

ORIGINAL RESEARCH

Open access

Retrieval-Augmented Generation System for Supporting Hospital Accreditation Preparation Using Policy Documents, Audit Findings, Quality Indicators, Regulatory Standards, and Departmental Evidence Files

Isabella Garcia^{1*}, Diego Herrera¹

Abstract

Hospital accreditation requires organizations to demonstrate that clinical, operational, safety, and governance processes are aligned with externally defined standards. This demonstration depends on extensive internal documentation, including current policies, procedure manuals, audit reports, quality dashboards, meeting records, training logs, and department-level evidence files. Accreditation preparation is often conducted through manual document searches across fragmented repositories. This process is slow, resource-intensive, and vulnerable to missed evidence, outdated documents, inconsistent interpretations, and duplication of staff effort. This article proposes a retrieval-augmented generation system designed to support hospital accreditation preparation without replacing human judgment. The system indexes internal accreditation-relevant documents and allows authorized users to ask natural-language questions that receive synthesized, citation-backed responses. The proposed system includes a multi-source ingestion pipeline, a metadata-rich vector database, a permissioned large language model, a citation-grounding layer, and a human review dashboard. Together, these components would support evidence retrieval, gap identification, traceability, and accreditation team collaboration. The system could assist accreditation teams by reducing time spent locating and cross-referencing evidence. It would also be expected to improve the completeness and consistency of preparation materials by grounding responses in authoritative internal sources. A specialized retrieval-augmented generation system could help hospitals move from episodic accreditation preparation toward continuous audit readiness. Its value would depend on document quality, governance safeguards, human review, and rigorous evaluation in real accreditation workflows.

Keywords Large language models, Retrieval-augmented generation, Regulatory compliance, Hospital accreditation, Healthcare quality, Evidence synthesis

*Correspondence:

Isabella Garcia

isabella.garcia@gmail.com

¹ Department of Digital Healthcare Analytics, Faculty of Medicine, University of Buenos Aires, Buenos Aires, Argentina

Introduction

Hospital accreditation is a strategic mechanism through which healthcare organizations demonstrate compliance with safety, quality, governance, and clinical performance expectations. Preparation for accreditation requires hospitals to assemble large bodies of documentary evidence that connect standards to policies, audit findings, quality indicators, staff

training records, and departmental practice. Systematic reviews of hospital accreditation have emphasized its relationship to quality improvement while also showing that accreditation is embedded in complex organizational processes rather than isolated inspection events [1, 2]. For this reason, accreditation readiness depends not only on the existence of policies, but also on the ability to retrieve, interpret, and connect evidence across the institution.

The current preparation process is often fragmented because the required evidence is distributed across policy management systems, shared drives, quality dashboards, audit archives, departmental folders, and individual staff knowledge. Studies on registration and quality monitoring burden indicate that documentation work can become a substantial operational load, particularly when staff must repeatedly collect, validate, and reformat information for external reporting or internal governance [3, 4]. Commentary on governance complexity further suggests that the perceived and real burden of documentation can increase when multiple oversight systems require similar but non-identical evidence [5]. In accreditation preparation, these dynamics can turn evidence assembly into a laborious search-and-reconciliation exercise rather than a focused quality review.

Retrieval-augmented generation offers a promising architecture for institutions that need trustworthy question answering over large, specialized document collections. Recent healthcare RAG studies and reviews show how retrieval systems can ground language model responses in selected source documents, making them more suitable for clinical, educational, and organizational contexts than unconstrained generation alone [6-9]. Medical RAG applications have been explored for laboratory regulations, radiology, liver disease, nephrology, nursing, and medical question answering, indicating that the architecture can be adapted to domain-specific knowledge bases [10-21]. Although accreditation preparation has distinctive governance requirements, these prior applications support the conceptual feasibility of using RAG for high-volume, high-stakes document synthesis.

This article proposes an Emerging AI system for hospital accreditation preparation that unifies policy documents, audit findings, quality indicators, regulatory standards, and departmental evidence files into a single queryable knowledge base. The system is not presented as an autonomous compliance engine, but as a governed decision-support layer that could assist accreditation coordinators, quality officers, department leaders, and survey preparation teams. Its central premise is that RAG can make institutional evidence more discoverable, interpretable, and traceable when combined with strict citation grounding, role-based access, and mandatory human review. The intended contribution is a conceptual system design that frames accreditation preparation as a retrieval, synthesis, and governance problem rather than merely a document management problem.

Background

Hospital accreditation and documentation requirements

Hospital accreditation systems evaluate whether healthcare organizations meet defined standards related to patient safety, clinical governance, infection prevention, medication management, emergency preparedness, quality improvement, and leadership accountability. The evidence submission process typically requires hospitals to demonstrate that formal policies are current, that practice is monitored through audits and indicators, and that corrective actions are documented. Reviews of accreditation show that its impact depends on how organizations integrate standards into daily quality systems rather than treating surveys as episodic events [1, 2]. Therefore, a digital system for accreditation support should be designed around evidence continuity, version control, and traceable alignment between standards and institutional documents.

The documentation burden and current evidence assembly

Evidence assembly remains burdensome because hospitals often maintain relevant materials in separate systems that were not designed around accreditation logic. A single accreditation question may require locating a policy in one repository, a related audit in another, a quality indicator trend in a dashboard, and meeting minutes in a departmental

folder. Research on quality registration burden shows that documentation tasks can consume professional attention and may affect staff perceptions of administrative workload and joy in work [3, 4]. A RAG system could assist by reducing repetitive searching, but it would still depend on the completeness and reliability of the underlying document ecosystem.

Retrieval-Augmented Generation (RAG) – core concepts

Retrieval-augmented generation combines a retriever that searches a defined corpus with a language model that generates an answer based on retrieved evidence. In healthcare applications, this architecture is attractive because it can ground responses in institutional or domain-specific texts rather than relying only on a model's internal parameters [7-9]. Reviews of biomedical RAG emphasize that retrieval, generation, and evaluation must be considered together, particularly when answers are used in settings where factual accuracy and source traceability matter [9, 21]. For accreditation preparation, the relevant corpus would be local institutional evidence and applicable standards rather than general web or textbook knowledge.

Prior NLP and AI applications in hospital compliance and quality

Prior NLP and AI work in healthcare has addressed question answering, document retrieval, summarization, guideline interpretation, and extraction of information from electronic records. RAG studies have demonstrated conceptual utility in answering medical questions, interpreting clinical regulations, summarizing electronic health record content, and supporting specialized biomedical education [6, 7, 14, 22]. Embedding-based retrieval research in electronic health records also indicates that retrieval performance depends on model choice, pooling strategy, and the representational fit between queries and documents [23]. These findings are relevant to accreditation because policy language, audit terminology, and regulatory standards often require precise retrieval rather than broad semantic similarity alone.

Challenges of AI in high-stakes governance contexts

AI systems used for hospital governance must address risks related to hallucination, misinterpretation, privacy, accountability, and overreliance. Ethical and safety analyses of healthcare AI emphasize that accountability cannot be delegated to automated systems, particularly when outputs influence clinical or organizational decisions [24, 25]. Privacy scholarship also highlights the need to protect health information and institutional data when AI systems process sensitive internal documents [26, 27]. For an accreditation RAG system, these concerns require strict access controls, audit logs, source-grounding validation, and human approval before any generated material is used in formal submissions.

System Overview

High-level architecture

The proposed architecture begins with a secure ingestion pipeline that extracts text and metadata from accreditation-relevant documents, chunks them into retrievable units, embeds them, and stores them in a vector database. When a user submits a query, the system embeds the query, retrieves relevant chunks, optionally re-ranks them, and passes only the selected evidence to a permissioned language model for synthesis. Healthcare RAG studies show that retrieval design, context construction, and grounding constraints are central to producing useful responses in specialized domains [8, 12, 15, 18]. In the accreditation setting, the generated answer would include inline citations to source chunks, explicit uncertainty when evidence is insufficient, and links back to original documents for review.

Figure 1 illustrates how accreditation-relevant institutional documents can be converted into a permissioned, citation-grounded, human-reviewed RAG workflow for continuous audit readiness.

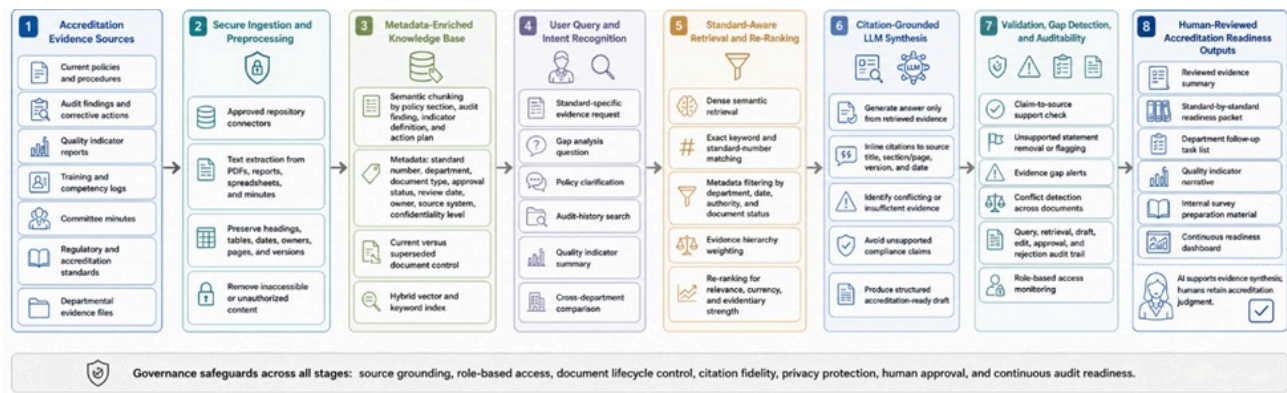


Figure 1. Governed Retrieval-Augmented Generation Workflow for Hospital Accreditation Preparation.

Core document sources

The system would index several classes of institutional evidence: clinical policies, standard operating procedures, internal audit findings, external survey reports, quality indicator dashboards, trend reports, training logs, committee minutes, equipment checks, and departmental evidence folders. It would also include regulatory and accreditation standards used by the organization, such as standards from accrediting bodies, government payers, or state authorities, subject to licensing and access requirements. RAG applications in laboratory regulation and specialty-specific medical domains suggest that narrow, authoritative corpora are especially important when the system must answer questions about rules, procedures, and evidence requirements [14, 16, 17]. For hospital accreditation, the corpus must therefore distinguish approved policies from drafts, current documents from retired versions, and organization-wide requirements from department-specific artifacts.

Design principles

The design principles are exhaustive source grounding, full auditability, role-based access, human-centered workflow integration, and mandatory review before external use. These principles reflect the high-stakes nature of accreditation, where an apparently fluent but unsupported answer could misrepresent compliance status or obscure a real safety gap. Reviews of healthcare RAG and broader AI safety literature indicate that grounding, validation, and human oversight are essential when systems are used in sensitive clinical or governance contexts [21, 24, 25]. Accordingly, the proposed system would operate as an evidence navigation and drafting assistant rather than as an independent compliance authority.

Knowledge Base Construction – Document Ingestion and Indexing

Document harvesting and preprocessing

Document harvesting would use secure connectors to policy libraries, shared drives, quality databases, audit archives, and departmental repositories approved for accreditation work. Preprocessing would extract text from PDFs, word-processing files, spreadsheets, and reports while preserving headings, tables, dates, document owners, and page references needed for citation fidelity. Work on clinical information extraction and RAG-based summarization from electronic health records illustrates the importance of transforming heterogeneous source materials into structured, retrievable representations before generation occurs [7]. In accreditation preparation, preprocessing should also preserve the difference between narrative evidence, tabular quality indicators, policy requirements, and corrective action records.

Chunking strategy and metadata enrichment

The chunking strategy should divide documents into semantically meaningful units such as policy sections, audit finding summaries, corrective action descriptions, indicator definitions, and meeting-minute decisions. Each chunk would receive metadata for standard number, department, document type, approval status, effective date, review date, owner, source system, and confidentiality level. Medical RAG research on long-context retrieval and specialized question answering suggests that retrieval quality depends on providing the generator with context that is specific enough to answer the question but broad enough to preserve meaning [8, 18]. Metadata enrichment would therefore allow the system to filter for current policies, prioritize authoritative evidence, and separate hospital-wide standards from local departmental procedures.

Table 1 defines the accreditation knowledge architecture required for a RAG system to retrieve evidence according to source authority, document currency, metadata integrity, and governance relevance.

Table 1. Accreditation RAG Knowledge Architecture: From Institutional Evidence Sources to Governed Retrieval Units

Knowledge layer	Accreditation-specific content	Metadata required for safe retrieval	Retrieval role in the RAG system	Governance value added
Regulatory and accreditation standards	External standards, survey requirements, interpretive guidance, government or payer requirements	Standard number, issuing body, effective date, version, licensing restrictions, applicability scope	Anchors user queries to specific compliance expectations and supports standard-aware retrieval	Prevents generic answers by tying evidence synthesis to the exact requirement being examined
Approved policies and procedures	Hospital-wide policies, clinical procedures, administrative protocols, emergency preparedness plans	Policy owner, approval status, effective date, review date, version, department, retired/superseded status	Provides formal institutional intent and documented process evidence	Distinguishes current authoritative documents from drafts, outdated policies, or local informal practices
Internal audit findings	Audit reports, tracer findings, compliance checks, corrective action records, closure documentation	Audit date, department, finding type, severity, responsible owner, corrective action status, closure evidence	Retrieves evidence of monitoring, nonconformance detection, and follow-up action	Shows whether policies are operationally monitored rather than merely stored
Quality indicator evidence	Dashboards, metric definitions, trend reports, performance summaries, improvement plans	Indicator name, numerator/denominator, reporting period, responsible committee, benchmark, action plan linkage	Connects accreditation standards to measurable quality performance and improvement activity	Supports readiness narratives that show monitoring, review, and action rather than isolated metric reporting
Training and competency records	Staff training logs, competency checklists, orientation materials, annual education records	Staff group, training topic, completion period, department, trainer, competency validation method	Retrieves evidence that required practices are communicated and reinforced	Helps identify whether a policy requirement is supported by workforce training evidence

Committee and governance records	Meeting minutes, board reports, quality committee reviews, leadership decisions	Committee name, meeting date, agenda item, decision, responsible party, follow-up status	Links operational evidence to oversight and leadership accountability	Demonstrates that quality and compliance information is reviewed through governance structures
Departmental evidence files	Unit-specific checklists, local audits, equipment checks, readiness binders, service-line documentation	Department, source folder, document owner, evidence type, date range, confidentiality level	Supplies localized proof for standards that require department-level demonstration	Prevents hospital-wide policies from being mistaken as sufficient evidence of local implementation
Generated response records	RAG-generated summaries, cited answer drafts, reviewer edits, rejected outputs, approved readiness materials	Query, retrieved chunks, citation map, reviewer identity, edit history, approval status, timestamp	Creates a traceable record of how evidence was assembled and reviewed	Makes the preparation process auditable and accountable rather than opaque

Embedding and vector store

Each chunk would be embedded using a domain-suitable representation model and indexed in a vector store that supports semantic retrieval, keyword matching, metadata filtering, and re-ranking. For accreditation questions, hybrid retrieval is conceptually important because users may search by exact standard numbers, regulatory phrases, department names, or natural-language descriptions of evidence. Research comparing embedding strategies in electronic health record retrieval shows that retrieval effectiveness is not generic and should be tuned to the structure and language of healthcare documents [23]. The vector database should therefore be treated as a governed accreditation knowledge infrastructure rather than a simple storage layer.

Continuous update and document lifecycle management

Accreditation knowledge bases must remain current because policies are revised, indicators change, standards are updated, and evidence files are added throughout the survey cycle. The system would integrate with document management systems to detect newly approved documents, superseded versions, retired policies, and updated standards, while preserving prior versions for auditability. Healthcare AI privacy and accountability discussions indicate that lifecycle governance is a core safety requirement, not an optional technical enhancement [24, 26, 27]. In this system, document lifecycle management would help prevent outdated evidence from being presented as current proof of compliance.

RAG Architecture – Retrieval and Generation for Accreditation Queries

Query understanding and intent recognition

The query layer would classify user requests into categories such as standard-specific evidence requests, gap analysis questions, policy clarification requests, audit-history searches, quality indicator summaries, or cross-department

comparisons. This classification would guide retrieval filters, source prioritization, and response format so that a request for “evidence for emergency preparedness training” is handled differently from a request for “current policy owner for emergency generator testing.” RAG studies in medical reasoning and question answering show that retrieval must be aligned with the user’s intent and the type of answer expected [6, 11, 18]. In accreditation work, intent recognition would be expected to reduce irrelevant retrieval and improve the usefulness of generated evidence summaries.

Retrieval with standard-aware ranking

Standard-aware retrieval would first identify candidate chunks through dense and keyword search, then re-rank them according to relevance to the standard, authority of the source, document currency, and evidentiary strength. A current approved policy would generally be weighted differently from a draft procedure, while a completed audit with corrective action closure would be interpreted differently from a planning note. RAG applications in laboratory regulations and radiology illustrate the importance of domain-specific retrieval constraints when users need answers grounded in authoritative materials [14, 15]. For accreditation, re-ranking should therefore incorporate the evidence hierarchy used by quality and compliance teams.

LLM generation with strict instruction

The generation layer would be instructed to synthesize answers only from retrieved chunks, avoid unsupported claims, identify conflicting evidence, and state when the available documents do not establish compliance. The response would be structured around the accreditation question, the relevant standard or requirement, supporting evidence, source citations, possible gaps, and suggested items for human review. Biomedical RAG reviews and clinical implementations emphasize that strict grounding is necessary to reduce hallucination and make outputs inspectable in high-stakes settings [9, 20, 21]. The system should therefore privilege cautious, citation-rich synthesis over confident but weakly supported narrative.

Handling multi-standard queries and cross-referencing

Many accreditation questions span multiple standards because a single domain, such as infection prevention or emergency preparedness, may involve policies, staff training, equipment checks, performance indicators, committee oversight, and incident review. The system would map broad user questions to multiple standards or document categories, retrieve evidence for each, and highlight areas where documents align, conflict, or leave gaps. Medical RAG systems developed for specialty-specific reasoning suggest that cross-document synthesis is one of the key advantages of retrieval-grounded language models when evidence is dispersed across a corpus [10, 11, 16, 19]. In accreditation preparation, this cross-referencing capability could help teams move from isolated document collection toward integrated readiness review.

Evidence Synthesis, Citation, and Answer Grounding

Inline citation and source traceability

The system would link every factual claim in a generated response to the exact retrieved source chunk from which it was derived, including the document title, version, page or section, approval date, and source repository. This traceability is essential because accreditation specialists must be able to verify not only that evidence exists, but also that it is current, authoritative, and applicable to the standard under review. Scoping and systematic reviews of medical RAG emphasize citation accuracy, evidence grounding, and retrievability as core requirements for safe use in healthcare knowledge tasks [21, 28]. In this accreditation system, inline citations would therefore function as both a usability feature and a governance safeguard.

Evidence gap identification

When retrieved documents do not provide sufficient support for a standard element, the system should explicitly report that the evidence base appears incomplete rather than generating a plausible but unsupported answer. For example, if a policy mentions staff training but no current training logs or competency records are retrieved, the response would identify the policy as partial evidence and flag the missing departmental documentation for human follow-up. Studies of RAG in healthcare question answering show that retrieval-grounded systems should be evaluated not only for correct answers, but also for their ability to avoid unsupported synthesis when relevant evidence is absent or ambiguous [8, 9, 18]. In accreditation preparation, this conservative behavior would be central to distinguishing genuine readiness from superficial document availability.

Aggregation and trend analysis of quality indicators

The system could retrieve quality indicator definitions, dashboard exports, committee reports, and improvement plans, then generate plain-language summaries that connect performance trends to accreditation standards. Such summaries would be especially useful when survey preparation requires evidence that the hospital monitors performance, reviews variation, and acts on quality concerns through governance processes. Research on documentation burden and quality indicator reporting shows that collecting and explaining indicator data can impose substantial work on healthcare professionals, especially when information must be reformatted for governance or external review [3, 4]. A RAG system would not replace statistical interpretation, but it could assist teams by locating relevant indicator narratives and connecting them to documented action plans.

Safety, Auditability, and Human-in-the-Loop Review

Guardrails against hallucination and misrepresentation

A post-generation validator would check whether each claim in the response is supported by one or more retrieved chunks, and unsupported statements would be removed, highlighted, or routed for human review. This mechanism would be particularly important when the model summarizes compliance status, because an unsupported assertion could create false confidence or obscure a gap requiring corrective action. Healthcare AI safety literature stresses that accountability, external validation, and responsible oversight are necessary when algorithmic outputs may influence organizational decisions [24, 29]. In the proposed system, guardrails would be designed to make uncertainty visible rather than to automate accreditation judgment.

Collaborative review dashboard

The review dashboard would allow accreditation coordinators, quality officers, department leaders, and subject matter experts to inspect generated answers alongside retrieved evidence. Users could edit the narrative, reject weak evidence, request additional retrieval, assign follow-up tasks, and approve a final version for inclusion in internal readiness materials. Human-centered RAG design is consistent with healthcare AI scholarship that frames these systems as tools for augmenting expert work rather than replacing accountable professional judgment [21, 24, 25]. The dashboard would therefore serve as the operational boundary between automated evidence synthesis and human accreditation decision-making.

Audit trail and evidence of the preparation process

The system would maintain a complete audit trail of user queries, retrieved documents, generated drafts, citations, human edits, approvals, and rejected outputs. This log would support internal quality assurance by showing how evidence was assembled, what gaps were identified, and which staff members reviewed the final materials. Privacy and governance analyses of healthcare AI emphasize that sensitive institutional and patient-related information must be protected through access control, monitoring, and accountability mechanisms [26, 27]. In accreditation preparation, the audit trail could also help demonstrate that the organization used a structured and reviewable process to prepare for survey activities.

Integration into the Accreditation Preparation Workflow

Embedding in accreditation project management

The system would be integrated into the accreditation project timeline so that each standard, chapter, department, or readiness workstream has a linked evidence workspace. Team members could query evidence for specific requirements, identify missing documentation, attach reviewed responses to task records, and monitor whether standards have sufficient supporting material. Prior research on accreditation and documentation burden suggests that the preparation process is organizationally distributed and requires coordination across clinical, administrative, and quality teams [1-5]. Embedding RAG into project management would therefore help convert evidence retrieval from an ad hoc search activity into a structured readiness workflow.

From preparation to continuous readiness

Beyond a formal survey cycle, the system could support continuous readiness by monitoring newly approved policies, updated departmental evidence, audit findings, and quality indicator narratives for relevance to accreditation standards. When a policy changes or a quality report signals a potential compliance concern, the system could prompt responsible staff to review affected standards and update evidence links. RAG studies in specialized medical domains show that retrieval-grounded systems are most useful when they are tied to a maintained, authoritative knowledge base rather than treated as one-time answer generators [10, 14, 15, 17]. In this way, accreditation preparation could evolve from periodic document assembly into an ongoing knowledge management process.

Evaluation Strategy

Answer quality and citation accuracy

Evaluation should begin with expert review of generated answers for factual correctness, completeness, relevance to the accreditation question, and accuracy of citations to source documents. Accreditation specialists could compare RAG-assisted responses with manually assembled responses, focusing on whether the system retrieved appropriate evidence and avoided unsupported claims. Reviews of healthcare RAG emphasize the need for domain-specific evaluation criteria that examine answer quality, citation fidelity, and safety rather than relying only on generic language model benchmarks [9, 21, 28]. For this proposed system, evaluation should be framed as validation of a governed evidence-support workflow rather than measurement of autonomous compliance performance.

Table 2 proposes an evaluation and governance framework that treats accreditation RAG performance as a combined problem of retrieval quality, citation accuracy, evidence completeness, workflow fit, and accountable human review.

Table 2. Evaluation and Governance Framework for Accreditation-Focused Retrieval-Augmented Generation

Evaluation domain	Core assessment question	Suggested evaluation approach	Failure mode detected	Governance response
Retrieval relevance	Does the system retrieve evidence that directly addresses the accreditation question or standard?	Expert review of top retrieved chunks for standard alignment, department relevance, and evidence type appropriateness	Retrieval of semantically similar but non-applicable documents	Improve standard-aware filters, metadata quality, and re-ranking logic
Citation fidelity	Does every generated factual claim link to the correct source document,	Claim-by-claim citation audit by accreditation	Incorrect citation, missing citation, or	Require citation validation before reviewer approval;

	section, page, version, and date?	specialists or quality officers	citation to weak evidence	suppress unsupported claims
Evidence currency	Does the system prioritize current approved documents over drafts, retired policies, or superseded evidence?	Comparison of retrieved sources against document lifecycle records	Outdated policy or retired evidence presented as current support	Strengthen lifecycle controls and metadata-based exclusion rules
Evidence completeness	Does the answer identify all major evidence types needed for the standard?	Compare RAG-assisted response against a manually prepared gold-standard evidence packet	Partial evidence presented as sufficient readiness	Require explicit “missing evidence” sections in generated responses
Gap identification	Does the system clearly flag missing, conflicting, or insufficient evidence?	Scenario testing with intentionally incomplete evidence sets	Hallucinated readiness claim or failure to detect missing logs, audits, or action plans	Route gap findings to human follow-up tasks and readiness workstreams
Cross-document consistency	Does the system detect contradictions across policies, audits, indicators, and departmental files?	Expert review of multi-source queries involving known inconsistencies	Conflicting documents synthesized into a falsely coherent answer	Require conflict alerts and reviewer adjudication before use
Access control and privacy	Does the system restrict retrieval and generation according to user role, confidentiality level, and source permissions?	Permission testing across user roles and document classes	Unauthorized exposure of sensitive institutional or patient-related information	Enforce role-based access, logging, and restricted-context generation
Human review quality	Do reviewers understand, edit, approve, or reject outputs appropriately?	Review dashboard analytics, edit tracking, user interviews, and approval audits	Overreliance on AI-generated text or approval without source inspection	Require reviewer attestation and source-inspection checkpoints
Workflow efficiency	Does the system reduce repetitive evidence searching while preserving expert oversight?	Time-motion comparison of manual versus RAG-assisted evidence assembly	Faster output with reduced verification rigor	Pair efficiency metrics with citation accuracy and reviewer-quality metrics
Continuous readiness impact	Does the system help teams maintain updated evidence between formal survey cycles?	Longitudinal monitoring of evidence gaps, updated policies, task closure, and readiness packet revisions	System used only as an episodic document search tool	Integrate alerts into accreditation project management and document lifecycle workflows

Efficiency and user satisfaction

The evaluation should also examine whether accreditation coordinators and quality officers perceive the system as useful, trustworthy, and aligned with their workflow. Representative accreditation queries could be used to compare manual evidence search with RAG-assisted search, but the results should be interpreted conceptually and operationally

rather than as proof of guaranteed efficiency gains. Research on quality registration burden indicates that staff experience, perceived workload, and documentation complexity are important dimensions of governance processes in hospitals [3-5]. User evaluation should therefore include whether the system reduces search frustration, clarifies evidence ownership, and supports more meaningful review of gaps.

Operational impact on accreditation outcomes

A later evaluation phase could examine whether the system helps teams identify evidence gaps earlier, improve the consistency of readiness materials, and support more coherent preparation across departments. Any assessment of operational impact should be conducted cautiously because accreditation outcomes are influenced by leadership, culture, staffing, survey scope, and the maturity of existing quality systems. Studies of accreditation impact and healthcare AI validation both caution against attributing complex organizational outcomes to a single intervention without careful contextual analysis [1, 2, 29]. The appropriate evaluation question is therefore whether the RAG system contributes to a more transparent, reviewable, and continuously updated preparation process.

Limitations

Document quality and completeness

The proposed system would be only as reliable as the documents it ingests, because outdated policies, missing audit files, incomplete dashboards, or inconsistent departmental records would directly affect retrieval and synthesis. Even a well-designed RAG architecture cannot infer evidence that was never documented or correct governance weaknesses embedded in the source corpus. Studies of RAG in healthcare repeatedly show that retrieval quality and corpus construction are central determinants of answer usefulness [7-9, 23]. For accreditation preparation, document hygiene, version control, and source ownership would therefore be prerequisites for safe and meaningful deployment.

Organizational trust and change management

Accreditation teams may initially resist relying on an AI-assisted system because compliance work is high-stakes, reputation-sensitive, and closely tied to professional accountability. A phased introduction would be needed, beginning with internal evidence discovery and parallel manual review before any RAG-assisted content is used in formal survey preparation. Ethical analyses of healthcare AI emphasize that trust should be earned through transparency, oversight, validation, and clear assignment of responsibility [24-26]. The system should therefore be positioned as a support tool for expert review rather than as a substitute for accreditation judgment.

Conclusion

The proposed retrieval-augmented generation system would support hospital accreditation preparation by transforming fragmented institutional evidence into a searchable, traceable, and reviewable knowledge base. It would allow authorized users to ask natural-language questions about standards, policies, audits, quality indicators, regulatory requirements, and departmental evidence files. Its purpose would be to assist accreditation teams in locating and synthesizing evidence, not to independently determine compliance.

The main strength of the system is its ability to unify evidence that is normally scattered across multiple repositories and organizational units. By grounding generated answers in retrieved source documents, it could make preparation materials more transparent and easier to verify. By identifying gaps and conflicts, it could also help teams focus attention on substantive readiness issues rather than repetitive document searching.

Important challenges would remain. The system would depend on accurate, current, and well-governed documents, as well as careful access control for sensitive institutional and patient-related information. It would also require cultural

acceptance, expert oversight, and rigorous validation before being trusted in real accreditation workflows.

Pilot implementations should be considered in hospitals actively preparing for accreditation and willing to evaluate RAG-assisted preparation alongside current manual processes. Such pilots should focus on answer quality, citation accuracy, user trust, workflow fit, and the system's ability to support continuous readiness. With appropriate governance, a specialized RAG system could help shift accreditation from a stressful document scramble toward a more systematic and continuously audit-ready process.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 01 Apr 2026

Revised: 29 May 2026

Accepted: 20 Jun 2026

Published online: 20 July 2026

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

- Alhawajreh MJ, Paterson AS, Jackson WJ. Impact of hospital accreditation on quality improvement in healthcare: a systematic review. *PLoS One*. 2023;18(12):e0294180.
- Hussein M, Pavlova M, Ghalwash M, Groot W. The impact of hospital accreditation on the quality of healthcare: a systematic literature review. *BMC Health Serv Res*. 2021;21(1):1057.
- Zegers M, Veenstra GL, Gerritsen G, Verhage R, Van Der Hoeven HJ, Welker GA, et al. Perceived burden due to registrations for quality monitoring and improvement in hospitals: a mixed methods study. *Int J Health Policy Manag*. 2020;11(2):183-92.

Hesselink G, Verhage R, Hoiting O, Verweij E, Janssen I, Westerhof B, et al. Time spent on documenting quality indicator data and associations between the perceived burden of documenting these data and joy in work among professionals in intensive care units in the Netherlands: a multicentre cross-sectional survey. *BMJ Open*. 2023;13(3):e062939.

Jeurissen PP, Klazinga N, Hagenaars L. Complex governance does increase both the real and perceived registration burden: the case of the Netherlands. Comment on "Perceived burden due to registrations for quality monitoring and improvement in hospitals: a mixed methods study". *Int J Health Policy Manag*. 2021;11(4):533-5.

Das S, Ge Y, Guo Y, Rajwal S, Hairston J, Powell J, et al. Two-layer retrieval-augmented generation framework for low-resource medical question answering using Reddit data: proof-of-concept study. *J Med Internet Res*. 2025;27:e66220.

Alkhalaf M, Yu P, Yin M, Deng C. Applying generative AI with retrieval augmented generation to summarize and extract key clinical information from electronic health records. *J Biomed Inform*. 2024;156:104662.

Zhang G, Xu Z, Jin Q, Chen F, Fang Y, Liu Y, et al. Leveraging long context in retrieval augmented language models for medical question answering. *NPJ Digit Med*. 2025;8(1):239.

Amugongo LM, Mascheroni P, Brooks S, Doering S, Seidel J. Retrieval augmented generation for large language models in healthcare: a systematic review. *PLOS Digit Health*. 2025;4(6):e0000877.

Miao J, Thongprayoon C, Suppadungsuk S, Garcia Valencia OA, Cheungpasitporn W. Integrating retrieval-augmented generation with large language models in nephrology: advancing practical applications. *Medicina (Kaunas)*. 2024;60(3):445.

Wang D, Liang J, Ye J, Li J, Li J, Zhang Q, et al. Enhancement of the performance of large language models in diabetes education through retrieval-augmented generation: comparative study. *J Med Internet Res*. 2024;26:e58041.

Ke YH, Jin L, Elangovan K, Abdullah HR, Liu N, Sia AT, et al. Retrieval augmented generation for 10 large language models and its generalizability in assessing medical fitness. *NPJ Digit Med*. 2025;8(1):187.

Fukui Y, Kawata Y, Kobashi K, Nagatani Y, Iguchi H. Evaluation of a retrieval-augmented generation system using a Japanese institutional nuclear medicine manual and large language model-automated scoring. *Radiol Phys Technol*. 2025;18(3):861-76.

Nanua S, Steward R, Neely B, Datto M, Youens K. Retrieval-augmented generation for interpreting clinical laboratory regulations using large language models. *J Pathol Inform*. 2025;37:100520.

Weinert DA, Rauschecker AM. Enhancing large language models with retrieval-augmented generation: a radiology-specific approach. *Radiol Artif Intell*. 2025;7(3):e240313.

Ge J, Sun S, Owens J, Galvez V, Gologorskaya O, Lai JC, et al. Development of a liver disease-specific large language model chat interface using retrieval-augmented generation. *Hepatology*. 2024;80(5):1158-68.

Fu Z, Fu S, Huang Y, He W, Zhong Z, Guo Y, et al. Application of large language model combined with retrieval enhanced generation technology in digestive endoscopic nursing. *Front Med (Lausanne)*. 2024;11:1500258.

Jeong M, Sohn J, Sung M, Kang J. Improving medical reasoning through retrieval and self-reflection with retrieval-augmented large language models. *Bioinformatics*. 2024;40(Suppl 1):i119-29.

Perkins G, Anderson NW, Spies NC. Retrieval-augmented generation salvages poor performance from large language models in answering microbiology-specific multiple-choice questions. *J Clin Microbiol*. 2025;63(3):e01624-24.

Xiong L, Zeng Q, Luo W, Liu R. NurRAG: retrieval augmented generation for nursing question answering with large language models. *Int J Nurs Sci*. 2025. [Epub ahead of print].

- Liu S, McCoy AB, Wright A. Improving large language model applications in biomedicine with retrieval-augmented generation: a systematic review, meta-analysis, and clinical development guidelines. *J Am Med Inform Assoc.* 2025;32(4):605-15.
- Xu X, Liu S, Zhu L, Long Y, Zeng Y, Lu X, et al. Development and evaluation of a retrieval-augmented large language model framework for enhancing endodontic education. *Int J Med Inform.* 2025;203:106006.
- Myers S, Miller TA, Gao Y, Churpek MM, Mayampurath A, Dligach D, et al. Lessons learned on information retrieval in electronic health records: a comparison of embedding models and pooling strategies. *J Am Med Inform Assoc.* 2025;32(2):357-64.
- Habli I, Lawton T, Porter Z. Artificial intelligence in health care: accountability and safety. *Bull World Health Organ.* 2020;98(4):251.
- Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. *PLoS Med.* 2018;15(11):e1002689.
- Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Med Ethics.* 2021;22(1):122.
- Price WN, Cohen IG. Privacy in the age of medical big data. *Nat Med.* 2019;25(1):37-43.
- Miao Y, Zhao Y, Luo Y, Wang H, Wu Y. Improving large language model applications in the medical and nursing domains with retrieval-augmented generation: scoping review. *J Med Internet Res.* 2025;27:e80557.
- Habib AR, Lin AL, Grant RW. The Epic Sepsis Model falls short—the importance of external validation. *JAMA Intern Med.* 2021;181(8):1040-1.