** (83.4 – 93.7%)	**10.1%** (5.90 – 15.8%)	**0.6%** (0.02 – 3.45%)	**91.2%** (85.7 – 95.1%)	**8.2%** (4.43 – 13.6%)	**0.6%** (0.02 – 3.45%)
Radiologist	73.6% (66.0 – 80.3%)	22.0% (15.8 – 29.3%)	4.4% (1.79 – 8.86%)	83.0% (76.3 – 88.5%)	15.7% (10.4 – 22.3%)	1.3% (0.15 – 4.47%)

Reader	Report No Change Sign-Off	Report 0–1 Change Sign-Off	Clinically Significant Errors
MUSCULOSKELETAL AND ABDOMEN RADIOGRAPH · N=107			
VLM	**66.4%** (56.6 – 75.2%)	**88.8%** (81.2 – 94.1%)	**3.7%** (1.03 – 9.30%)
Radiologist	57.0% (47.1 – 66.5%)	90.7% (83.5 – 95.4%)	6.5% (2.67 – 13.0%)

95% Clopper-Pearson confidence intervals.

04 OPERATIONAL EFFICIENCY & WORKFLOW VALUE

PROSPECTIVE EFFICIENCY GAINS

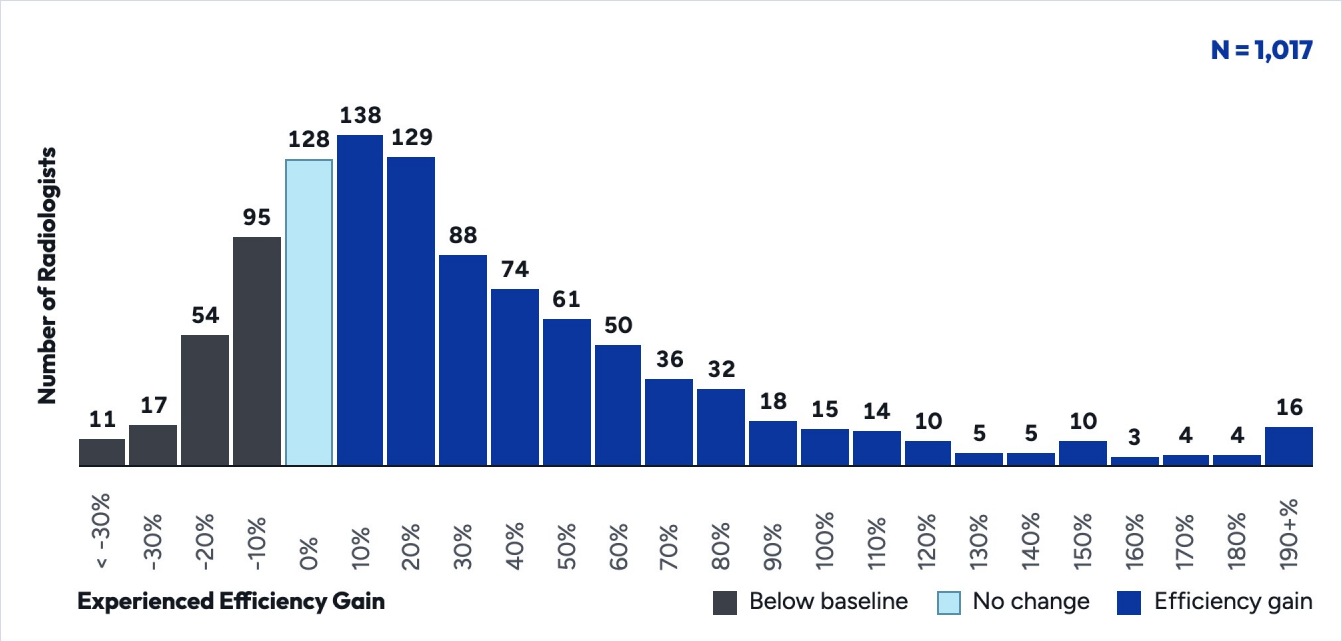
While the results above focus on retrospective evaluation, we also wanted to understand the real-world impact of radiologists working collaboratively with the VLMs. Table 4 shows the total number of examinations drafted for Chest Radiographs, Abdomen Radiographs and Musculoskeletal Radiographs through Oct. 22, 2026, with overall volume-weighted radiograph drafting efficiency gains[†] of 31% among participating radiologists in the Radiology Partners Clinical Services practice. Figure 3 demonstrates significant efficiency gains for approximately 77% of radiologists using drafting capabilities for both radiographic and CT examinations.

Table 4. Efficiency Impact of Drafting by Exam Type

Exam Type	Drafted Exams (All Time)	Efficiency Gain (Radiologists with >50 Drafted Exams)
Chest Radiograph	3.42M	+28%
Abdomen Radiograph	0.36M	+29%
Musculoskeletal Radiograph	2.55M	+36%

[†]Source: Radiology Partners Clinical Transformation data. Efficiency was measured as the average time-based work units (TBWUs) completed per minute of reading time among radiologists who had read at least 50 drafted exams, compared with each radiologist's own performance before drafting was available.

Figure 3. **Distribution of Mosaic Efficiency Impact, including drafted radiographs and CTs.**



N = 1,017 radiologists. Each bar shows a +/-5% range around its label.

Baseline of the same radiologists reading before the implementation of the drafting solution.

05 CLINICAL AI GOVERNANCE

A comprehensive lifecycle clinical AI governance process is in place for radiograph drafting (Figure 4), which includes implementation of continuous monitoring of radiologist agreement rates with VLM output and report signing times to capture data on radiologist-AI interaction. This allows us to evaluate the real-world distribution of radiologist-AI behavior to assess model failure modes, automation bias, and outlier behavior.

Our AI continuous monitoring dashboard (Figure 5) tracks two AI-radiologist interaction metrics for each tracked high-priority pathology on drafted radiographs. The Positive Concordance rate represents the percentage of AI positive drafted cases for a specific finding that are also confirmed to be positive in the final signed radiologist reports and is a proxy for AI precision. The Negative Discordance rate represents the percentage of AI negative drafted cases for a specific finding that are found to be positive in the final signed radiologist reports and is a proxy for the AI miss rate.

Figure 4. Clinical AI Governance Process

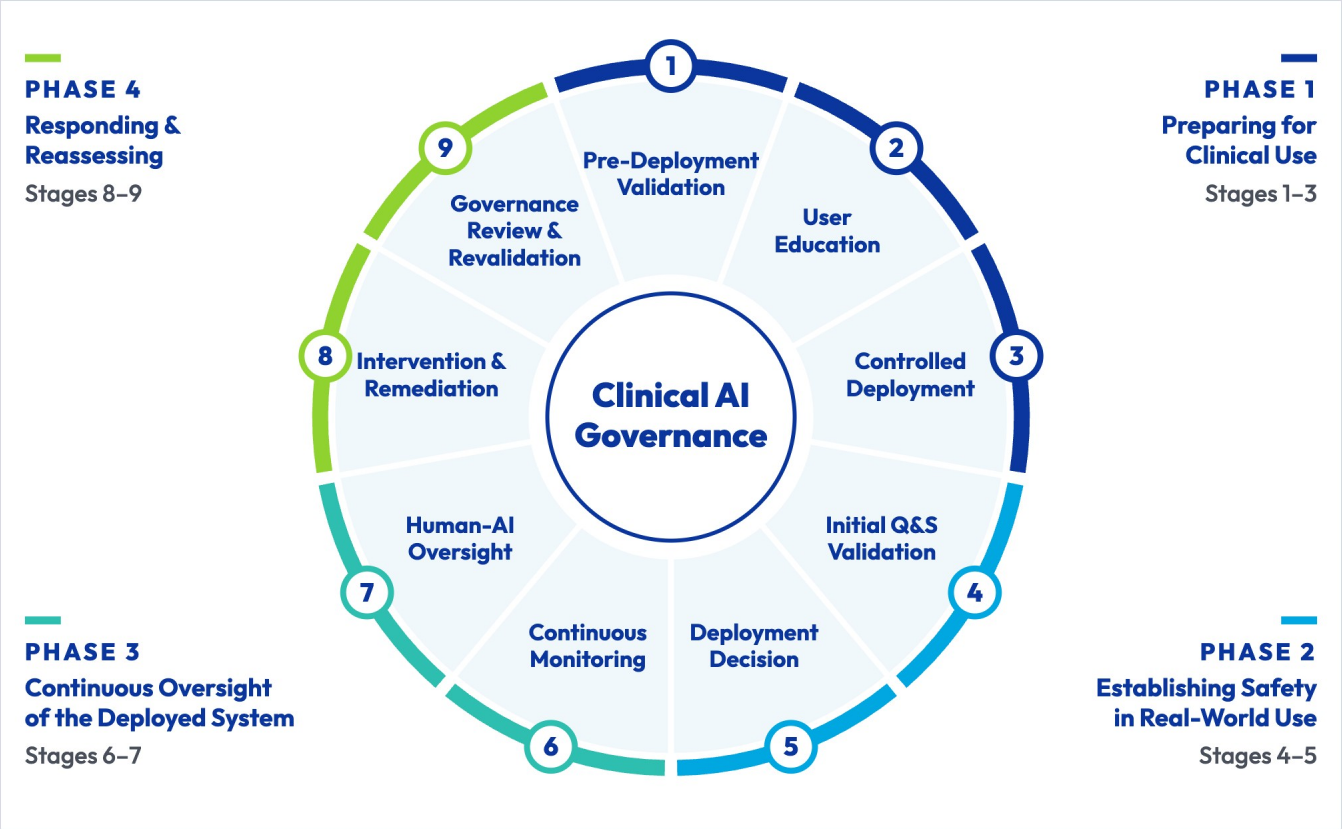
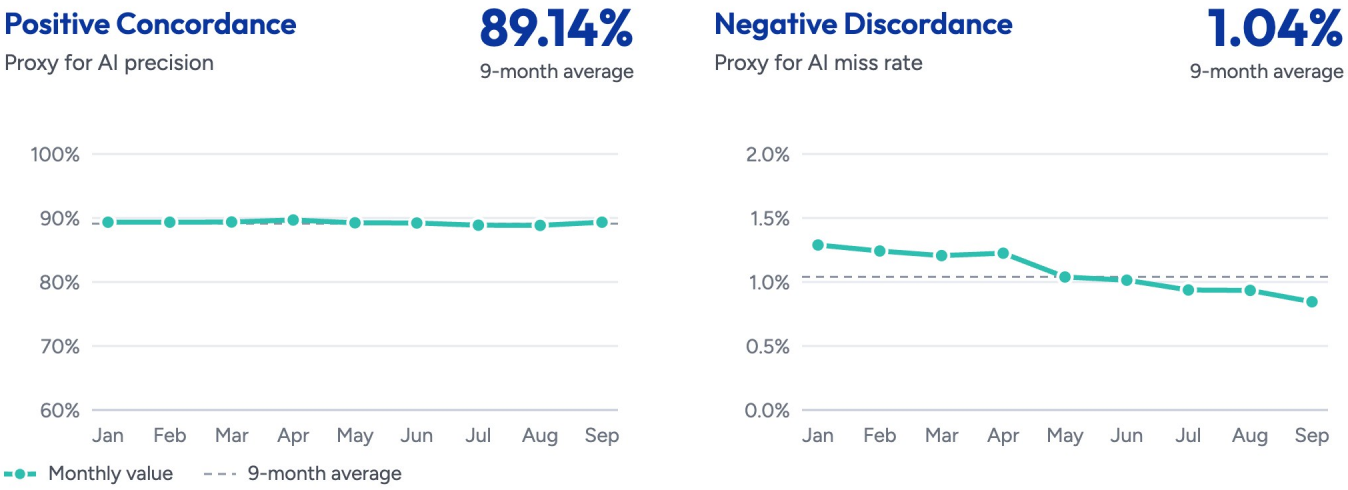


Figure 5. Drafted Radiographs Continuous Monitoring Dashboard, January 1 – September 30, 2026.



Aggregated post-deployment monitoring data from 15 tracked high-priority pathologies across Chest Radiographs, Abdomen Radiographs and Musculoskeletal Radiographs (see Tables 1 and 2 for the pathology list). Average Positive Concordance rate was 89.14%. Average Negative Discordance rate was 1.04%.

06 CONCLUSIONS

Comprehensive VLMs represent a significant advancement in radiology AI, moving beyond isolated detection tasks toward clinically useful, examination-level assistance. Mosaic's ongoing research validation results demonstrate strong real-world performance across Chest Radiographs, Abdomen Radiographs and Musculoskeletal Radiographs, supporting the potential for VLM-powered report drafting to enhance radiologist diagnostic performance, efficiency, and clinical workflow.

The opportunity is not simply better AI, but better radiology through human-AI collaboration. With encouraging clinical validation results and growing pressure on the radiology workforce, the time to begin thoughtfully integrating these capabilities into clinical practice is now. Mosaic is actively pursuing FDA regulatory authorization while continuing rigorous validation, real-world performance monitoring, and model refinement to support safe, scalable deployment.

07 REFERENCES

[1] Jing AB, Garg N, Zhang J, Brown JJ. AI solutions to the radiology workforce shortage. *npj Health Syst*. 2025;2(1):20. doi:10.1038/s44401-025-00023-6

[2] Mamdouh D, Attia M, Osama M, et al. Advancements in radiology report generation: a comprehensive analysis. *Bioengineering (Basel)*. 2025;12(7):693. doi:10.3390/bioengineering12070693

[3] Seah JCY, Tang JSN, Tran A. Drafting the future: the dawn of AI report generation in radiology. *Radiology*. 2025;316(1):e243378. doi:10.1148/radiol.243378

[4] Blankemeier L, Kumar A, Cohen JP, et al. Merlin: a computed tomography vision–language foundation model and dataset. *Nature*. 2026;652(8112):1318-1328. doi:10.1038/s41586-026-10181-8

[5] Uppuluri PC, Yang CW. The emergence of foundation models in U.S. radiology: a narrative review of clinical utility, safety, and evaluation. *Acad Radiol*. 2026;33(3):797-814. doi:10.1016/j.acra.2025.10.063

[6] Larson DB, Poff JA, Krishnan S, et al. Predicting the value of radiology artificial intelligence applications: large-scale predeployment evaluation of a portfolio of models. *AJR Am J Roentgenol*. Published online 2026. doi:10.2214/AJR.25.34340

ABOUT MOSAIC CLINICAL TECHNOLOGIES, INC.

Mosaic Clinical Technologies, Inc., a wholly owned subsidiary of Radiology Partners, Inc., develops advanced technology and proprietary vision-language AI models for medical imaging, built in-house in collaboration with radiologists in the Radiology Partners Clinical Services practice. Its MosaicOS platform brings diagnostic tools, cutting-edge AI and radiology workflows together in a cloud-native operating system designed to support radiologists and the healthcare organizations they serve. Learn more at MosaicClinical.ai and connect with us on LinkedIn and YouTube.

ABOUT COGNITA IMAGING, INC.

Cognita Imaging, Inc., Mosaic Clinical Technologies, Inc.'s advanced AI lab, develops proprietary vision-language foundation models and other generative AI capabilities for radiology. Its work is focused on helping radiologists deliver more accurate, efficient care and expanding access to high-quality healthcare. Cognita seeks to elevate the performance of all radiologists to the level of the best expert subspecialists and, longer term, to empower radiologists with digital imaging information not easily perceptible to the human eye. Learn more at cognita.ai.

RSNA 2026 EVENT HUB

Download both research papers and the results summary: mosaicclinical.ai/rsna-event-hub
Contact Mosaic: mosaicclinical.ai/contact

REGULATORY COMPLIANCE NOTE

CAUTION: Investigational device. Not for sale in the United States.