Reprinted from *Canadian Healthcare Technology,* April 2026 issue

CANADIAN Healthcare Technology
www.canhealth.com

Measuring AI performance in radiology using the AI quality framework

BY JEFF VACHON

Across academic and enterprise health systems, AI tools are now routinely embedded within imaging workflows, supporting triage, image detection, quantification, reporting, and operational coordination. The more pressing challenge today is not adoption but understanding how well these systems perform once they become part of daily clinical reality.

As AI assumes a more central role in radiologic practice, it increasingly influences diagnostic decisions, workflow and clinical management. In this environment, traditional approaches to AI evaluation, often based on static, retrospective accuracy metrics, are no longer sufficient.

Radiology now requires a framework that can continuously assess how intelligently AI systems operate within real-world clinical environments. This need has given rise to the concept of the AI Quality Framework (AIQ): an evidence-driven approach to measuring AI performance as it is experienced by radiologists, departments, and patients.

From point solutions to embedded intelligence: Early generations of radiology AI were characterized by narrowly focused algorithms designed to address specific diagnostic tasks. These tools often existed outside the core imaging workflow and were accessed selectively, requiring additional interfaces or manual steps. While clinically promising, their impact was inherently constrained.

Today, AI has migrated into the foundational infrastructure of radiology. Algorithms are increasingly embedded directly into AI platforms, reporting systems and enterprise workflow engines.

Rather than acting as optional add-ons, AI systems now operate continuously in the background – prioritizing studies, generating structured findings, surfacing quantitative measurements and supporting operational decision-making.

Jeff Vachon explains the benefits of Bialogics and the AI quality framework at last fall's RSNA conference in Chicago. The AI-driven system can assess the performance of other AI applications.

In many departments, AI has effectively become part of the digital backbone of imaging services.

As AI becomes pervasive, radiology must move beyond asking whether an algorithm is accurate in isolation and instead evaluate how intelligently it performs within the broader context of a complex diagnostic system guided by human expertise.

Why traditional AI metrics are no longer sufficient: Most AI systems enter clinical use on the strength of validation studies conducted in controlled research environments. These studies typically report sensitivity, specificity, and related statistical accuracy measures derived from curated datasets with known ground truth.

While essential for regulatory clearance and early assessment, these metrics represent only potential performance.

Clinical radiology, however, is dynamic. Image quality varies, disease prevalence shifts, biases become transparent, patient populations differ across jurisdictions and over time.

Workflow integration influences how AI outputs are interpreted and acted upon.

Human judgment remains central, and radiologist interpretation itself introduces variability. Under these conditions, AI performance cannot be assumed to remain static after deployment.

An algorithm that performs well during development may drift, degrade, or behave unpredictably once exposed to real-world complexity. Static accuracy metrics cannot capture these changes.

Intelligence, by contrast, implies consistency, adaptability, and contextual awareness. Measuring AI intelligence therefore requires continuous observation within clinical practice rather than episodic retrospective review.

Defining the AI Quality Framework: The AI Quality Framework (AIQ) reframes how radiology evaluates AI performance. Rather than focusing solely on technical accuracy, AIQ assesses how consistently AI outputs align with radiologist-reported findings, how reliably the system avoids false positives and false negatives, and how much incremental value it adds to clinical interpretation.

AIQ is derived from a composite of clinically meaningful performance indicators calculated directly from routine imaging workflows. By comparing AI-generated findings with finalized radiology reports on a case-by-case basis, AIQ reflects how AI behaves in the same clinical context as the radiologist. Importantly, the framework allows institutions to weight these indicators according to clinical priorities, recognizing that intelligence may be defined differently for screening, triage, or diagnostic augmentation.

In this way, AIQ functions not as a single abstract score, but as a structured measure of real-world clinical intelligence.

Measuring operational impact as a component of AI intelligence: As AI becomes embedded in daily workflows, its value must also be assessed through its impact on radiology performance and productivity. Within AIQ, operational metrics are intentionally evaluated alongside diagnostic KPIs to provide a unified view of how AI influences both patient care and departmental function.

Radiologist read-time is one of the most direct indicators of AI's operational impact. AI systems that automatically identify findings, generate measurements, or populate structured report elements are designed to reduce cognitive and manual workload.

Within AIQ, changes in read time are interpreted in parallel with concordance-based KPIs such as overall concordance, positive and negative concordance rate. Consistent reductions in read times with stable or improving AIQ metrics indicate that AI is effectively supporting clinician decision-making, while drops in diagnostic alignment may point to unsafe speed-ups or workflow issues.

Turnaround time offers a broader view of AI's influence on care delivery. AI-driven triage and prioritization tools are intended to surface critical findings earlier in the workflow, particularly in emergency and high-acuity settings.

Improvements in turnaround time that coincide with stable AIQ scores indicate successful workflow integration. Conversely, gains in speed accompanied by declining negative concordance or increasing false-positive alerts may reflect prioritization inefficiencies that warrant governance review.

Clinical radiology is dynamic, and image quality varies, disease prevalence shifts, and biases become transparent.

At the departmental level, throughput metrics such as studies interpreted per shift or relative value units (wRVU) generated provide insight into AI's effect on capacity management.

Within an AIQ-aligned model, throughput is never evaluated in isolation. Increases in volume are assessed alongside AIQ trends, drift indicators, and radiologist interaction metrics to ensure that productivity gains do not come at the expense of diagnostic intelligence or clinician trust.

Augmented diagnostic capability and clinical value: Beyond efficiency, AI's most meaningful contribution may be its ability to enhance diagnostic capability. AIQ captures this dimension through measures such as augmented findings rate, which quantify AI's incremental contribution relative to baseline interpretation. These metrics provide evidence that AI is not merely accelerating workflows but actively improving diagnostic insight.

The need for continuous monitoring: A defining characteristic of radiology is that it evolves over time. AI systems are no exception. Changes in patient populations, imaging protocols, scanner technology, or clinical workflows can subtly influence performance long after deployment. Without continuous monitoring, these changes may remain undetected until they manifest as clinical risk.

AIQ addresses this challenge by enabling longitudinal tracking of AI performance across diagnostic and operational dimensions. Trends in AIQ scores, concordance metrics, and productivity indicators provide early warning of drift or instability.

This allows radiology departments to intervene proactively – adjusting workflows, retraining models, or restricting use when necessary – rather than reacting to failures retrospectively.

Implications for radiologists and radiology leadership: As AI becomes more deeply embedded in practice, the role of the radiologist is evolving from sole generator of findings to supervisor of AI-augmented interpretation. For this model to succeed, trust must be supported by evidence. AIQ provides that evidence by making AI behavior transparent and measurable.

For radiology executives, AIQ offers a strategic governance tool. It supports informed decisions about AI procurement, deployment, scaling, and retirement, while providing defensible documentation for quality oversight and regulatory readiness. By integrating clinical and operational intelligence into a single framework, AIQ enables leadership to manage AI as a core component of radiology performance rather than as an isolated technology initiative.

AIQ as the foundation of intelligent radiology: Radiology is no longer questioning whether AI works, but how intelligently it performs in real-world clinical practice. AIQ offers a framework to ensure AI remains trustworthy, clinically aligned, and beneficial to patient care as AI diagnostics become the norm.

Jeff Vachon is the president of Bialogics.