bialogics

Innovation Spotlight | Radiology AI

Bialogics AI Quality (AIQ) Framework - Evidence-Driven, Continuous, AI Quality Assessment in Radiology

Setting a New Standard for Assessment of AI In Diagnostic Imaging

The AI Quality (AIQ) Framework - Evidence-Driven, Continuous, AI Quality Assessment in Radiology

SECTION 1: AI QUALITY ASSESSMENT OVERVIEW

Introduction: The integration of artificial intelligence (AI) into radiologic practice has the potential to enhance diagnostic accuracy and operational efficiency. However, its widespread clinical adoption is constrained by the absence of standardized, evidence-based quality assurance mechanisms. This paper introduces Bialogics' AI Quality (AIQ) framework—a continuous assessment of AI performance across key diagnostic and operational performance indicators. By providing the objective measurement of AI algorithm performance, AIQ enables institutions to evaluate the clinical impact, track operational benefits, and foster confidence while deploying AI technologies in radiology practice.

Challenges in Present-Day AI Performance Assessment: AI performance is largely characterized by manual, retrospective review where radiologists compare AI findings to their own interpretations and the clinical history of the patient. Early adopter institutions are experimenting with natural language processing (NLP) to automate performance assessment by extracting diagnoses from radiology reports as ground truth and comparing these with AI outputs to measure concordance retrospectively. Despite these advances, clinical integration of AI remains hindered by the absence of commercially available, evidence-based validation platforms.

Scientific Validation vs. Real-World Clinical Performance: Scientific AI accuracy analysis is typically performed in controlled research environments where algorithms are tested against curated, high-quality datasets with known ground truth. These studies measure performance using statistical metrics such as accuracy, sensitivity, or predictive value to validate algorithm reliability under ideal conditions. In contrast, real-time clinical practice introduces dynamic variables—workflow integration, image quality variability, patient complexity, and human interpretation—that influence AI performance in daily use. Measuring accuracy in this environment requires continuous monitoring of Positive and Negative Compliance Rates**,** which compare AI-generated findings with radiologist-reported outcomes to determine how consistently the algorithm aligns with real-world clinical interpretations and decision-making.

Bialogics' AIQ Framework: AIQ integrates directly into radiology workflows, supporting the clinically validated, transparent, and scalable deployment of AI solutions. AIQ solves two critical problems plaguing AI in radiology today; 1) the lack of a standardized, quantitative process for evaluating the performance/accuracy of AI algorithms individually and collectively and 2) tracking

and reporting on the radiology workflow benefits due to the use of AI. AIQ continuously analyzes the clinical findings of each AI algorithm with the clinical findings in each diagnostic report.

By analyzing the clinical findings of AI versus the clinical findings in the diagnostic report for an exam and by collecting and analyzing all the operational and clinical data related to that exam via HL7 and DICOM, AIQ calculates key performance indicators (KPIs), including Concordance, Positive Concordance Rate, Negative Concordance Rate, Positive Predictive Value (PPV), Negative Predictive Value (NPV), Augmented Findings Rate (AFR), and Workflow Impact. AIQ also computes the AIQ Score which is an algorithmic calculation based on a weighted sum of the KPIs.

AIQ Key Performance Indicators: The AIQ KPIs are calculated based on four key AI accuracy measurements: True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN). These values are determined based on comparing the AI findings and the radiologist's findings in the Diagnostic Report.

	AI Findings (+)	AI Findings (-)
Diagnostic Report (+)	TP	FN
Diagnostic Report (-)	FP	TN

AIQ reports eight core KPI metrics which will help users track the current and ongoing performance of their AI systems.

Key Performance Indicators (KPIs) for AIQ		
Concordance	Agreement between AI findings and radiologist-reported diagnoses.	TP+TN /Total Exams
Positive Concordance Rate (PCR)	Measures the degree of agreement between the AI's positive findings and the radiologist's confirmed positive findings in the diagnostic report or ground truth dataset.	TP / (TP + FN)
Negative Concordance Rate (NCR)	Measures the degree of agreement between the AI and the radiologist in identifying cases or regions with no abnormal findings	TN / (TN + FP)
Positive Predictive Value (PPV)	The proportion of AI-detected positives that are concordant with all positive findings	TP / (TP + FP)
Negative Predicted Value (NPV)	The proportion of AI-detected negatives that are concordant with negative findings	TN / (TN + FN)
Augmented Findings Rate (AFR)	AI's capability to identify and reduce false positives and false negatives	Positive Finding Rate before AI vs. Positive Finding Rate after AI

AIQ Score: The 6KPIs listed above each measure a specific use case of AI performance and each KPI is relevant for specific users of AI. What doesn't currently exist though, is a single measurement, a score, that allows users of AI algorithms to define how they want to measure AI success. The AIQ score is calculated using a weighted sum of the normalized KPI values, where each weight is user-defined and reflective of the clinical priorities for a given use case.

The AIQ Score Formula – Example: AIQ = (25% * Concordance) + (15% * PCR) + (15% * NCR) + (10% * PPV) + (10% * NPV) + (15% * AFR)

The AIQ Score offers a structured, quantitative analysis of an AI model's clinical performance. Each metric—such as concordance, PCR, NCR, etc. - is weighted according to its clinical significance, then normalized and multiplied to calculate its individual contribution. In this example, the model achieves a high AIQ score of **90.9 out of 100**, signalling strong diagnostic alignment with radiologist findings and low error deviation.

Metric	Weight (%)	AI Score	Normalized	Weighted Contribution	AI Q Score Contribution
Concordance	**25%**	**92%**	**0.92**	**0.2300**	✅ **High**
PCR	**20%**	**94%**	**0.94**	**0.1880**	✅ **High**
NCR	**20%**	**89%**	**0.89**	**0.1780**	✅ **Medium**
PPV	**10%**	**92%**	**0.92**	**0.00920**	✅ **High**
NPV	**10%**	**97%**	**0.97**	**0.0970**	✅ **High**
AFR	**15%**	**80%**	**0.80**	**0.12**	⚠️ **Slight Deviation**
AIQ Score	100%			=	**90.5%**

By integrating multiple performance dimensions into a single, interpretable score, the AIQ Score allows clinicians and administrators to objectively evaluate, compare, and monitor AI systems with confidence, making it a crucial tool for informed AI adoption in radiology.

In radiology AI, assigning standardized weightings to performance metrics such as Positive Concordance Rate (PCR), Negative Concordance Rate (NCR), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Concordance, and Augmented Findings Rate (AFR) is crucial for evaluating algorithm quality. However, there is no universally accepted standard for these weightings; they often depend on the clinical context and the specific application of the AI tool. Each user will have the ability to select the weightings for each variable thereby defining their own success thresholds.

AIQ Metric Weighting Framework

Metric	Clinical Priority Context	Importance Weighting	Notes
Positive Concordance Rate (PCR)	High-risk diagnoses (e.g., cancer screening)	Very High	Prioritize to minimize missed cases.
Negative Concordance Rate (NCR)	Resource-limited settings, overdiagnosis risk	High	Essential to reduce false positives.
Positive Predictive Value (PPV)	Resource-limited, low-tolerance for false positives	High	Optimize to ensure positives are true.
Negative Predictive Value (NPV)	High-risk diagnoses, low-prevalence diseases	Very High	Critical for ruling out disease safely.
Concordance	General AI validation across populations	Moderate to High	Agreement with radiologists.
Augmented Findings Rate (AFR)	AI augmentation (supporting radiologists)	High	Measures AI's incremental value

By customizing metric weighting to the specific diagnostic use case, AIQ ensures that the AIQ Score accurately mirrors real-world clinical expectations and operational requirements. This approach supports nuanced, context-specific evaluation of AI system performance, improving the reliability and clinical relevance of continuous monitoring.

Bridging AI Performance with Radiologic Workflow: While diagnostic accuracy is essential, the clinical value of AI in radiology also depends on its ability to improve operational efficiency and workflow integration. The AIQ framework extends performance evaluation beyond interpretive metrics to systematically assess the operational impact of AI deployment.

Productivity metrics, including RVUs and study volumes per shift, offer quantitative evidence of AI's impact on radiologist throughput. When tracked longitudinally, these indicators can reveal trends in workload distribution, potential reductions in burnout, and support resource allocation strategies within departments. Moreover, assessing AI's impact on workflow coordination, such as study prioritization, automated flagging, and structured reporting, helps identify areas where AI can drive sustainable operational improvements.

SECTION 2: BIALOGICS AIQ COMMERCIAL PROGRAM

AIQ Framework Solution Overview: Bialogics offers a vendor-neutral, interoperable platform designed to consolidate and manage all radiology-related data within a unified database. This includes data from Radiology Information Systems (RIS), Picture Archiving and Communication

Systems (PACS), and structured reference tables such as Relative Value Units (RVUs) and subspecialty classifications. The platform also integrates output data from any AI algorithm, regardless of vendor.

Engineered for deployment either on virtual machine (VM) infrastructure or cloud-based environments, the Bialogics platform ensures that all patient data remains securely hosted within the client's network. This architecture supports compliance with institutional data governance and privacy policies by maintaining local data residency and preventing external data exposure.

AI output data is ingested directly from the algorithm using multiple standardized methods. These include secondary capture mechanisms embedded within the Structured Report Information (SRI) field of the DICOM header, direct HL7 messages generated by the AI model, or integration via third-party AI platform.

In the real-time radiology workflow, images are transmitted directly from the modality to the AI algorithm for processing. Once analyzed, the AI-generated results are simultaneously routed to the PACS and to Bialogics, with clinical findings embedded within the DICOM header and displayed in real time on the radiologist's worklist. As the radiologist reviews and interprets the study, the AI output is available alongside the images to support decision-making. Upon finalization of the radiology report, results are sent to both the RIS and Bialogics. The Bialogics AIQ Framework then compares the AI-generated findings with the finalized radiologist report, enabling continuous measurement and calculation of the AIQ Score to assess AI performance and clinical alignment.

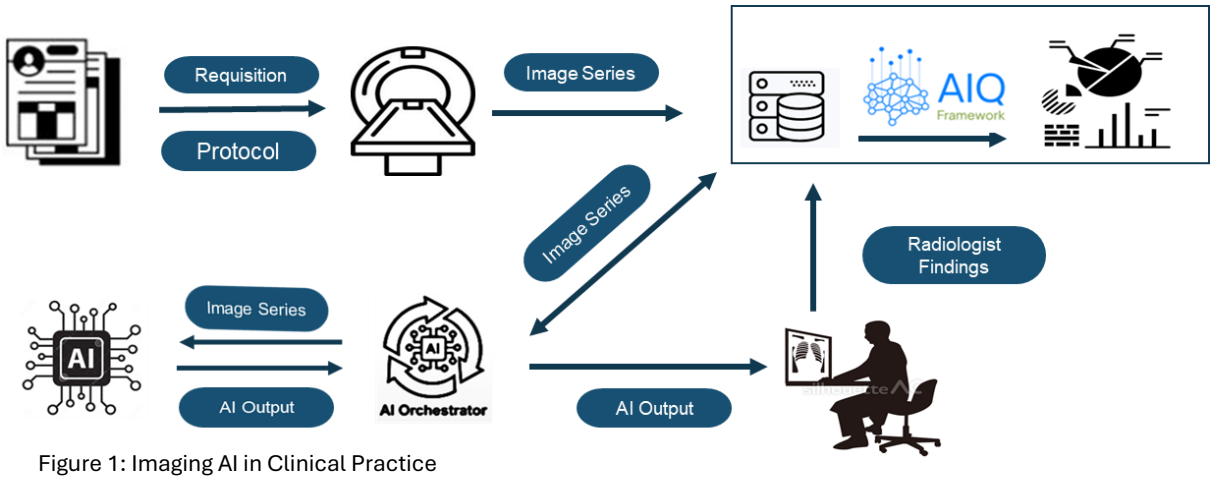

Figure 1: Imaging AI in Clinical Practice

The AIQ Framework is offered as a stand-alone solution or as a component of the DImax Analytics platform.

AIQ Framework Platform – Features and Capabilities

The **AIQ Framework Platform** is purpose-built to support the needs of both clinical AI users and data scientists, offering a comprehensive suite of tools for AI research, performance monitoring, and operational analysis within radiology environments.

1. **Advanced Cohort Identification**
   The platform includes robust cohort search functionality, enabling users to define and retrieve patient populations that meet specific clinical criteria. Searches can be conducted using SNOMED CT concepts, structured clinical parameters, or free-text queries, supporting precision cohort creation for AI model testing, validation, and retrospective analysis.

2. **Dynamic Visualization and Reporting**
   A configurable, user-friendly dashboard environment allows users to build customized visualizations and reports. These tools provide in-depth insights into AI performance, productivity impact, and workflow integration, offering clear transparency into how AI is influencing radiology operations at both the case and population levels.

3. **AIQ Framework Metrics and Scoring**
   The platform includes a suite of preconfigured analytic tools and scoring methodologies to support standardized evaluation of AI algorithms. Key metrics include:

   - **Concordance/Discordance** with radiologist interpretations
   - **Positive Concordance Rate (PCR) / Negative Concordance Rate (NCR),** with accuracy and F1 analysis
   - **Negative Predictive Value (NPV) / Positive Predictive Value (PPV)**
   - **Drift Reporting,** trending of accuracy and AIQ Score over time and forecasting drift
   - **Augmented Findings Rate**, assessing AI's contribution to diagnostic yield
   - **AI Prevalence Analysis**, evaluating AI-triggered case distributions
   - **AIQ Score Analysis**, quantifying overall AI performance and value across multiple dimensions

Integrating such metrics within the AIQ framework allows for a unified, evidence-based approach to evaluating AI's clinical and operational value. By quantifying not only diagnostic accuracy but also efficiency and workflow optimization, AIQ facilitates a multidimensional understanding of AI performance, aligning technical innovation with real-world radiologic practice.

Real-Time Monitoring and Continuous AI Assessment: A core strength of the AIQ framework is its ability to enable real-time, continuous performance monitoring, ensuring that AI models sustain high diagnostic accuracy and operational efficiency after deployment. The continuous assessment process includes several key components:

- **Live AI Performance Tracking:** AI predictions are validated against radiologist-ground truth diagnoses to measure ongoing accuracy.
- **Anomaly Detection and Alerts:** Any significant drop in AIQ score triggers an alert, prompting immediate human review and intervention.
- **Comparative Benchmarking:** AIQ enables institutions to compare multiple AI models, ensuring the selection of high-performing algorithms.
- **Feedback Loop for AI Optimization:** Continuous AIQ monitoring informs model refinement, reducing biases and improving adaptability.
- **Support for Regulatory Compliance:** offering a standardized and measurable approach to AI evaluation, ensuring alignment with approval processes and future regulatory processes

Through these mechanisms, AIQ transforms AI quality assurance from a static, retrospective process into a dynamic, evidence-driven model capable of maintaining clinical integrity across the AI lifecycle.

Conclusion: This dynamic, intelligent integration positions the AIQ Score not merely as a performance audit tool but as a clinical intelligence layer that enhances radiologist reporting, augments diagnostic accuracy, and ensures that AI systems operate with contextual concordance and high clinical value. It supports transparent decision-making for AI governance, fosters clinician trust, and elevates the role of AI from isolated diagnostic support to a fully embedded, learning partner in patient care.

About Bialogics: Bialogics stands at the forefront of precision analysis in Diagnostic Imaging data and patient workflows. Their innovative approach establishes a real-time environment aimed at harnessing clinical and operational data to enhance business performance and clinical outcomes. Through the integration of multiple imaging data sources and deep learning ML, Bialogics has developed a practical solution for all sizes of Diagnostic Imaging organizations. This solution offers ongoing evaluation of Business and Clinical intelligence analytics as well as AIQ Framework and AI Scoring of AI algorithms in current use or for retrospective research, enabling real-time monitoring of AI metrics.

Contact: info@bialogics.com