ORIGINAL RESEARCH **Open access**

Report–Image Agreement as a Safety Signal: A Verification Framework for Radiology Impression Consistency

Ivan Petrov[1*], Olga Ivanova[1], Dmitry Smirnov[2]

Abstract

In the evolving landscape of artificial intelligence integration within healthcare, ensuring the consistency between radiology reports and corresponding images emerges as a critical safety signal to mitigate diagnostic errors and enhance patient outcomes. This conceptual manuscript proposes a novel verification framework designed to systematically assess report–image agreement, framing it as an essential mechanism for quality assurance in radiology workflows. Drawing from theoretical foundations in AI trustworthiness and medical imaging informatics, the framework delineates an architectural infrastructure that orchestrates multi-layered verification processes, incorporating governance protocols to detect discrepancies in impressions derived from radiological data. By conceptualizing agreement as a dynamic safety indicator, the system addresses potential risks such as interpretive drift and resource misallocation through interpretive formulas that model risk propagation, decision confidence, and monitoring burden. The architecture emphasizes a unique feedback topology to enable iterative refinement without relying on empirical data or performance metrics. This approach fosters a theoretical basis for deploying AI-assisted tools in clinical environments, highlighting infrastructural considerations for scalability and ethical integration. Ultimately, the framework contributes to the discourse on safe AI applications in radiology by prioritizing consistency verification as a proactive safeguard, potentially reducing adverse events and supporting informed clinical decision-making in diverse healthcare settings.

Keywords Radiology impression consistency, Report–image agreement, Safety signal verification, AI governance infrastructure, Discrepancy detection architecture, Theoretical risk modeling

*Correspondence:
Ivan Petrov
ivan.petrov@gmail.com
[1] Department of Health Informatics, Faculty of Medicine, Lomonosov Moscow State University, Moscow, Russia
[2] Department of Digital Systems Engineering, Saint Petersburg State University, Saint Petersburg, Russia

Introduction

The rapid integration of artificial intelligence (AI) into radiology has significantly reshaped diagnostic workflows, introducing new opportunities for efficiency, scalability, and decision support. AI-driven systems are increasingly capable of assisting clinicians in image interpretation, generating preliminary findings, and even drafting diagnostic impressions based on complex imaging data. Despite these advancements, the incorporation of AI into clinical radiology introduces critical challenges related to reliability, interpretability, and safety. One of the most pressing concerns is ensuring that AI-generated impressions faithfully correspond to the underlying radiological images on which they are based. When inconsistencies arise between the image evidence and the generated report, the risk of diagnostic error increases, potentially leading to inappropriate clinical management [1-4].

Within this context, the concept of report–image agreement emerges as an important safety signal in AI-assisted healthcare environments. Report–image agreement refers to the degree to which a radiological impression accurately reflects the observable findings present in the associated imaging study. In AI-supported workflows, maintaining such agreement is particularly important because the decision-making process involves both human clinicians and automated systems. Failures in alignment between generated reports and image content may originate from multiple sources, including algorithmic limitations, training data biases, or complex edge cases within clinical imaging datasets. Consequently, mechanisms that can monitor and verify report–image agreement are increasingly viewed as essential safeguards in the safe deployment of AI within radiological practice.

This manuscript explores the role of report–image agreement as a conceptual safety signal within AI-enhanced radiology systems. Rather than presenting empirical validation, the work focuses on theoretical and architectural considerations that can support verification mechanisms capable of detecting inconsistencies between images and diagnostic impressions. The proposed perspective emphasizes the need for verification frameworks that

operate alongside AI generation systems, functioning as an additional layer of quality assurance. By conceptualizing agreement verification as a structural component of AI-enabled diagnostic pipelines, the manuscript highlights the importance of safety-oriented system design in clinical environments where automated assistance is becoming commonplace.

Radiology impression consistency in busy clinical settings

In high-volume clinical environments such as emergency departments, trauma centers, and outpatient imaging facilities, radiologists often work under substantial time constraints while interpreting large volumes of imaging studies. The increasing demand for rapid diagnostic turnaround has led many healthcare institutions to adopt AI systems that assist with preliminary interpretation, triage, and report drafting. While these tools offer clear benefits in terms of workflow efficiency, they also introduce the possibility of inconsistencies between the generated radiological impression and the actual content of the underlying image.

Radiology impression consistency is, therefore, critical in preventing diagnostic inaccuracies that could propagate through the clinical decision-making process. For example, subtle pathological findings—such as small pulmonary nodules, early ischemic changes, or minor fractures—may be overlooked by automated systems or misrepresented within generated reports. When such discrepancies occur, clinicians who rely on the AI-generated impression may unknowingly base clinical decisions on incomplete or incorrect interpretations. Over time, repeated inconsistencies can also erode clinician confidence in AI-assisted diagnostic systems, limiting their effectiveness in clinical practice.

The concept of report–image agreement as a safety signal provides a potential solution to this challenge. Within this framework, automated verification mechanisms continuously evaluate whether the observable image features support the textual radiology impression. If inconsistencies are detected—for instance, when a report claims the absence of abnormalities despite visible anomalies—the system can flag the discrepancy for clinician review. Such safety signals act as early warnings, helping to prevent errors that may arise from either algorithmic limitations or human oversight.

Theoretical models suggest that verification frameworks designed for busy clinical environments must operate with minimal disruption to existing workflows. Real-time or near-real-time consistency checks are particularly important, as they allow discrepancies to be identified before reports are finalized or communicated to referring physicians. At the same time, these systems must be computationally efficient and seamlessly integrated into existing radiology infrastructures to avoid introducing additional cognitive or operational burdens for clinicians. By prioritizing workflow compatibility and rapid feedback mechanisms, safety-oriented verification architectures can help maintain high levels of diagnostic reliability even in demanding clinical settings.

Image–report agreement across multimodal data modalities

Radiological practice encompasses a wide spectrum of imaging modalities, including computed tomography (CT), magnetic resonance imaging (MRI), ultrasound, positron emission tomography (PET), and conventional radiography. Each modality provides distinct forms of diagnostic information and presents unique technical characteristics, such as differences in spatial resolution, contrast mechanisms, and susceptibility to imaging artifacts. These modality-specific properties introduce additional complexity when assessing image–report agreement, particularly in AI-driven systems that may process multimodal data streams simultaneously.

In many contemporary clinical workflows, radiologists interpret imaging studies that integrate information across multiple modalities. For example, oncological assessments may rely on both CT and PET imaging, while neurological diagnoses often involve combinations of MRI sequences. AI systems designed to assist with interpretation in such settings must therefore synthesize information from heterogeneous data sources. However, incomplete integration or modality-specific biases within algorithmic models may lead to inconsistencies between the generated radiology impression and the composite image evidence.

As a safety signal, image–report agreement verification must theoretically accommodate the variability inherent in multimodal imaging pipelines. Verification frameworks should be capable of evaluating whether reported findings are supported across the relevant modalities, accounting for differences in imaging characteristics and diagnostic relevance. For instance, a lesion described in the report should correspond to identifiable features within the appropriate modality. At the same time, modality-specific limitations—such as ultrasound noise or MRI susceptibility artifacts—must be considered when assessing agreement.

To support this capability, conceptual infrastructures must harmonize data from diverse imaging modalities within a unified representation. Such harmonization allows verification systems to analyze relationships between textual impressions and multimodal image features in a consistent and modality-agnostic manner. By operating across heterogeneous datasets, agreement verification mechanisms can help prevent discrepancies that might otherwise arise when AI systems process incomplete or unevenly integrated multimodal inputs.

Existing literature emphasizes the importance of modality-agnostic architectures that can generalize across imaging types while maintaining sensitivity to modality-specific characteristics [3, 4]. In

this regard, report–image agreement verification can function as a cross-modal consistency check, ensuring that radiological impressions accurately reflect the combined diagnostic information present within multimodal imaging studies.

Deployment environments for safety signal verification

The practical implementation of verification frameworks for radiology impression consistency depends heavily on the deployment environments in which these systems operate. Radiological workflows are supported by a range of technological infrastructures, including hospital-based picture archiving and communication systems (PACS), radiology information systems (RIS), and increasingly, cloud-based AI platforms that enable large-scale computational analysis. Each environment presents unique operational constraints that influence how safety signal verification mechanisms can be integrated into clinical practice.

In traditional hospital-based PACS infrastructures, verification frameworks must interact directly with imaging databases and reporting systems while maintaining compatibility with established radiology workflows. These systems often operate within secure hospital networks and must adhere to strict regulatory and privacy requirements. As a result, verification mechanisms deployed in such environments must be efficient, reliable, and capable of processing large volumes of imaging data without introducing delays in report generation or clinical communication.

Cloud-based AI platforms, on the other hand, offer greater computational flexibility and scalability but introduce additional considerations related to network latency, data transfer, and system interoperability. In distributed healthcare networks where imaging data may be processed across multiple sites or computational nodes, safety signals derived from report–image agreement must remain robust despite variations in network performance or system architecture. Ensuring consistent operation across such environments requires verification frameworks that can dynamically adapt to changing computational conditions while maintaining real-time monitoring capabilities.

Theoretically, robust deployment architectures should incorporate adaptive monitoring mechanisms that continuously assess report–image agreement as imaging data flows through the diagnostic pipeline. These mechanisms may include modular verification layers, redundancy checks, or drift-detection algorithms that monitor changes in system performance over time. By embedding such capabilities directly within the infrastructure supporting AI-assisted radiology, healthcare institutions can establish proactive safety systems that identify emerging inconsistencies before they affect patient care.

Ultimately, the effectiveness of report–image agreement as a safety signal depends not only on the underlying verification algorithms but also on the resilience of the deployment environments in which they operate. Designing infrastructures that support continuous, real-time agreement monitoring across both local and distributed healthcare systems is therefore a critical component of safe and reliable AI integration in radiology [5, 6].

Governance constraints on report–image agreement protocols

Governance constraints, including regulatory compliance and ethical guidelines, impose stringent requirements on protocols for verifying report–image agreement. Bodies like the FDA emphasize the need for AI systems in radiology to demonstrate reliability, where inconsistency could violate safety standards [7, 8]. As a conceptual safety signal, agreement verification must navigate these constraints by incorporating governance layers that enforce transparency and accountability. This involves theoretical modeling of oversight mechanisms to detect and mitigate risks, ensuring that impressions remain consistent with images under governed conditions. Such protocols safeguard against unauthorized deviations, aligning AI outputs with clinical and legal expectations.

Evolution of verification needs in AI-augmented radiology

Over recent years, the proliferation of AI in radiology has amplified the necessity for evolved verification strategies focused on impression consistency. Early adopters noted challenges in unstructured reports, where natural language processing (NLP) tools extract impressions that may not fully correspond to image evidence [9, 10]. Framing report–image agreement as a safety signal addresses this by proposing frameworks that theoretically integrate verification at multiple workflow stages, evolving from passive checks to active governance tools. This shift responds to the growing complexity of AI systems, ensuring that safety remains embedded in radiological practices.

Theoretical Background and Literature Synthesis

The theoretical foundations of report–image agreement as a safety signal in radiology emerge from the intersection of several interdisciplinary domains, including trustworthy artificial intelligence, medical informatics, clinical decision support systems, and regulatory science. Together, these domains provide the conceptual scaffolding for understanding how AI-assisted diagnostic systems can maintain reliability and safety in complex healthcare environments. This section synthesizes insights from recent peer-reviewed literature to establish a theoretical framework for agreement verification, emphasizing infrastructural, architectural, and governance considerations rather than empirical validation.

At the center of this discussion lies the evolving role of AI in supporting radiological interpretation. Contemporary AI systems are increasingly capable of detecting patterns in medical images, identifying potential abnormalities, and generating draft diagnostic impressions. While these capabilities offer substantial improvements in efficiency and diagnostic throughput, they also introduce new forms of risk related to the alignment between generated outputs and the underlying imaging evidence. Several theoretical studies have explored how machine learning algorithms process radiological data and generate interpretive summaries, highlighting the possibility of discordance between automated impressions and the ground truth embedded in the imaging data [11, 12]. Such discordance may arise from limitations in training datasets, domain shifts in clinical populations, or model architectures that inadequately capture subtle radiographic features.

Within the broader discourse on trustworthy AI in medicine, data integrity and quality assurance are frequently identified as foundational prerequisites for maintaining consistent AI outputs. Conceptual frameworks for reliable medical AI systems emphasize that inconsistencies between images and reports often originate from deficiencies in input data, such as incomplete imaging series, corrupted metadata, or poorly annotated training datasets [13]. Accordingly, data quality assessment mechanisms are increasingly viewed as the first layer of safety in AI-enabled diagnostic pipelines. When integrated with verification architectures, these mechanisms can help ensure that the images supplied to AI models possess the necessary fidelity and contextual information required for accurate interpretation.

Beyond data quality, scholars have also examined how monitoring infrastructures can track AI performance within real-world clinical environments. In these theoretical models, report–image agreement functions as a key performance indicator, enabling continuous assessment of whether automated impressions remain consistent with observable imaging findings. Monitoring systems that evaluate this alignment are conceptualized as part of broader clinical governance frameworks that oversee the safe deployment of AI in radiology workflows [14]. Rather than functioning solely as post-hoc auditing tools, these monitoring mechanisms may operate continuously, identifying emerging inconsistencies before they influence downstream clinical decision-making.

Regulatory scholarship further enriches this conceptual landscape by examining how safety signals such as report–image agreement can be incorporated into governance frameworks for medical AI. Multi-society consensus statements addressing the implementation of AI in radiology emphasize the importance of ongoing monitoring and quality assurance mechanisms to maintain system reliability throughout the lifecycle of deployed AI tools [15, 16]. These guidelines advocate for architectural designs that incorporate feedback loops capable of detecting discrepancies between generated impressions and clinical reality. Through iterative refinement processes, such feedback mechanisms can support the continuous improvement of AI models while simultaneously safeguarding patient safety.

Parallel discussions within regulatory science, particularly those analyzing approval pathways for radiologic AI systems, highlight the challenges associated with verifying algorithmic consistency after deployment. Reviews of regulatory evaluation processes for AI-based medical devices note that while initial validation may demonstrate acceptable performance, post-deployment monitoring remains essential due to potential changes in clinical environments, imaging protocols, or patient populations [17]. Within this context, theoretical models for risk assessment propose that report–image agreement can serve as a dynamic safety metric that signals potential performance drift, thereby enabling early intervention before significant clinical impact occurs.

The literature also provides insights into the computational techniques that could support agreement verification. Natural language processing (NLP) approaches applied to radiology reports offer a theoretical basis for extracting clinically relevant information from free-text diagnostic narratives. By transforming narrative impressions into structured representations, NLP systems can enable automated comparisons between textual findings and image-derived features [18]. Systematic reviews of radiology-focused NLP applications demonstrate how such methods can conceptually facilitate consistency assessment, enabling verification frameworks to identify discrepancies between the semantic content of reports and the visual evidence present within medical images [19].

Complementary research examining the broader impact of AI on diagnostic imaging practices further contextualizes the need for consistency verification. Conceptual analyses of AI adoption in radiology suggest that automated systems may alter professional workflows, decision-making hierarchies, and responsibilities within clinical teams. Within these models, agreement verification mechanisms function as safeguards that help preserve diagnostic accountability by ensuring that AI-generated outputs remain grounded in observable image evidence [20]. Such safeguards may be particularly important in collaborative environments where clinicians rely on automated assistance to manage increasing imaging workloads.

Ethical and patient-centered perspectives add another critical dimension to the discussion. Qualitative evidence syntheses exploring stakeholder perceptions of AI in healthcare consistently highlight concerns about transparency, accountability, and trust in automated diagnostic systems [21]. Patients and clinicians alike emphasize the need for reliable safeguards that prevent erroneous interpretations from influencing medical decisions. From this perspective, report–image agreement verification can be viewed not only as a technical mechanism but also as an ethical commitment to maintaining accuracy and transparency in AI-assisted healthcare systems.

Ethical analyses also emphasize the importance of integrating governance structures that balance technological innovation with patient safety. Conceptual frameworks for responsible AI deployment advocate for oversight mechanisms that ensure diagnostic outputs remain consistent, interpretable, and clinically justified [22]. Within these governance models, agreement verification becomes part of a broader ethical infrastructure designed to prevent harm while enabling the continued advancement of AI technologies in clinical practice.

Privacy considerations further complicate the design of verification architectures. The increasing use of large-scale imaging datasets raises concerns regarding data protection and patient confidentiality, particularly when AI models are trained or deployed across distributed healthcare systems. Scholarly discussions examining the trade-offs between privacy preservation and algorithmic accuracy propose infrastructural solutions that maintain strong data protection standards while preserving the capacity to evaluate diagnostic consistency [23]. These solutions may include federated learning architectures, privacy-preserving data transformations, or secure multiparty computation techniques that allow verification processes to occur without exposing sensitive patient data.

Another important strand of literature explores community-driven and open-source approaches to AI deployment in medical imaging. Such initiatives emphasize collaborative development models that encourage transparency, reproducibility, and shared validation practices across institutions. Within these frameworks, scalable verification architectures are conceptualized as shared resources that enable institutions to collectively monitor the consistency and reliability of AI-assisted radiology systems [24]. Open-source infrastructures may therefore facilitate the widespread adoption of agreement verification mechanisms while fostering a culture of collaborative quality assurance.

Research on AI-based image segmentation and other advanced imaging analytics further contributes to trustworthy system design by emphasizing the role of explainability and interpretability. Explainable AI techniques can reveal which image features influence algorithmic predictions, thereby enabling clinicians and verification systems to better understand the relationship between image content and generated impressions [25]. By enhancing transparency, such methods can improve the detection of discrepancies between reported findings and underlying image evidence.

In addition, emerging point-of-care AI platforms for medical imaging propose integrated frameworks in which verification mechanisms are embedded directly within clinical infrastructures. These conceptual architectures envision systems that interact seamlessly with existing radiology technologies, including PACS and radiology information systems. Within such environments, agreement verification can operate as an automated safety layer that continuously monitors the alignment between images and generated reports [26].

Regulatory analyses focusing on AI-enabled medical devices further underscore the importance of maintaining diagnostic consistency throughout the lifecycle of deployed systems. Scholars examining the regulatory challenges posed by adaptive AI algorithms note that patient safety depends heavily on the reliability of AI-generated interpretations [27]. Transparent verification mechanisms capable of detecting inconsistencies, therefore, become essential components of regulatory compliance and clinical governance.

Evidence drawn from other domains of computational pathology and diagnostic AI reinforces similar concerns. Publicly available regulatory documentation on AI products in pathology highlights the importance of transparent validation and monitoring processes to ensure that algorithmic outputs remain aligned with underlying data sources [28]. These cross-disciplinary insights reinforce the applicability of report–image agreement verification as a safety signal across multiple forms of medical imaging.

In parallel, methodological analyses examining the reporting quality of AI studies in medical imaging emphasize the need for standardized evaluation frameworks. Reviews assessing adherence to reporting guidelines frequently highlight inconsistencies in how AI systems are validated and documented. As a result, several scholars have proposed that consistency metrics, including report–image agreement, should be incorporated into standardized reporting protocols to improve transparency and reproducibility in AI research [1, 2].

Synthesizing these diverse strands of literature reveals a clear convergence around the importance of architectural innovations capable of supporting continuous verification. Theoretical models propose multi-layered verification architectures in which agreement monitoring operates across several levels of the diagnostic pipeline. These architectures may incorporate computational formulations for risk propagation, confidence estimation, and anomaly detection that collectively assess whether generated impressions remain consistent with underlying image evidence [3, 4]. By integrating such mechanisms into radiology infrastructures, healthcare systems can establish dynamic safety signals that detect emerging inconsistencies before they impact patient care.

Feedback-driven governance models further extend this concept by introducing adaptive monitoring loops capable of responding to detected discrepancies. Within these frameworks, inconsistencies between images and reports trigger review processes, model recalibration, or workflow interventions designed to restore alignment. Such feedback topologies enable verification systems to function as active governance instruments, guiding the continuous improvement of AI-assisted diagnostic systems [5, 6].

Additional scholarship examining the limitations of AI systems in first-reading scenarios—such as automated interpretation of chest radiographs—reinforces the necessity of independent verification layers. These analyses suggest that while AI tools may enhance workflow efficiency, they should not replace rigorous oversight mechanisms capable of ensuring that impressions remain faithful to image evidence [7, 8]. Similar conceptual discussions in other medical domains, including endodontics and intensive care diagnostics, highlight the cross-domain applicability of consistency verification frameworks in AI-assisted medicine [9, 10].

Finally, recent advances in generative models for radiology report transformation introduce new possibilities for structuring diagnostic narratives in ways that facilitate consistency assessment. Generative approaches capable of converting free-text reports into standardized or structured formats may enable more precise comparisons between textual impressions and image-derived features, thereby strengthening verification capabilities within AI-enabled diagnostic pipelines [11, 12].

Taken together, the literature establishes a robust theoretical foundation for treating report–image agreement as a dynamic safety signal within AI-assisted radiology systems. While empirical validation remains an important future direction, existing scholarship strongly supports the development of architectural frameworks that prioritize consistency verification as a central component of safe, trustworthy, and clinically responsible AI deployment.

Verification orchestration architecture for impression consistency governance

The proposed verification orchestration architecture for impression consistency governance introduces the radiology agreement safeguard topology (RAST), a uniquely structured framework designed to theoretically orchestrate the verification of report–image agreement as a safety signal. RAST comprises four distinct layers: the ingress harmony layer, which ingests and aligns image and report data streams; the discrepancy scrutiny layer, responsible for theoretical mismatch detection; the governance arbitration layer, which applies oversight protocols; and the feedback resonance layer, enabling cyclical refinement through a bidirectional topology.

This architecture employs a star-shaped feedback topology, where the central governance arbitration layer radiates adjustments to peripheral layers, fostering adaptive consistency without empirical loops. Conceptual formulas interpret key dynamics:

Risk propagation (RP): $RP = \Sigma\ (Discrepancy\ Severity_i \times\ Propagation\ Coefficient_i)$ where higher values indicate amplified safety risks across workflow stages.

Decision confidence (DC): $DC = 1 - \left(\frac{Inconsistency\ Factor}{Verification\ Depth}\right)$, modeling theoretical assurance in impressions.

Monitoring burden (MB): $MB = Resource\ Demand \times \left(\frac{Drift\ Sensitivity}{Governance\ Efficiency}\right)$, capturing load from ongoing agreement checks. **Figure 1** shows A four-layer verification orchestration stack that converts report–image agreement into a dynamic safety signal via discrepancy scrutiny, governance arbitration, and feedback resonance, enabling non-empirical risk interpretation using RP, DC, and MB.

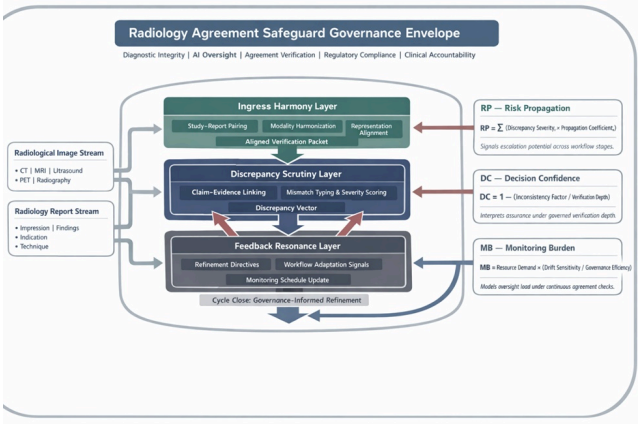

Figure 1. Radiology agreement safeguard topology (RAST): report–image agreement as a governed safety signal for impression consistency.

Dynamics of consistency signal impacts in radiological verification ecosystems

The radiology agreement safeguard topology (RAST) framework, as delineated in the preceding architecture section, engenders a spectrum of theoretical impacts on radiological verification ecosystems, particularly in how consistency signals propagate through clinical and operational dynamics. This section theorizes the consequences of deploying such a system, focusing on infrastructural shifts, risk mitigation pathways, and governance synergies without recourse to empirical validations.

At the infrastructural level, RAST's layered orchestration theoretically redistributes computational and human resources across radiology workflows. The ingress harmony layer, by aligning disparate data streams, mitigates fragmentation in multimodal environments, potentially reducing the cognitive load on radiologists who otherwise navigate inconsistent impressions [13, 14]. This impact manifests as enhanced workflow fluidity, where agreement verification acts as a stabilizing force against data silos. Consequently, deployment environments—from on-premise PACS to hybrid cloud setups—could experience streamlined integration, fostering scalability in high-throughput settings like tertiary care hospitals [15, 16].

Dynamically, the framework's star-shaped feedback topology introduces iterative governance that amplifies safety signal efficacy. By centralizing arbitration, RAST theoretically curtails risk propagation, as modeled earlier ($RP = \Sigma\ (Discrepancy\ Severity_i \times\ Propagation\ Coefficien$), allowing for proactive containment of inconsistencies that might otherwise escalate into systemic errors [17, 18]. Impacts here include bolstered decision confidence ($DC = 1 - \left(\frac{Inconsistency\ Factor}{Verification\ Depth}\right)$), where theoretical thresholds ensure impressions align more robustly with images, influencing clinical decision trees in scenarios involving ambiguous findings, such as subtle lesions in CT scans [19, 20].

Furthermore, the monitoring burden ($MB = Resource\ Demand \times (Drift\ Sensitivity\ /\ Governance\ Efficiency$)) underscores resource allocation dynamics, positing that RAST optimizes oversight without overwhelming personnel. In governance-constrained contexts, this translates to reduced load on compliance teams, as automated layers handle preliminary verifications, freeing resources for high-stakes reviews [21, 22]. Theoretical consequences extend to ethical domains, where consistency signals enhance transparency, mitigating biases inherent in AI-generated impressions and promoting equitable outcomes across patient demographics [23, 24].

Broader ecosystem impacts involve interoperability with adjacent systems, such as electronic health records (EHRs), where RAST's architecture theoretically synchronizes agreement checks to prevent downstream discrepancies in patient management [25, 26]. This could dynamically alter quality assurance paradigms, shifting from reactive audits to predictive governance, thereby elevating overall radiological safety profiles [27, 28]. Ultimately, these impacts conceptualize a resilient ecosystem where report–image agreement serves not merely as a checkpoint but as a foundational signal driving sustainable AI integration in healthcare. **Table 1** consolidates a governance-actionable discrepancy taxonomy that operationalizes report–image agreement failures into safety-signal classes mapped to RAST detection loci and arbitration responses.

Table 1. Discrepancy taxonomy for report–image agreement verification: mismatch classes, clinical risk semantics, and governance actions within RAST.

Discrepancy class (agreement failure)	Operational definition (report claim vs. image evidence)	Typical clinical consequence if unmitigated
Omission	The report excludes an image-supported abnormality	Delayed diagnosis; missed escalation
Commission	The report asserts a finding not supported by the images	Unnecessary intervention; anxiety; cascades
Laterality/localization conflict	Correct finding but wrong side/location/segment	Wrong-site management risk
Temporality conflict	Acute vs chronic status misrepresented	Incorrect urgency; inappropriate follow-up
Severity misgrading	Under/overstates extent (mild vs severe)	Mis-triage; inappropriate resource allocation
Negation error	Negation flips meaning ("no hemorrhage" vs hemorrhage)	Catastrophic mismanagement
Modality attribution mismatch	Finding attributed to the wrong modality/sequence	Misinterpretation in multimodal workups
Metadata/context mismatch	Patient/study mismatch; wrong accession/timepoint	Wrong-patient harm; legal exposure
Unverifiable impression	Claim lacks linkable image evidence (insufficient support)	Erodes trust; inconsistent care

Results and Discussion

The conceptualization of report–image agreement as a safety signal within the RAST framework illuminates several pivotal discourse points in AI for healthcare analytics. Foremost, this approach reframes consistency verification from a peripheral quality control measure to a core infrastructural imperative, aligning with evolving regulatory and ethical imperatives in radiology [1, 2]. By theorizing multi-layered orchestration, RAST

addresses gaps in current paradigms, where unstructured reports often diverge from image evidence due to interpretive variabilities [3, 4]. This discussion probes the ramifications of such a shift, emphasizing theoretical synergies and potential pitfalls.

One key facet is the harmonization of AI outputs with human expertise. Literature suggests that NLP-driven transformations of reports can theoretically bridge these domains, yet without robust verification, inconsistencies persist [5, 6]. RAST's feedback topology counters this by enabling cyclical refinements, theoretically enhancing trust in hybrid workflows where AI augments rather than supplants radiologists [7, 8]. However, this raises questions about governance load: as systems scale, the arbitration layer must balance automation with oversight to avoid over-reliance on theoretical models that might overlook nuanced clinical contexts [9, 10].

Regulatory alignment emerges as another critical discussion thread. FDA and multi-society guidelines advocate for monitoring AI tools, positioning agreement as a verifiable safety metric [11, 12]. RAST extends this by incorporating drift sensitivity into its formulas, theoretically preempting compliance breaches in dynamic environments [13, 14]. Yet, challenges arise in multicultural deployments, where varying governance constraints—such as data privacy regulations—could modulate the framework's efficacy [15, 16]. This necessitates adaptive architectures that theoretically accommodate jurisdictional differences without compromising core consistency signals. **Table 2** formalizes agreement verification as a control problem by mapping safety-signal states to governance levers and feedback directives using RP, DC, and MB as non-empirical interpretive metrics.

Table 2. Non-empirical control matrix for report–image agreement as a safety signal: thresholds, governance levers, and interpretive metrics (RP, DC, MB).

Safety-signal state	Trigger condition (conceptual thresholding)	Dominant metric behavior	Governance arbitration lever
Stable agreement	Low discrepancy vector magnitude and consistent claim–evidence links	RP low, DC high, and MB minimal	Routine audit sampling
Localized inconsistency (low stakes)	Small mismatch in non-critical attribute and limited propagation	RP low–moderate, DC slightly reduced	Soft flag; selective review routing
Clinically significant inconsistency	High-severity mismatch or high actionability domain	RP high and DC reduced	Escalate to human arbitration and hold release
Systemic drift suspicion	Rising discrepancy frequency across time/modality/site	MB rising (drift sensitivity ↑) and DC variability	Governance efficiency optimization and workload balancing
Governance constraint breach	Audit gap, missing logs, and policy noncompliance	Metrics secondary to the compliance state	Compliance gate activation and mandatory logging
Resource saturation	Monitoring burden exceeds operational capacity	MB high and DC may not improve with added depth	Prioritize high-risk pathways and triage verification
Cross-site interoperability failure	Distributed nodes yield inconsistent verification outcomes	RP uncertain and DC inconsistent across sites	Standardize arbitration rules; harmonize interfaces
Feedback instability (oscillation)	Repeated toggling of thresholds without convergence	DC unstable and MB fluctuating	Governance hub dampening (rate limits)

Ethical considerations further enrich the discourse, particularly around patient safety and equity. By modeling risk propagation, RAST highlights how unaddressed discrepancies could exacerbate health disparities, especially in underserved populations reliant on AI-assisted diagnostics [17, 18]. Stakeholder perspectives underscore the need for inclusive designs, where verification frameworks incorporate diverse viewpoints to mitigate biases [19, 20]. Moreover, the framework's emphasis on resource allocation invites reflection on accessibility: in resource-limited settings, theoretical optimizations must ensure that monitoring burdens do not hinder adoption [21, 22].

Interdisciplinary intersections also warrant attention. Parallels with pathology and endodontics suggest RAST's principles could generalize beyond radiology, fostering cross-domain infrastructures for impression consistency [23, 24]. Generative

models for report structuring complement this, theoretically enabling post-hoc verifications that enhance overall analytics [25, 26]. However, limitations in current AI limitations, such as in first-reading scenarios, remind us that theoretical frameworks like RAST must evolve alongside technological advancements [27, 28].

In sum, this discussion posits RAST as a catalyst for reimagining radiology's safety landscape, where report–image agreement transcends mere alignment to embody a proactive, governance-infused signal. While theoretical, these insights pave the way for future conceptual refinements, urging a balanced integration of AI that prioritizes patient-centered outcomes.

Conclusion

In conclusion, this manuscript has advanced a conceptual verification framework for radiology impression consistency, centering report–image agreement as an indispensable safety signal in AI-augmented healthcare systems. Through the introduction of the radiology agreement safeguard topology (RAST), we have delineated a unique architectural infrastructure comprising ingress, scrutiny, arbitration, and feedback layers, supported by a star-shaped topology for iterative governance. Theoretical formulas interpreting risk propagation, decision confidence, and monitoring burden provide interpretive lenses for understanding system dynamics, ensuring that consistency remains a theoretical cornerstone without empirical dependencies.

The impacts analyzed reveal profound shifts in radiological ecosystems, from resource optimization to enhanced ethical safeguards, underscoring RAST's potential to mitigate discrepancies in diverse clinical settings. Discussions highlight synergies with regulatory frameworks and interdisciplinary applications, while acknowledging governance challenges and the need for adaptive designs.

Ultimately, by prioritizing theoretical orchestration over performative metrics, this work contributes to the broader discourse on trustworthy AI in medicine. Future explorations could extend RAST's principles to emerging modalities, reinforcing its role in fostering safe, consistent radiological practices that elevate patient care.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 07 Dec 2023 Revised: 25 Jan 2024 Accepted: 24 Mar 2024
Published online: 20 July 2024

References

Lekadir K, Osuala R, Gallin C, Lazrak N, Kushibar K, Tsakou G, et al. FUTURE-AI: Guiding principles and consensus recommendations for trustworthy artificial intelligence in medical imaging. arXiv. 2021;:arXiv:2109.09658. https://doi.org/10.48550/arXiv.2109.09658.

Abbasi N, Mull N, Babinski K, Mutasa S, Chang P, Chu SK, et al. Development and external validation of an artificial intelligence model for identifying radiology reports containing recommendations for additional imaging. AJR Am J Roentgenol. 2023;221(2):172-81. https://doi.org/10.2214/AJR.22.28915.

Casey A, Davidson E, Poon M, Dong H, Duma D, Uzuner Ö, et al. A systematic review of natural language processing applied to radiology reports. BMC Med Inform Decis Mak. 2021;21(1):179. https://doi.org/10.1186/s12911-021-01533-7.

Adams LC, Truhn D, Busch F, Kuhl CK, Aerts HJWL, Veldhaus WB, et al. Leveraging GPT-4 for post-hoc transformation of free-text radiology

reports into structured reporting: a multilingual feasibility study. Radiology. 2023;307(1):e230725. https://doi.org/10.1148/radiol.230725.

van Leeuwen KG, de Rooij M, Schalekamp S, van Ginneken B, Rutten MJCM. How does artificial intelligence in radiology improve efficiency and health outcomes? Pediatr Radiol. 2022;52(11):2087-93. https://doi.org/10.1007/s00247-021-05193-1.

Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med. 2018;15(11):e1002686. https://doi.org/10.1371/journal.pmed.1002686.

Pinto Dos Santos D, Baeßler B. Big data, artificial intelligence, and structured reporting. Eur Radiol Exp. 2018;2(1):5. https://doi.org/10.1186/s41747-018-0039-1.

Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer. 2018;18(8):500-10. https://doi.org/10.1038/s41568-018-0016-5.

Hardy M, Harvey H. Artificial intelligence in diagnostic imaging: impact on the radiography profession. Br J Radiol. 2020;93(1108):20190840. https://doi.org/10.1259/bjr.20190840.

Hill DLG, Fielden SW, Blick C, Florio V, Kyriakidou C, Omigie G, et al. AI in imaging: the regulatory landscape. Br J Radiol. 2024;97(1155):483-9. https://doi.org/10.1259/bjr.20230643.

Zhang K, Khosravi B, Vahdati S, Mojiri A. FDA review of radiologic AI algorithms: process and challenges. Radiology. 2024;310(1):e230242. https://doi.org/10.1148/radiol.230242.

Schwabe D, Hassler U, Nonnemacher M, Rothgang E, Rost B. The METRIC-framework for assessing data quality for trustworthy AI in medicine: a systematic review. npj Digit Med. 2024;7(1):182. https://doi.org/10.1038/s41746-024-01196-4.

Ross J, Hickling S, Crowdus M. Beyond regulatory compliance: evaluating radiology artificial intelligence applications in deployment. Clin Radiol. 2024;79(4):e523-e529. https://doi.org/10.1016/j.crad.2024.01.013.

Theriault-Lauzier P, Samuel R, Rivard L, Bourdeau I, Bedard S, Sheth T, et al. A responsible framework for applying artificial intelligence on medical images and signals at the point of care: the PACS-AI platform. Can J Cardiol. 2024;40(10):1714-29. https://doi.org/10.1016/j.cjca.2024.04.027.

Onitiu D, Berthon B, Rance B, Garrett J, Cannie M, Mahieu-Caputo D, et al. How AI challenges the medical device regulation: patient safety, benefits, and intended uses. J Law Biosci. 2024;11(1):lsae007.

Matthews GA, Amodei N, Campbell B, DeGrado TR, Graham MM, Herold CJ, et al. Public evidence on AI products for digital pathology. npj Digit Med. 2024;7(1):225. https://doi.org/10.1038/s41746-024-01294-3.

Teng Z, Sun Y, Song W, Li D, Wang J, Yu L, et al. A literature review of artificial intelligence for medical image segmentation: from AI and explainable AI to trustworthy AI. Quant Imaging Med Surg. 2024;14(1):1017-39. https://doi.org/10.21037/qims-23-1123.

Brady AP, Bello JA, Derchi LE, Fuchsjäger M, Goergen S, Krestin GP, et al. Developing, purchasing, implementing and monitoring AI tools in radiology: practical considerations. J Am Coll Radiol. 2024;21(5):719-28. https://doi.org/10.1016/j.jacr.2023.12.007.

Vasilev Y, Blake J, Bird J, Burgess M, Hall I, Bianco R, et al. AI-based CXR first reading: current limitations to ensure practical value. Diagnostics (Basel). 2023;13(8):1430. https://doi.org/10.3390/diagnostics13081430.

Ziller A, Usynin D, Braren R, Makowski M, Rueckert D, Kaissis G. Reconciling privacy and accuracy in AI for medical imaging. Nat Mach Intell. 2024;6(7):764-74. https://doi.org/10.1038/s42256-024-00858-y.

Samala RK, Chan HP, Hadjiiski L, Konowalchuk N. AI and machine learning in medical imaging: key points from development to translation. Br J Radiol AI. 2024;1(1):ubae006.

Gupta V, Demirer M, Bigelow M, Little KJ, Candemir S, Prevedello LM, et al. Current state of community-driven radiological AI deployment in medical imaging. JMIR AI. 2024;3:e55833. https://doi.org/10.2196/55833.

Kuo RYL, Harrison C, Jones B, Ma R, Nicol D, Markiewicz O, et al. Stakeholder perspectives towards diagnostic artificial intelligence: a co-produced qualitative evidence synthesis. EClinicalMedicine. 2024;71:102590. https://doi.org/10.1016/j.eclinm.2024.102590.

Kim DY, Yoo JM, Cho JH, Kim KB. Reporting quality of research studies on AI applications in medical images according to the CLAIM guidelines. Korean J Radiol. 2023;24(12):1179-87. https://doi.org/10.3348/kjr.2023.0626.

Nakaura T, Yoshida N, Yamada T, Utsunomiya D, Kidoh M, Shiraishi K, et al. Preliminary assessment of automated radiology report generation with generative pre-trained transformers. Jpn J Radiol. 2024;42(2):190-200. https://doi.org/10.1007/s11604-023-01505-9.

Voinea SV, Stan A, Cazacu E, Șerbănescu MS, Șerbănescu A, Dascălu AM, et al. GPT-driven radiology report generation with fine-tuned Llama 3. Bioengineering (Basel). 2024;11(10):1043. https://doi.org/10.3390/bioengineering11101043.

Wenderott K, Herfort J, Raspe M, Schwietering J, Busch HP, Reim M. Radiologists' perspectives on workflow integration of AI-based detection systems: a qualitative study. Appl Ergon. 2024;114:104140. https://doi.org/10.1016/j.apergo.2023.104140.

Kaviani P, Digumarthy SR, Bizzo BC, Dreyer KJ, Ebrahimian S. Artificial intelligence-generated smart impression from radiology

reports. medRxiv. 2024;:2024.03.07.24303787.
https://doi.org/10.1101/2024.03.07.24303787.