Technological Innovation & Considerations for Post Market Performance Monitoring in Radiology

Nina Kottler, MD, MS, FSIIM
Associate CMO, Clinical AI @Radiology Partners
Associate Fellow @Stanford AIMI Center

Nov 21, 2024

radiology partners™

Disclosures

- Consultant for ES3, Inc (aerospace)
- Consultant for Synapsica Healthcare
- Partner (equity owner) at Radiology Partners (RP)
- Sole or partial owner of several radiology practices managed by RP
- RP has a minority interest in aidoc
- RP has an indirect minority interest in Rad AI
- Speaking honorariums from Philips and Bayer for RSNA 2024
- Associate Fellow Stanford AIMI Center
- Hold several volunteer positions at RSNA, SIIM, and RadEqual
- Volunteer with Sai Aashraya, Radiology Division

radiology partners™

How Is GenAI Used In Radiology Today?

1

text

summarize
translate

text

2

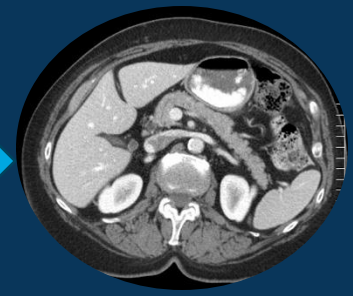
text

3

text

text

4



radiology partners™

continuous Validation

LOADING...

Agenda

1. Monitoring a GenAI LLM In Radiology Practice
   - Types of Errors
   - Error Mitigation
   - Required Oversight by Qualified End User
2. A Framework for Continuous Monitoring

radiology partners™

GenAI Reporting

Report Preamble

- Exam
- History
- Comparison
- Technique

FINDINGS:

Description of exam findings

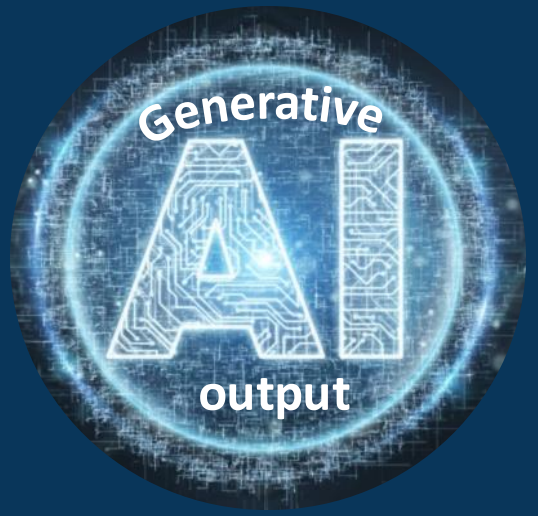
IMPRESSION:

Edited summarization

radiology partners™

GenAI Reporting Validation Process

- Random selection of 3,000 radiology reports from across scaled practice
- Compared GenAI output to the Final Radiology Report using LLM



VS

Radiologist edited output

- Results reviewed by panel of radiologists to validate & determine significance

radiology partners™

GenAI Validation Results

GenAI Model	GenAI Clinically Significant Error Rate	Rad Edited Clinically Significant Error Rate
Impression Generation	4.8%	

radiology partners™

GenAI Reporting Pitfalls

	GenAI Error Type	Error Rate
1	Numerosity	0.2%
2	Laterality / Body Site	0.6%
3	Recommendation	1.0%
4	Missed Finding	1.3%
5	Other Hallucination	1.7%

GenAI Error – Numerosity

	GenAI Error Type
1	Numerosity
2	Laterality / Body Site
3	Recommendation
4	Missed Finding
5	Other Hallucination

FINDINGS:	GenAI IMPRESSION:	Final IMPRESSION:
CAROTID ARTERIES: Calcified plaque causes 70% stenosis of the proximal right internal carotid artery. Calcified plaque causes 30% stenosis of the proximal left internal carotid artery. Calcified plaque causes 40% stenosis of the origin of the left common carotid artery.	70% stenosis of the proximal right internal carotid artery. 40% stenosis of the proximal left internal carotid artery. 40% stenosis of the origin of the left common carotid artery.	70% stenosis of the proximal right internal carotid artery. 30% stenosis of the proximal left internal carotid artery. 40% stenosis of the origin of the left common carotid artery.

radiology partners™

GenAI Error – Numerosity

	GenAI Error Type
1	Numerosity
2	Laterality / Body Site
3	Recommendation
4	Missed Finding
5	Other Hallucination

FINDINGS:

Irregular perihilar 5.8 x 3.6 cm right lung mass is decreased from 10/25/2022 chest CT.

Left lower lobe solid 0.8 cm pulmonary nodule is stable back to 03/14/2019 chest CT.

GenAI IMPRESSION:

Perihilar right lung mass is decreased from 10/25/2019 chest CT.

Final IMPRESSION:

Perihilar right lung mass is decreased from 10/25/2022 chest CT.

radiology partners™

GenAI Error – Laterality ("Fill in the Blank")

	GenAI Error Type
1	Numerosity
2	Laterality / Body Site
3	Recommendation
4	Missed Finding
5	Other Hallucination

Initial FINDINGS:	GenAI IMPRESSION:	Final IMPRESSION:	Final FINDINGS:
body site, stability, severity, size, clinical indication			
underlying minimal cortical irregularity worrisome for osteomyelitis.	cortical irregularity worrisome for osteomyelitis.	minimal cortical irregularity worrisome for osteomyelitis.	There is new underlying minimal cortical irregularity worrisome for osteomyelitis.

radiology partners™

GenAI Error – Recommendation

	GenAI Error Type
1	Numerosity
2	Laterality / Body Site
3	Recommendation
4	Missed Finding
5	Other Hallucination

INITIAL FINDINGS:

Within the proximal posterior tibia there is a cortically based nonaggressive appearing lucent lesion with a well-defined sclerotic rim measuring up to 2.2 cm.

GEN AI IMPRESSION:

Nonaggressive appearing lucent lesion within the proximal posterior tibia. Consider further evaluation with MRI.

FINAL IMPRESSION:

Nonaggressive appearing cortically based lesion with a well-defined sclerotic rim in the posterior tibia. This is favored to represent a nonossifying fibroma.

radiology partners™

GenAI Error – Missed Finding

	GenAI Error Type
1	Numerosity
2	Laterality / Body Site
3	Recommendation
4	Missed Finding
5	Other Hallucination

FINDINGS INITIAL:

There is moderate bilateral hydronephrosis and hydroureter without obvious obstructing calculus noted.

Bladder is markedly distended.

IMPRESSION INITIAL:

Moderate bilateral hydronephrosis and hydroureter without obvious obstructing calculus.

??

IMPRESSION FINAL:

Moderate bilateral hydronephrosis and hydroureter without obvious obstructing calculus.

Markedly distended bladder.

radiology partners™

GenAI Error – Hallucination ("Nearest Neighbor")

	GenAI Error Type
1	Numerosity
2	Laterality / Body Site
3	Recommendation
4	Missed Finding
5	Other Hallucination

CT Abdomen and Pelvis. Indication: 55-year-old male patient with pelvic pain

Initial FINDINGS:	GenAI IMPRESSION:	Final IMPRESSION:
Small amount of free fluid in the pelvis.	Small amount of free fluid in the pelvis, possibly due to recent ovarian cyst rupture.	Small amount of free fluid in the pelvis, nonspecific.

radiology partners™

GenAI Error – Hallucination (“Contradiction”)

	GenAI Error Type
1	Numerosity
2	Laterality / Body Site
3	Recommendation
4	Missed Finding
5	Other Hallucination

FINDINGS:

Water-soluble contrast was slowly infused into the colon via gravity opacifying the rectum, and descending colon. There was no contrast extravasation to suggest a colovaginal fistula.

GenAI IMPRESSION:

No evidence of a colovaginal fistula.

A contrast enema was not performed due to patient discomfort.

Final IMPRESSION:

No evidence of a colovaginal/rectovaginal fistula.

The patient was transferred to CT for further evaluation.

radiology partners™

GenAI Error – Hallucination (”Redundancy”)

	GenAI Error Type
1	Numerosity
2	Laterality / Body Site
3	Recommendation
4	Missed Finding
5	Other Hallucination

FINDINGS INITIAL:

LUNG BASES: There is a 4 mm ground-glass nodule in the right lower lobe (axial image 20). 6 mm ground-glass nodule noted in the left lower lobe (axial image 2).

IMPRESSION INITIAL:

4 mm right lower lobe and 6 mm left lower lobe ground-glass pulmonary nodules.

6 mm left ground-glass pulmonary nodule.

IMPRESSION FINAL:

4 mm right lower lobe and 6 mm left lower lobe ground-glass pulmonary nodules. Recommend follow-up to resolution.

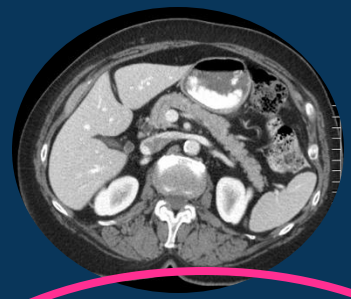
radiology partners™

Agenda

1. Monitoring a GenAI LLM In Radiology Practice
   - Types of Errors
   - Error Mitigation
   - Required Oversight by Qualified End User
2. A Framework for Continuous Monitoring

radiology partners™

Continuous Evaluation of Model Performance



data

3 ensembling

AI results

concordant
discordant

Rad report

4 In or out of distribution to training data

document

+/- clinical adjudication

1,2 Automated or expert in-the-loop monitoring

radiology partners™

GenAI Monitoring Strategies

	Strategy	”Ground Truth” Comparison
1	Automated Monitoring	Radiology Report
2	Expert-in-the-loop Monitoring	Radiology Consensus
3	Ensemble Monitoring	Other Gen AI Models
4	IN vs. OOD Data Monitoring	GenAI Training Data
5	Model Stress Testing	Ground Truth Output
6	User Feedback	“Radiology Report”

GenAI Monitoring Strategies

	Strategy	Strengths	Weaknesses
1	Automated Monitoring	Scalable, Continuous	Inherent error rate, delayed results for low volume/ prevalence sites
2	Expert-in-t Monitoring		xpert uous)
3	Ensemble M		without re- cal difficulty
4	IN vs. OOD Monitoring		ith accuracy; ta
5	Model Stress Testing	can test low prevalence findings	y test what is e to data drift unless refresh ground truth; discontinuous
6	Expert Feedback	Fast	Sampling bias

Can choose or combine approaches based on model and model risk

Take Home Points

	GenAI Error Type
1	Numerosity
2	Laterality / Body Site
3	Recommendation
4	Missed Finding
5	Other Hallucination

- GenAI is being used today in healthcare and should be monitored post deployment for accuracy (error patterns)
- There are several ways this can be done at scale
- Our validation suggests autonomous GenAI reporting on its own is probably not safe, but expert in-the-loop GenAI reporting is likely a *net benefit for patient safety*
- Users benefit from education on the specific error patterns associated with GenAI reporting

radiology partners™

Thank You!

@RadKottler
nina.kottler@radpartners.com
Radiology Partners
El Segundo, CA

radiology partners™

Questions