Algorithmic Drift and Lifecycle Quality Assurance in Radiology AI: A Framework for Ongoing Performance Monitoring and Re-Validation

Colin Wu[1], Edward Pettyjohn, MD[2], Youyou Cheng[1], Ravi Patel[1], Sina Ashraf[1], Justin Kuyn[3], Shreyas Meruga[4], Zhuangwei Kang, PhD[5], Sarah Pettyjohn, MD[6]

[1] University of the Incarnate Word School of Osteopathic Medicine; [2] University of Illinois College of Medicine Peoria (UICOMP); [3] University of Colorado Boulder; [4] UT Health San Antonio Department of Molecular Medicine; [5] Independent Researcher; [6] Rutgers Robert Wood Johnson Medical School

Introduction

AI tools are increasingly used in radiology, yet most institutions treat them as fixed systems. In reality, model accuracy and calibration decline as patient populations change, scanners are upgraded, protocols shift, disease prevalence evolves, or vendors introduce silent updates. This drift can affect patients through missed findings, unnecessary follow-up, inconsistent triage, and widening performance gaps across demographic groups. Unlike imaging hardware, clinical AI does not undergo routine quality assurance. This project developed a radiology-specific framework for continuous monitoring, drift detection, and scheduled re-validation to protect patient safety and diagnostic reliability.

Methods

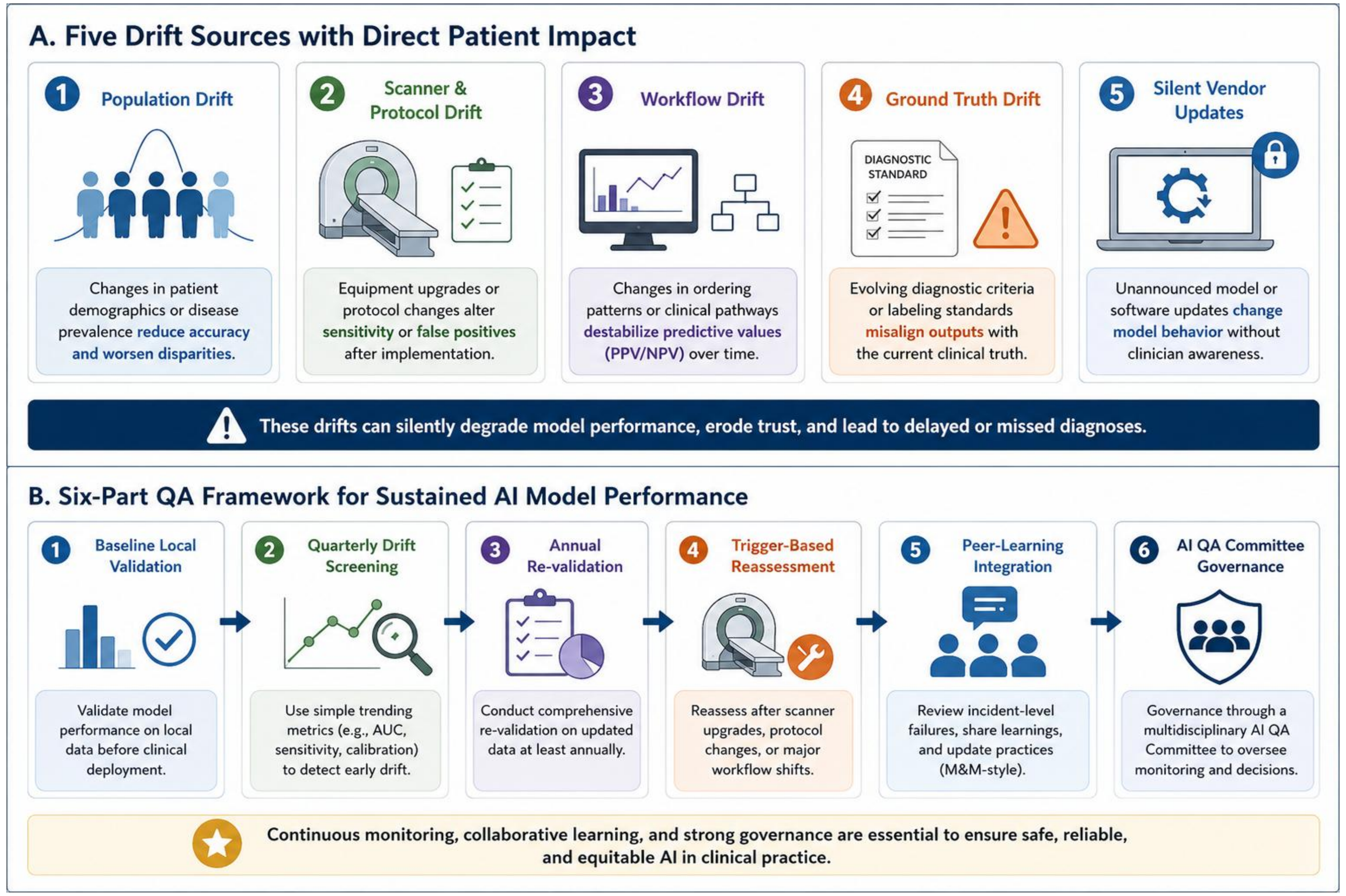
We synthesized evidence on drift mechanisms, radiology QA practices, silent vendor update behavior, human factors risks, and real-world performance degradation in breast CAD, lung nodule triage, hemorrhage detection, and TACE prediction. These inputs informed a practical Algorithmic Lifecycle QA cycle modeled on hardware QA standards

Results

We identified five drift sources with direct patient impact:

1. Population drift reducing accuracy and worsening disparities.
2. Scanner and protocol drift altering sensitivity or false positives after equipment changes.
3. Workflow drift destabilizing predictive values as ordering patterns evolve.
4. Ground truth drift misaligning outputs with new diagnostic standards.
5. Silent vendor updates changing model behavior without clinician awareness.

We propose a six-part QA Framework: baseline local validation, quarterly drift screening using simple trending metrics, annual re-validation, trigger-based reassessment after scanner or protocol changes, peer-learning integration for incident-level failures, and governance through an AI QA Committee.



Conclusion

Algorithmic drift is unavoidable and can compromise patient care if unmonitored. A structured lifecycle QA program ensures that clinical AI remains safe, calibrated, and reliable throughout real-world use.

References

•Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D.
Key challenges for delivering clinical impact artificial intelligence.
BMC Med. 2019;17(1):195.
doi:10.1186/s129with16-019-1426-2

•Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK.
Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs.
PLoS Med. 2018;15(11):e1002683.
doi:10.1371/journal.pmed.1002683

•Wiens J, Saria S, Sendak M, et al.
Do no harm: a roadmap for responsible machine learning for health care.
Nat Med. 2019;25(9):1337-1340.
doi:10.1038/s41591-019-0548-6

•Shimoda W, Kishida Y, Kiyuna T, et al.
Monitoring and maintenance of deep learning models in clinical practice.
Jpn J Radiol. 2021;39(11):1055-1062.
doi:10.1007/s11604-021-01145-8

•U.S. Food and Drug Administration.
Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan.
Published January 2021.