

ORIGINAL RESEARCH

Open access

Artificial Intelligence-Based Radiology Operations System for Prioritizing Critical Imaging Reports Using Free-Text Radiology Findings, Order Urgency, Patient Risk Profiles, and Historical Escalation Patterns

Ravi Kumar^{1*}, Neha Sharma¹, Aniket Deshmukh², Arjun Nair¹, Meera Pillai²

Abstract

Radiology reporting generates critical findings that require urgent communication across increasingly complex clinical workflows. Static worklist triage does not fully exploit report content, patient risk, order context, or prior escalation behaviour. Manual prioritisation and rule-based STAT flags may overlook subtle or unexpected abnormalities embedded in free-text reports. These approaches also fail to adapt to real-time patient vulnerability and institutional communication patterns. This article proposes an AI-based radiology operations framework that continuously extracts clinically actionable findings from narrative reports. The framework fuses those findings with order urgency, patient risk profiles, and historical escalation patterns to support dynamic worklist prioritisation. The system comprises an NLP report analysis module, an urgency-risk fusion engine, a historical escalation learner, a prioritisation decision engine, and a real-time notification dashboard. Each component is designed to support explainable, auditable, and workflow-sensitive prioritisation. The proposed system could reduce delays in acknowledging critical imaging results while balancing radiologist workload. Its value would depend on careful governance, transparent priority justifications, and human-in-the-loop feedback. An adaptive AI-based worklist system offers a pathway toward safer, data-driven radiology operations. By integrating narrative findings, clinical context, and historical escalation behaviour, such a framework could strengthen critical result management.

Keywords Artificial intelligence, Clinical natural language processing, Radiology operations, Critical results, Worklist prioritisation, Escalation learning

*Correspondence:

Ravi Kumar

ravi.kumar@outlook.com

¹ Department of Digital Healthcare Technologies, Faculty of Engineering, IIT Delhi, New Delhi, India

² Department of Clinical Informatics and Health Analytics, Faculty of Medicine, IIT Bombay, Mumbai, India

Introduction

Critical radiology findings require timely recognition, reporting, acknowledgment, and clinical action, yet communication remains vulnerable to interruptions, handoffs, competing workload, and variation in local escalation practice. AI-enabled radiology triage has shown conceptual relevance for urgent imaging conditions because it can help identify studies that warrant faster review or communication than routine queue order would provide [1-3]. Critical result management is therefore not only an image interpretation challenge but also an operations problem involving report generation, clinician

notification, and reliable acknowledgment [4, 5]. A system-oriented approach is needed because the safety risk arises at the intersection of imaging findings, reporting workflows, and downstream clinical response.

STAT ordering alone is an incomplete prioritisation mechanism because unexpected critical findings may appear in studies originally ordered as routine, outpatient, or low urgency. Free-text radiology reports contain nuanced descriptions of actionable findings, uncertainty, chronicity, and follow-up recommendations that are not always available to worklist logic or operational dashboards [6-8]. NLP-based extraction of report meaning could allow an operations system to identify clinically important signals after dictation and before communication delays accumulate [9, 10]. This makes narrative report content a key operational asset rather than merely a documentation endpoint.

AI methods could support critical report prioritisation by extracting actionable findings from free text and combining them with contextual signals that shape clinical urgency. Prior work on radiology NLP, incidental findings, follow-up recommendations, and abnormality classification suggests that language-based and image-adjacent AI methods can identify clinically meaningful content that would otherwise remain embedded in narrative documentation [11-15]. However, an AIF article should not treat prioritisation as a stand-alone model output; it should conceptualise prioritisation as a workflow decision informed by findings, patient risk, order context, and escalation history [16-18]. Historical escalation patterns are especially important because they reflect how institutions actually respond to critical findings under operational constraints.

This article proposes an AI-based radiology operations system that could dynamically prioritise imaging reports for radiologist review and clinical escalation based on free-text finding severity, order urgency, patient risk, and prior escalation patterns. The framework is positioned as an explainable decision-support system embedded within PACS, reporting, notification, and dashboard infrastructure rather than as an autonomous replacement for radiologist judgment. It draws on implementation principles from clinical AI, human-AI collaboration, and explainability to ensure that priority changes can be inspected, overridden, and audited. The central thesis is that critical report prioritisation should be adaptive, context-aware, and governed as a clinical operations system.

Background

Radiology workflow and critical result management

Radiology workflow begins with order entry and image acquisition, then proceeds through interpretation, dictation, report finalisation, notification, acknowledgment, and downstream clinical action. Bottlenecks may arise when urgent studies compete with high-volume routine work, when findings are unexpected relative to order urgency, or when communication depends on manual escalation outside the normal reporting system [1, 2, 19]. Critical result management is therefore susceptible to latent errors that may not originate in interpretation itself but in the operational transition from report content to accountable clinician response [4, 5]. An AI-based prioritisation framework should address these transition points by connecting report meaning to worklist order and notification pathways.

NLP for free-text radiology reports

Free-text radiology reports require NLP methods capable of recognising findings, anatomical locations, temporality, negation, uncertainty, and recommendations. Prior studies have shown that machine learning and language models can classify radiology reports, identify actionable language, detect incidental findings, and extract follow-up recommendations from narrative text [6-9]. These capabilities are essential because criticality is often expressed through phrases that require clinical interpretation, such as new haemorrhage, enlarging mass, possible pneumothorax, or recommendation for urgent follow-up [10, 20, 21]. For an operations system, NLP should function not merely as document classification but as a structured translation layer between narrative reporting and prioritisation logic.

Order urgency and patient risk as prioritisation signals

Order urgency provides an important but incomplete signal because STAT, urgent, and routine labels reflect ordering intent rather than the full clinical severity of the eventual finding. Patient context, including care setting, comorbid burden, previous imaging history, and vulnerability to deterioration, can modify the operational significance of the same radiologic finding [13–15]. AI systems designed for clinical radiology should therefore combine textual finding extraction with structured clinical context rather than relying on either signal in isolation [16, 17]. This integrated perspective supports a risk-aware priority estimate in which a finding is interpreted in relation to the patient's current clinical state.

Historical escalation patterns and clinician response

Historical escalation records can reveal which report types, service lines, times of day, or clinical contexts have previously required urgent notification or experienced delayed acknowledgment. Communication logs, critical result flags, and clinician feedback could provide a learning signal for identifying reports that are not only clinically important but also operationally vulnerable [4, 5, 22]. Such patterns should be interpreted cautiously because they may encode institutional habits, staffing constraints, and alert fatigue rather than ideal clinical practice [23, 24]. A historical escalation learner should therefore be governed as a reflective operations tool rather than as an unquestioned reproduction of past behaviour.

AI-driven worklist prioritisation in radiology operations

AI-driven worklist prioritisation has emerged as a promising approach for bringing urgent studies or reports to radiologist attention earlier than static queue ordering might allow. Prior work on chest radiograph triage, intracranial haemorrhage detection, pneumothorax detection, and abnormality classification illustrates how AI could support prioritisation when clinically significant findings require faster review [1–3, 16, 19]. Nevertheless, many existing systems focus on condition-specific detection or image-level triage rather than a broader, multimodal operations framework that integrates report text, order urgency, patient risk, and escalation behaviour [18, 25, 26]. The proposed AIF addresses this gap by positioning prioritisation as an adaptive decision structure rather than a single-task detection tool.

System Overview

High-level architecture

The proposed system begins by ingesting free-text radiology report content as it becomes available, then applies NLP to identify findings, uncertainty, recommendations, and potential criticality. These extracted signals are fused with order urgency, scan context, patient risk profile, and care setting before being passed to a historical escalation learner that estimates communication vulnerability [6, 7, 9, 13]. The prioritisation engine then produces an explainable priority state that can reorder the worklist, flag reports for review, and trigger notification components when escalation criteria are met [1, 2, 19]. This architecture treats priority as a continuously updated operational state rather than as a fixed property assigned at order entry.

Figure 1 illustrates the proposed AI-based radiology operations architecture for converting free-text report findings, order urgency, patient risk, and historical escalation behaviour into explainable worklist prioritisation and governed clinical notification support.

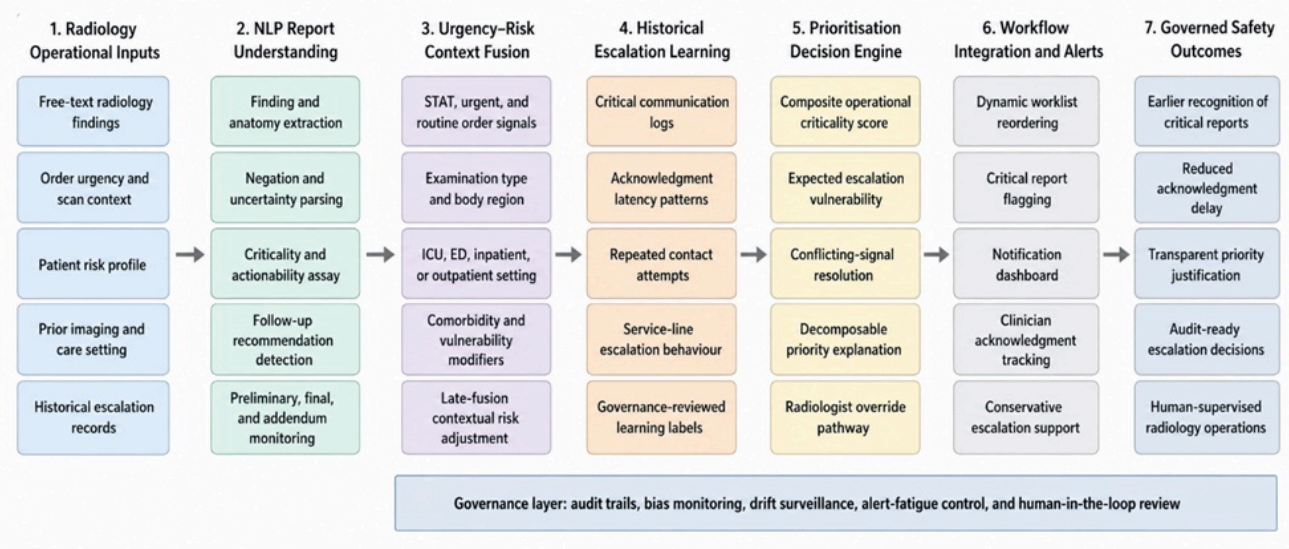


Figure 1. AI-Based Radiology Operations System for Context-Aware Critical Imaging Report Prioritisation

Core assumptions

The framework assumes that radiology report text is available soon after dictation or preliminary interpretation, that order-entry metadata can be accessed, and that patient risk features can be derived from clinical information systems. It also assumes that historical escalation logs, acknowledgment timestamps, or critical communication records exist in a form that can be linked to prior reports and workflow events [4, 5, 22]. These assumptions are realistic for many digitally mature radiology environments but may require interface development, terminology harmonisation, and governance agreements before deployment [23, 25]. The system should therefore be designed as a configurable framework rather than as a universally portable product.

Design principles

The system should be explainable, auditable, real-time, bias-aware, and embedded within normal radiology work rather than separated into an additional monitoring burden. Explainability is important because radiologists and clinicians need to understand whether a priority change was driven by the finding, the patient context, the order urgency, or a history of delayed escalation [24, 27]. Auditability is equally important because priority decisions may affect workload distribution, notification burden, and patient safety governance [22, 23]. The design should preserve human authority by allowing radiologist override, clinician acknowledgment, and continuous feedback into future system behaviour.

NLP for Free-Text Radiology Findings

Finding extraction and criticality assay

The NLP module should extract clinically meaningful finding phrases from report text and classify their potential criticality in relation to anatomy, acuity, and actionability. Transformer-based and machine learning approaches could be used conceptually to identify findings such as intracranial haemorrhage, pneumothorax, free air, pulmonary embolic concern, suspicious mass, or urgent follow-up recommendation [6, 9, 11]. The module should distinguish between acute and chronic descriptions, incidental and expected findings, and descriptive abnormalities that carry different escalation implications [10, 12, 15]. Its output should not be treated as a final clinical judgment but as a structured signal for downstream prioritisation.

Negation, uncertainty, and recommendation parsing

Radiology language contains negation, hedging, uncertainty, and conditional recommendations that can substantially alter operational urgency. Expressions such as “no evidence of,” “cannot exclude,” “possible,” “new since prior,” or “recommend urgent follow-up” should modify how a finding contributes to prioritisation [7, 8, 20]. Recommendation parsing is particularly important because follow-up language often identifies actionable abnormalities that require communication even when the report is not framed as an emergency [21, 28]. The system should therefore represent uncertainty explicitly rather than forcing every textual signal into a binary critical or non-critical category.

Continuous report monitoring and incremental updating

Radiology reports may change between preliminary dictation, attending finalisation, addendum creation, and corrected report release. The NLP module should re-evaluate report text whenever the report changes so that the priority state reflects the most current language and does not remain anchored to an earlier draft [5, 6, 9]. Incremental monitoring would be particularly relevant when a later amendment clarifies uncertainty, adds a critical finding, or changes a recommendation from routine follow-up to urgent clinical action [7, 8]. This design supports a living operational representation of report risk rather than a one-time text classification event.

Order Urgency and Patient Risk Profiling

Incorporating order urgency and scan context

The urgency-risk fusion engine should ingest order priority, examination type, body region, clinical indication, care setting, and timing information from the imaging requisition and workflow system. These structured signals are useful because a STAT neuroimaging order from the emergency department carries a different baseline operational expectation than a routine outpatient follow-up examination [2, 3, 19]. However, the system should avoid treating order urgency as determinative because serious findings may be discovered incidentally on routine studies and low-acuity labels may reflect incomplete information at ordering time [10, 15, 16]. Order context should therefore shape, but not dominate, the prioritisation logic.

Patient risk profile construction

Patient risk profiles should integrate clinically relevant vulnerability signals such as age, acuity setting, comorbidities, recent procedures, intensive care status, and prior critical imaging history. These variables could help determine whether the same textual finding should be prioritised differently across patients with different probabilities of deterioration or different consequences of delay [13, 14, 17]. For example, a small abnormality may warrant higher operational priority when it appears in a patient with substantial clinical vulnerability or a history of rapidly evolving disease. The profile should remain transparent and clinically interpretable so that users can see why patient context influenced the priority assignment [23, 24].

Fusion of structured context with NLP-derived findings

The system should use a late-fusion approach in which NLP-derived finding severity is combined with structured order context and patient risk after each signal has been independently represented. This approach would allow the engine to preserve the semantic meaning of the report while still adjusting priority according to patient vulnerability and operational urgency [6, 9, 16]. A finding such as a possible pneumothorax could receive different priority treatment depending on whether the patient is ventilated, postoperative, outpatient, or already under close inpatient monitoring [3, 15]. The goal is not to replace clinical judgment but to produce an explainable operational recommendation that reflects both textual severity and contextual risk.

Table 1 defines how each signal domain contributes distinct operational value to critical report prioritisation while introducing separate interpretability and governance requirements.

Table 1. Signal-Level Contribution of Multimodal Inputs to Critical Imaging Report Prioritisation

Input signal domain	Operational meaning	Prioritisation contribution	Main interpretability requirement	Key governance risk
Free-text report findings	Narrative evidence of abnormality, acuity, uncertainty, and recommendation language	Identifies clinically actionable content that may not be captured by order status alone	Show extracted phrases, finding category, negation status, and uncertainty level	NLP misclassification or over-weighting ambiguous language
Order urgency and scan context	Ordering intent, modality, body region, indication, and timing	Provides baseline operational expectation for review speed	Display whether priority was influenced by STAT, urgent, or routine status	Treating order urgency as determinative despite unexpected findings
Patient risk profile	Vulnerability based on care setting, acuity, comorbidity, prior imaging history, and deterioration risk	Modifies urgency according to patient-specific consequences of delay	Show which patient-context factors increased or decreased priority	Unequal prioritisation if risk variables encode biased care patterns
Historical escalation behaviour	Prior communication, acknowledgment, repeated contact attempts, and delay patterns	Estimates operational vulnerability and likelihood of delayed response	Display comparable historical escalation features without exposing irrelevant details	Reproducing local habits, staffing constraints, or documentation bias
Real-time workflow state	Current report status, addenda, acknowledgment status, and worklist burden	Updates priority dynamically as report language or communication status changes	Show timestamped reason for priority change	Alert fatigue, excessive reordering, or workflow disruption
Human override and feedback	Radiologist acceptance, downgrade, upgrade, annotation, or correction	Provides supervised correction and governance review signal	Preserve override reason and user role in audit trail	Mistaking disagreement for error or reinforcing individual preferences

Historical Escalation Pattern Learning

Labelling critical escalations from retrospective data

Historical escalation learning would begin by linking prior reports to communication logs, notification records, acknowledgment timestamps, radiologist escalation actions, and clinician feedback. These linked records could help identify which report types were considered critical in practice, which communications required repeated contact attempts, and which cases experienced potentially delayed acknowledgment [4, 5, 22]. Labels should be constructed with governance oversight because historical escalation data may reflect local habits, missing documentation, or inconsistent thresholds for direct communication [23, 24]. The resulting labels should support operational learning while remaining open to review and correction by radiology leadership and clinical stakeholders.

Modeling escalation likelihood and latency

The escalation learner would conceptually estimate whether a report is likely to require critical communication and whether acknowledgment may be delayed under current workflow conditions. Inputs could include NLP-derived finding categories, care setting, order urgency, patient risk, time of day, service line, and prior escalation behaviour associated with similar reports [5, 7, 8]. The output should be used to rank operational vulnerability rather than to claim deterministic prediction of clinician response [22, 23]. Such a learner could help the prioritisation engine elevate reports that combine clinically serious content with historically fragile communication pathways.

Detecting shifts in escalation culture and feedback adaptation

Escalation culture may change as new clinical policies, staffing models, notification tools, or radiologist behaviours alter how critical findings are communicated. The system should therefore monitor feedback and overrides so that it can adapt to evolving local practice without reinforcing outdated or biased escalation patterns [24–26]. Human-in-the-loop feedback is essential because radiologists may identify over-prioritised reports, under-recognised critical patterns, or alert categories that create unnecessary burden [27, 29]. Continuous adaptation should be paired with audit trails so that changes in prioritisation logic remain explainable and accountable.

Prioritization Engine and Decision Logic

Criticality score construction

The prioritisation engine should construct a composite criticality score that combines finding severity, patient vulnerability, order context, and expected escalation latency. This score would not represent diagnostic certainty alone but rather operational urgency: the degree to which a report should move upward in the queue or trigger additional communication support [1, 2, 19]. Findings associated with severe or time-sensitive clinical consequences should contribute more strongly when paired with high-risk patient profiles or historically delayed acknowledgment pathways [4, 5]. The score should remain decomposable so users can inspect whether prioritisation was driven primarily by report text, patient risk, order urgency, or escalation history [24, 27].

Combining NLP risk with structured urgency and history

The decision engine should fuse NLP-derived critical finding signals with structured order urgency, patient risk features, and historical escalation patterns. A report containing language suggestive of an actionable abnormality could be prioritised differently depending on whether it comes from an emergency, inpatient, ICU, or outpatient context [3, 13, 14]. Historical communication patterns should then adjust the priority state when similar reports have previously required escalation or experienced delayed clinician acknowledgment [5, 22]. This layered logic would allow the system to move beyond STAT-only ordering and toward context-sensitive prioritisation [16, 17].

Handling uncertainty and conflicting signals

The system should explicitly handle uncertainty rather than suppress it, because radiology language often contains hedging, negation, and conditional recommendations. When a report includes uncertain language such as “cannot exclude” but the patient is high risk, the system could elevate priority while showing the uncertainty as part of the explanation [7, 8, 20]. Conversely, a clearly abnormal phrase in a clinically stable outpatient context may require review without automatic high-intensity escalation [10, 15]. Clinician override should be built into the decision logic so that radiologists can correct priority states and help the system learn from disputed cases [23, 27].

Real-Time Alerting and Clinical Workflow Integration

Dynamic worklist reordering and visual prioritisation

The system should update the radiology worklist whenever new report text, order metadata, patient risk information, or escalation status becomes available. Priority changes should be displayed through clear worklist indicators, concise explanations, and access to the factors that drove the recommendation [1, 19, 25]. Visual prioritisation should support radiologist attention without creating unnecessary interruption or alarm fatigue, especially when multiple AI tools operate within the same clinical environment [26, 29]. The dashboard should therefore function as a workflow aid rather than as a competing parallel queue.

Alert escalation workflow for critical findings

When the composite priority state exceeds a critical threshold and acknowledgment is absent, the system could recommend or initiate escalation through institutionally approved notification pathways. Such pathways may include ordering clinicians, covering teams, emergency department contacts, or intensive care staff depending on local policy and patient location [4, 5]. The alert should include a concise explanation of the finding, the relevant patient-risk modifier, and the reason the system considers the case operationally vulnerable [22, 24]. Escalation design should be conservative, auditable, and adjustable to prevent excessive alerts that undermine trust [23, 27].

System Governance and Feedback Loops

Radiologist override and human-in-the-loop feedback

Radiologists should be able to accept, downgrade, upgrade, or annotate system-generated priority states directly within the worklist interface. Each action should be logged as feedback, not as an error by default, because disagreement may reflect local context, incomplete data, or appropriate clinical judgment [23, 29]. Human-in-the-loop governance is especially important for a system that combines NLP interpretation, patient context, and historical escalation behaviour [24, 27]. Feedback should be reviewed periodically so that model updates reflect expert oversight rather than automatic reinforcement of operational habits.

Monitoring for bias and performance drift

The system should monitor whether prioritisation patterns differ across patient groups, care settings, imaging modalities, shift times, and service lines. Such monitoring is necessary because historical escalation logs may encode unequal communication practices, documentation quality, staffing patterns, or access to rapid acknowledgment [23, 25]. Drift surveillance should also examine whether report style, terminology, ordering behaviour, or clinical workflows change over time in ways that weaken the system's assumptions [11, 18]. Governance dashboards should therefore track not only technical behaviour but also fairness, usability, and institutional safety impact [22, 24].

Evaluation Strategy

Predictive and prioritisation performance

Evaluation should begin with retrospective simulation of worklist reordering using historical reports, order metadata, patient context, and escalation records. The goal would be to compare conceptual prioritisation behaviour against standard queue order and STAT-only triage without presenting the framework as a completed clinical performance study [1, 2, 19]. Evaluation should ask whether the system would be expected to bring clinically important reports to attention earlier and whether explanations align with expert review [5, 6, 9]. This phase should remain exploratory and safety-oriented rather than framed as proof of effectiveness.

Workflow and usability metrics

Workflow evaluation should examine whether radiologists understand the priority explanations, whether recommendations fit naturally into existing reading practices, and whether alerts increase or reduce cognitive burden. Usability assessment should include radiologist trust, perceived appropriateness of priority shifts, override patterns, and alert fatigue risk [26, 27, 29]. Because clinical AI often fails when it is technically plausible but operationally misaligned, implementation assessment should be treated as central rather than secondary [22, 23]. The evaluation should therefore focus on interaction quality, transparency, and workflow compatibility.

Prospective silent-mode and live pilot evaluation

A prospective silent-mode deployment would allow the system to generate priorities in parallel with normal operations without affecting clinical workflow. This stage should help determine whether the system identifies plausible critical reports, produces interpretable explanations, and behaves consistently across care settings and report types [16-18]. After governance review, a limited live pilot could test whether priority display and escalation support are acceptable to radiologists and downstream clinicians [22, 25]. Any live use should include monitoring, override capacity, and predefined procedures for pausing or modifying the system [23, 24].

Table 2 presents a staged evaluation matrix linking technical performance, workflow usability, escalation reliability, and governance safeguards for prospective assessment of the proposed system.

Table 2. Evaluation Matrix for an Explainable AI Radiology Operations Prioritisation System

Evaluation dimension	Primary question	Suggested assessment approach	Success indicator	Safety concern addressed
NLP validity	Does the system correctly extract clinically meaningful findings, uncertainty, negation, and recommendations?	Expert-reviewed comparison of extracted findings against report text	High agreement with radiologist-coded report meaning	False prioritisation from language misunderstanding
Prioritisation performance	Would the system bring critical or vulnerable reports forward earlier than static queue order?	Retrospective simulation against chronological and STAT-only workflows	Earlier ranking of cases requiring urgent acknowledgment	Delays caused by routine queue placement
Explanation quality	Can users understand why a report was elevated or downgraded?	Radiologist review of decomposed priority explanations	Priority rationale judged clinically plausible and inspectable	Black-box priority changes
Workflow fit	Does the system support radiologist attention without increasing cognitive burden?	Silent-mode review, usability testing, and live-pilot observation	Low friction, acceptable display design, manageable alert volume	Alert fatigue and workflow disruption
Escalation reliability	Does the system identify cases at risk for delayed acknowledgment?	Analysis of acknowledgment timestamps and contact-attempt history	Improved detection of communication-vulnerable reports	Missed or delayed critical-result communication
Override behaviour	Are human corrections meaningful and governable?	Review of upgrade, downgrade, and annotation patterns	Overrides reveal clinically useful	Unsafe automation or inappropriate model authority

			refinements rather than systematic mistrust	
Bias and drift	Does prioritisation remain fair and stable across patient groups, services, shifts, and report styles?	Stratified monitoring across demographics, care settings, modalities, and time periods	No unexplained disparity or performance degradation	Encoded inequity and temporal model decay
Prospective readiness	Is the system safe enough for limited live deployment?	Silent-mode deployment followed by governance review	Stable explanations, acceptable alert burden, clear pause criteria	Premature clinical implementation

Limitations

Report availability latency and NLP limitations

A report-based prioritisation system depends on the availability and quality of dictated or preliminary text, which may lag behind image acquisition. NLP may also misclassify ambiguous language, fail to capture nuanced context, or overreact to uncertain phrasing when clinical interpretation would be more cautious [8, 11, 20]. Recommendation extraction and incidental finding detection are especially vulnerable to local reporting style and incomplete documentation [7, 10, 21]. The system should therefore be viewed as an operations support layer, not as a substitute for radiologist interpretation or clinical communication judgment.

Generalizability and site-specific customisation

Escalation practices, report templates, ordering behaviour, patient populations, and notification policies vary substantially across institutions. A model that appears operationally sensible in one environment may require local adaptation before use in another because historical escalation behaviour is partly a product of institutional culture [25, 26, 28]. Commercial radiology AI tools and clinical AI implementations also show that integration, evidence, governance, and workflow fit are central to real-world value [22, 23]. The framework should therefore be locally configurable, prospectively evaluated, and continuously audited before being treated as a deployable operational standard.

Conclusion

An AI-based radiology operations system for prioritising critical imaging reports could support safer and more adaptive worklist management. By integrating report language, order urgency, patient risk, and historical escalation behaviour, the system would shift prioritisation from static queue order to context-aware operational decision support.

The framework's main strength is its integration of multiple signals that are usually handled separately in radiology operations. Real-time NLP of free-text findings, structured clinical context, escalation learning, and human feedback could jointly create a more explainable and responsive prioritisation environment.

Important challenges remain, including report latency, NLP uncertainty, local variation in escalation practice, and the risk of alert fatigue. The system would require careful governance, transparent explanations, and extensive clinical workflow testing before live use.

Future work should pursue multi-institutional silent-mode pilots and shared evaluation standards for radiology worklist AI. Standardised benchmarks would help clarify how adaptive prioritisation systems should be assessed, governed, and

responsibly integrated into clinical radiology operations.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 15 Sep 2021

Revised: 14 Oct 2021

Accepted: 26 Nov 2021

Published online: 25 February 2022

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

Annarumma M, Withey SJ, Bakewell RJ, Pesce E, Goh V, Montana G. Automated triaging of adult chest radiographs with deep artificial neural networks. *Radiology*. 2019;291(1):196-202.

<https://doi.org/10.1148/radiol.2018180922>.

Ginat DT. Implementation of machine learning software on the radiology worklist decreases scan view delay for the detection of intracranial hemorrhage on CT. *Brain Sci*. 2021;11(7):832.

<https://doi.org/10.3390/brainsci11070832>.

Hong W, Hwang EJ, Lee JH, Park J, Goo JM, Park CM. Deep learning for detecting pneumothorax on chest radiographs after needle biopsy: clinical implementation. *Radiology*. 2022;303(2):433-41.

<https://doi.org/10.1148/radiol.211860>.

Heilbrun ME, Chapman BE, Narasimhan E, Patel N, Mowery D. Feasibility of natural language processing–assisted auditing of critical findings in chest radiology. *J Am Coll Radiol*. 2019;16(9):1299-304.

<https://doi.org/10.1016/j.jacr.2019.04.013>.

Lauriola I, Lavelli A, Aiolfi F. An introduction to deep learning in natural language processing: models, techniques, and tools. *Neurocomputing*. 2022;470:443-56.
<https://doi.org/10.1016/j.neucom.2021.05.103>.

Chen MC, Ball RL, Yang L, Moradzadeh N, Chapman BE, Larson DB, et al. Deep learning to classify radiology free-text reports. *Radiology*. 2018;286(3):845-52.
<https://doi.org/10.1148/radiol.2017171111>.

Carrodegua E, Lacson R, Swanson W, Khorasani R. Use of machine learning to identify follow-up recommendations in radiology reports. *J Am Coll Radiol*. 2019;16(3):336-43.
<https://doi.org/10.1016/j.jacr.2018.09.035>.

Lou R, Lalevic D, Chambers C, Zafar HM, Cook TS. Automated detection of radiology reports that require follow-up imaging using natural language processing feature engineering and machine learning classification. *J Digit Imaging*. 2020;33(1):131-6.
<https://doi.org/10.1007/s10278-019-00251-7>.

Nakamura Y, Hanaoka S, Nomura Y, Nakao T, Miki S, Watadani T, et al. Automatic detection of actionable radiology reports using bidirectional encoder representations from transformers. *BMC Med Inform Decis Mak*. 2021;21(1):262.
<https://doi.org/10.1186/s12911-021-01615-6>.

Trivedi G, Dadashzadeh ER, Handzel RM, Chapman WW, Visweswaran S, Hochheiser H. Interactive NLP in clinical care: identifying incidental findings in radiology reports. *Appl Clin Inform*. 2019;10(4):655-69.
<https://doi.org/10.1055/s-0039-1693654>.

Sorin V, Barash Y, Konen E, Klang E. Deep learning for natural language processing in radiology—fundamentals and a systematic review. *J Am Coll Radiol*. 2020;17(5):639-48.
<https://doi.org/10.1016/j.jacr.2019.10.007>.

Trivedi G, Hong C, Dadashzadeh ER, Handzel RM, Hochheiser H, Visweswaran S. Identifying incidental findings from radiology reports of trauma patients: an evaluation of automated feature representation methods. *Int J Med Inform*. 2019;129:81-7.
<https://doi.org/10.1016/j.ijmedinf.2019.05.009>.

Fu S, Leung LY, Wang Y, Raulli AO, Kallmes DF, Kinsman KA, et al. Natural language processing for the identification of silent brain infarcts from neuroimaging reports. *JMIR Med Inform*. 2019;7(2):e12109.
<https://doi.org/10.2196/12109>.

Kehl KL, Elmarakeby H, Nishino M, Van Allen EM, Lepisto EM, Hassett MJ, et al. Assessment of deep natural language processing in ascertaining oncologic outcomes from radiology reports. *JAMA Oncol*. 2019;5(10):1421-9.
<https://doi.org/10.1001/jamaoncol.2019.1800>.

Kang SK, Garry K, Chung R, Moore WH, Iturrate E, Swartz JL, et al. Natural language processing for identification of incidental pulmonary nodules in radiology reports. *J Am Coll Radiol*. 2019;16(11):1587-94.
<https://doi.org/10.1016/j.jacr.2019.06.025>.

Tang YX, Tang YB, Peng Y, Yan K, Bagheri M, Redd BA, et al. Automated abnormality classification of chest radiographs using deep convolutional neural networks. *NPJ Digit Med*. 2020;3:70.
<https://doi.org/10.1038/s41746-020-0273-1>.

Nabulsi Z, Sellergren A, Jamsky S, Lau C, Santos E, Kiraly AP, et al. Deep learning for distinguishing normal versus abnormal chest radiographs and generalization to two unseen diseases tuberculosis and COVID-19. *Sci Rep*. 2021;11(1):15523.
<https://doi.org/10.1038/s41598-021-94771-7>.

Seah JC, Tang CH, Buchlak QD, Holt XG, Wardman JB, Aimoldin A, et al. Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. *Lancet Digit Health*.

2021;3(8):e496-e506.

[https://doi.org/10.1016/S2589-7500\(21\)00092-0](https://doi.org/10.1016/S2589-7500(21)00092-0).

Baltruschat I, Steinmeister L, Nickisch H, Saalbach A, Grass M, Adam G, et al. Smart chest X-ray worklist prioritization using artificial intelligence: a clinical workflow simulation. *Eur Radiol.* 2021;31(6):3837-45.

<https://doi.org/10.1007/s00330-020-07573-9>.

Callen AL, Dupont SM, Price A, Laguna B, McCoy D, Do B, et al. Between always and never: evaluating uncertainty in radiology reports using natural language processing. *J Digit Imaging.* 2020;33(5):1194-201.

<https://doi.org/10.1007/s10278-020-00352-8>.

Bozkurt S, Alkim E, Banerjee I, Rubin DL. Automated detection of measurements and their descriptors in radiology reports using a hybrid natural language processing algorithm. *J Digit Imaging.* 2019;32(4):544-53.

<https://doi.org/10.1007/s10278-018-0154-3>.

Sendak MP, Ratliff W, Sarro D, Alderton E, Futoma J, Gao M, et al. Real-world integration of a sepsis deep learning technology into routine clinical care: implementation study. *JMIR Med Inform.* 2020;8(7):e15182.

<https://doi.org/10.2196/15182>.

Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* 2019;17(1):195.

<https://doi.org/10.1186/s12916-019-1426-2>.

Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q Consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inform Decis Mak.* 2020;20(1):310.

<https://doi.org/10.1186/s12911-020-01332-8>.

van Leeuwen KG, Schalekamp S, Rutten MJ, van Ginneken B, de Rooij M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *Eur Radiol.* 2021;31(6):3797-804.

<https://doi.org/10.1007/s00330-020-07464-z>.

Liew C. The future of radiology augmented with artificial intelligence: a strategy for success. *Eur J Radiol.* 2018;102:152-6.

<https://doi.org/10.1016/j.ejrad.2018.03.019>.

Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019;25(1):44-56.

<https://doi.org/10.1038/s41591-018-0300-7>.

Nobel JM, Puts S, Bakers FC, Robben SG, Dekker AL. Natural language processing in Dutch free text radiology reports: challenges in a small language area staging pulmonary oncology. *J Digit Imaging.* 2020;33(4):1002-8.

<https://doi.org/10.1007/s10278-019-00280-2>.

Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJ. Artificial intelligence in radiology. *Nat Rev Cancer.* 2018;18(8):500-10.

<https://doi.org/10.1038/s41568-018-0016-5>.