

ORIGINAL RESEARCH

Open access

Generative Artificial Intelligence Framework for Producing Draft Clinical Governance Dashboards from Quality Metrics, Audit Reports, Incident Logs, Compliance Indicators, and Executive Reporting Templates

Rashid Al-Mahdi¹, Khalifa Al-Suwaidi^{1*}, Mariam Al-Kuwari²

Abstract

Hospital boards and quality committees depend on monthly clinical governance dashboards to oversee safety, quality, compliance, and organizational risk. Yet producing these reports often requires repeated manual consolidation of metrics, audit findings, incident summaries, and executive commentary from fragmented operational systems. Manual dashboard assembly can delay insight, increase administrative burden, and introduce errors when data are copied across spreadsheets, documents, and presentation templates. These workflows also make it difficult to maintain consistent language, trace every claim to its source, and produce timely board-ready narratives. This article proposes a generative AI framework that ingests quality metrics, audit reports, incident logs, compliance indicators, and executive reporting templates to produce a complete draft clinical governance dashboard. The framework is conceptual and intended to support first-draft preparation rather than autonomous publication. The proposed architecture includes a multi-source data ingestion layer, a template-guided large language model, a factual verification module, and a collaborative human-review interface. Together, these components could transform scattered institutional data into structured narrative sections, exception summaries, and draft governance commentary. The framework could reduce report preparation time, improve formatting consistency, and allow quality specialists to focus on interpretation, escalation, and improvement planning rather than repetitive assembly. Its value would depend on source grounding, auditability, privacy protection, and a clear approval workflow. Automated draft governance reporting is an audacious but feasible application of generative AI in complex health systems. With careful design, human oversight, and rigorous evaluation, such systems could support higher-integrity reporting without replacing clinical accountability.

Keywords Clinical Governance, Large language models, generative AI, Patient safety, Healthcare dashboards, Quality reporting

*Correspondence:

Khalifa Al-Suwaidi
khalifa.suwaidi@outlook.com

¹ Department of Intelligent Healthcare Systems, Faculty of Medicine, Qatar University, Doha, Qatar

² Department of Clinical Informatics Engineering, Faculty of Engineering, Hamad Bin Khalifa University, Doha, Qatar

Introduction

Clinical governance dashboards are central instruments for safety oversight because they convert operational data into board-level visibility on quality, risk, patient safety, and compliance. Their value lies not only in presenting indicators but

also in shaping how executives interpret trends, exceptions, and organizational accountability. Prior work on clinical dashboards has emphasized their role in supporting decision-making, audit and feedback, and health system monitoring, while governance research highlights the importance of board-level structures for quality oversight [1-4]. However, the administrative effort required to produce these dashboards remains substantial, particularly where metrics, incidents, audits, and compliance signals sit in separate systems [5, 6].

In many organizations, the monthly reporting process still depends on manual exports from analytics platforms, safety systems, audit repositories, and compliance trackers. Quality teams may copy values into spreadsheets, rewrite incident summaries into executive language, reconcile inconsistent terminology, and assemble narrative explanations for committees. Studies of patient safety reporting and dashboard implementation show that incident data, free-text narratives, and performance indicators require substantial interpretation before they can support governance decisions [7-10]. These manual steps create opportunities for delay, inconsistency, and loss of traceability between the final board paper and the underlying evidence [11, 12].

Generative AI and large language models have matured rapidly as tools for synthesizing structured and unstructured information into coherent clinical and administrative text. Research on biomedical large language models, clinical summarization, and automated document generation suggests that such systems could produce draft narratives when constrained by source data, templates, and domain rules [13-16]. Yet most published healthcare applications have focused on clinical knowledge, discharge summaries, or patient-facing communication rather than the recurring governance documents that drive executive oversight [17-19]. This gap motivates a framework-oriented approach to generative AI for clinical governance reporting rather than a performance-focused experimental system.

The thesis of this article is that a generative AI framework could produce a first draft of a clinical governance dashboard by transforming raw institutional data into a board-ready reporting structure. The proposed framework would ingest quality metric tables, audit findings, incident logs, compliance indicators, and executive reporting templates, then generate structured commentary, trend interpretation, exception flags, and draft recommendations for human review. It would not replace governance judgment but would act as a source-grounded drafting layer that supports timely, consistent, and auditable reporting. In this sense, generative AI becomes part of the administrative infrastructure of governance rather than a standalone author of clinical accountability.

Background

Clinical governance dashboards and their requirements

Clinical governance dashboards typically bring together quality metrics, adverse event summaries, audit findings, compliance indicators, and operational performance signals into a format that supports board, executive, and committee decision-making. Their requirements include regular cadence, clear visual hierarchy, reliable trend interpretation, exception highlighting, and traceability to underlying data sources. Reviews of healthcare dashboards show that dashboards are used across clinical and managerial settings but require careful design, workflow alignment, and evaluation to avoid becoming static displays rather than governance tools [3, 5, 6]. Governance literature also emphasizes that dashboards must support accountability, escalation, and organizational learning rather than merely describe performance [2, 20].

Current reporting workflows and bottlenecks

Current reporting workflows often require manual extraction, formatting, and narrative construction across multiple systems, creating a bottleneck between data availability and executive interpretation. Incident reports may contain structured fields and free-text descriptions, audit findings may appear as narrative documents, and compliance indicators may be stored in separate monitoring tools. Natural language processing studies of incident reporting show that extracting meaning from safety event narratives is already a complex informatics task, which helps explain why

governance reporting remains labor-intensive when performed manually [7-10]. A generative AI framework could therefore target not only document writing but also the coordination problem of assembling heterogeneous governance evidence into a coherent draft.

Large language models for automated report generation

Large language models can generate fluent domain-specific text, summarize clinical information, and assist in drafting structured healthcare documents when carefully bounded by task design and evidence sources. Biomedical research has shown that LLMs encode clinical knowledge and can support medical text generation, while recent work on discharge summaries illustrates how structured clinical data can be transformed into draft narrative documents [13, 14, 17, 19]. These capabilities are relevant to governance dashboards because the reporting task also involves translating structured and semi-structured data into a concise narrative for decision-makers. However, the same literature also cautions that fluency does not guarantee factual correctness, completeness, or appropriate organizational interpretation [15, 16, 18].

Template-guided and grounded generation

Template-guided and retrieval-augmented generation provide mechanisms for controlling the structure, terminology, and evidentiary basis of generated text. In a governance reporting context, templates would define required sections, approved headings, visual references, risk language, and escalation categories, while retrieval could bring in prior approved reports and relevant source excerpts. Recent work on retrieval-augmented generation in healthcare emphasizes the importance of grounding outputs in trusted evidence rather than relying on unconstrained model memory [21]. Ethical and security analyses further suggest that governance applications should preserve confidentiality, minimize unnecessary exposure of sensitive operational data, and maintain clear human accountability [22-24].

Evaluation and quality assurance of AI-generated administrative documents

AI-generated administrative documents should be evaluated for factual accuracy, completeness, readability, template adherence, and acceptability to the clinical leaders who will use them. Evidence from clinical summarization and dashboard evaluation suggests that technical correctness alone is insufficient; outputs must also be understandable, actionable, and compatible with existing decision workflows [4, 6, 12, 18]. For governance dashboards, evaluation should examine whether generated text correctly reflects the underlying metrics, represents incident and audit information without distortion, and supports appropriate escalation. Such evaluation should remain conceptual and prospective until pilot implementations can assess whether AI-assisted reporting improves report quality without weakening human oversight [15, 16, 25].

Framework Overview

High-level architecture

The proposed framework connects to data warehouses, incident reporting systems, audit repositories, compliance trackers, and executive reporting templates through a controlled ingestion layer. It then applies a template-based LLM pipeline that drafts each section of the governance dashboard while attaching source links, validation notes, and an audit trail. This architecture builds on prior dashboard literature showing the need for reliable data pipelines and implementation methods, while extending those principles into AI-assisted narrative generation [1, 3, 5, 6]. The output is not a final autonomous report but a draft dashboard document ready for structured review by quality, safety, and executive stakeholders.

Figure 1 presents the proposed source-grounded generative AI workflow for transforming quality metrics, audit reports, incident logs, compliance indicators, and executive templates into a verified draft clinical governance dashboard for human review and approval.

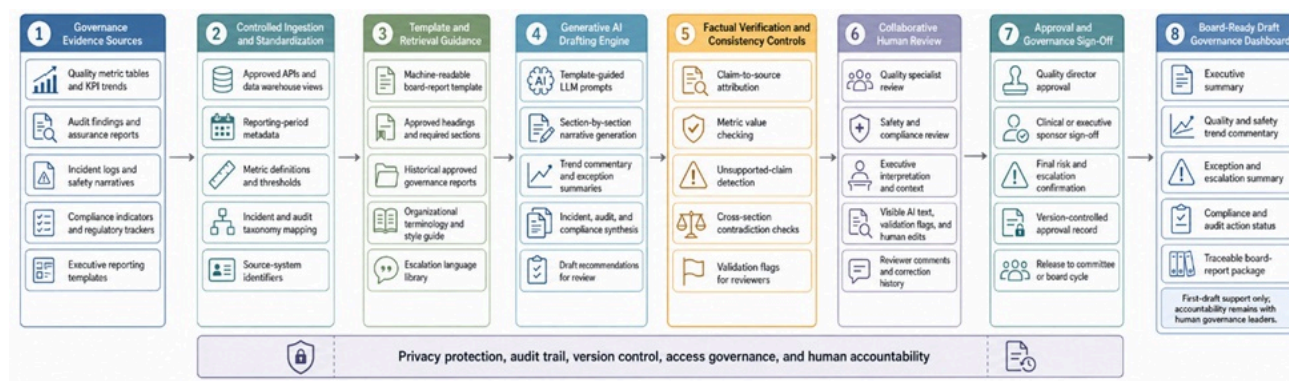


Figure 1. Source-Grounded Generative AI Workflow for Producing Draft Clinical Governance Dashboards from Quality, Safety, Audit, Compliance, and Executive Reporting Sources

Core inputs and outputs

The core inputs would include time-series quality metric tables, text summaries of recent audits, incident log excerpts, compliance indicator scores, and an executive reporting template specifying required sections and visualizations. The framework would produce a formatted draft governance report containing executive commentary, trend analysis, highlighted exceptions, and cross-references between metrics, incidents, audits, and compliance domains. Patient safety informatics studies show that incident logs can be classified and summarized, while dashboard reviews show that operational metrics must be presented in forms that support decision-making [4, 8, 9, 11]. Combining these inputs through a controlled generative layer could help governance teams move from fragmented source material to a coherent first draft.

Design principles

The design principles are source grounding, template fidelity, auditability, privacy preservation, and human-AI collaboration. Source grounding means every substantive claim should link back to a metric, audit excerpt, incident record, compliance indicator, or approved historical report, while template fidelity ensures that the draft follows the organization's established board-report structure. Privacy and ethical analyses of generative AI in healthcare underscore the need to protect sensitive data and prevent systems from being treated as independent authorities [22-24]. Human-AI collaboration is therefore central, because the framework should assist quality teams in drafting and checking reports rather than determine governance judgments on their behalf.

Data Ingestion and Standardization for Governance Reporting

Quality metrics and operational KPIs

The quality metrics ingestion layer would connect through approved APIs or data warehouse views to retrieve indicators such as mortality ratios, readmission rates, infection rates, waiting times, throughput measures, and other locally defined governance KPIs. These data would be standardized into a reporting schema that captures metric names, reporting periods, organizational units, thresholds, trend direction, and source-system identifiers. Dashboard research shows that effective clinical dashboards depend on reliable data structures, usable presentation, and integration with operational decision-making rather than isolated visualization [1, 3, 5, 26]. In the proposed framework, standardized KPI data would provide the factual backbone for generated trend commentary and exception summaries.

Audit reports and incident logs

Audit reports and incident logs would be ingested as complementary sources because audits provide formal assessments of practice and compliance, while incident logs capture safety events, contributing factors, and follow-up status. Structured incident fields such as event type, severity, date, and closure status could be combined with free-text summaries that explain context, contributing factors, and remedial action. Prior studies show that natural language processing can support classification and analysis of patient safety incident reports, but also that safety narratives require careful interpretation to avoid oversimplification [7-10, 27]. The framework should therefore treat incident and audit text as evidence for draft commentary, not as material to be freely paraphrased without validation.

Compliance indicators and regulatory data

Compliance indicators and regulatory data would be aggregated from accreditation dashboards, internal assurance systems, external benchmarks, and policy monitoring tools. These inputs could include hand hygiene compliance, mandatory training completion, audit standards, overdue actions, or other locally specified measures relevant to governance oversight. Clinical governance literature emphasizes that healthcare quality oversight requires alignment between organizational accountability, regulatory expectations, and board-level reporting routines [2, 20]. In the framework, compliance data would be normalized into categories such as compliant, at risk, overdue, or requiring escalation, while preserving links to the original evidence and reporting period.

Reporting templates and organizational style guide

The reporting template would be represented in a machine-readable form that specifies headings, required metrics, narrative sections, visual references, escalation language, and the expected executive tone. An organizational style guide would further define preferred terminology, plain-English explanations, abbreviations, and rules for describing risk, improvement actions, and exceptions. Work on clinical dashboards and large language model text generation suggests that structure and user expectations should be embedded into system design rather than left to unconstrained generation [6, 12, 14, 15]. Template-driven assembly would help ensure that drafts resemble established board papers while still allowing reviewers to adjust emphasis and interpretation.

Table 1 maps each governance evidence domain to its required standardization, AI-generated contribution, human interpretation requirement, and associated governance risk, clarifying why the proposed framework must operate as a controlled reporting architecture rather than a simple text-generation tool.

Table 1. Governance Evidence-to-Generated Output Architecture for an AI-Assisted Clinical Governance Dashboard Framework

Governance evidence domain	Typical source structure	Required standardization before generation	AI-generated draft contribution	Required human interpretation	Main governance risk if poorly controlled
Quality metrics and operational KPIs	Structured tables, time-series indicators, KPI extracts, data warehouse views	Metric names, reporting period, thresholds, service/unit identifiers, trend direction, denominator definitions, source-system identifiers	Draft trend commentary, threshold exceptions, metric-to-risk summaries, executive-level performance interpretation	Confirm whether trends reflect true operational risk, expected variation, data-quality issues, or contextual service pressures	Misstating performance direction, exaggerating risk, or generating commentary from unstable or incomplete metric definitions
Audit reports and	Narrative audit reports, action	Audit topic, date, standard assessed,	Draft audit summary,	Judge significance of findings,	Oversimplifying audit findings or

assurance findings	logs, compliance reviews, internal assurance documents	finding category, action owner, action status, evidence excerpt, review period	outstanding-action commentary, linkage between audit findings and governance priorities	determine whether actions are adequate, and clarify local accountability	converting nuanced assurance language into misleading executive claims
Incident logs and safety narratives	Structured incident fields plus free-text descriptions, severity ratings, contributing factors, closure status	Event type, severity, date, location, harm category, contributing factor taxonomy, remedial action status, de-identification status	Draft incident theme summary, safety-risk narrative, escalation flags, links between incidents and quality trends	Determine whether incident patterns require escalation, further investigation, or improvement planning	Distorting incident meaning, exposing sensitive information, or producing inappropriate causal interpretations
Compliance indicators and regulatory trackers	Compliance dashboards, accreditation measures, training records, regulatory monitoring tools	Compliance category, threshold, due date, responsible unit, current status, reporting period, evidence link	Draft compliance status commentary, overdue-action summaries, regulatory-risk statements, escalation prompts	Confirm regulatory significance, assess proportionality of language, and decide escalation pathway	Creating false reassurance, overstating non-compliance, or failing to identify material regulatory exposure
Executive reporting templates	Board-paper templates, committee reporting formats, required headings, approved visual references	Required sections, ordering rules, terminology, tone, escalation categories, sign-off fields, approved abbreviations	Structured report assembly, section sequencing, consistent executive language, required narrative components	Confirm that generated emphasis aligns with governance priorities and current organizational context	Producing a polished but incomplete dashboard that follows format without supporting meaningful oversight
Historical approved reports and style guides	Prior board reports, approved committee papers, organizational language guides, escalation phrasing	Retrieval indexing, approval status, reporting cycle metadata, document type, reusable wording boundaries	Consistent phrasing, institutional tone, examples of acceptable risk language, template-conforming commentary	Ensure retrieved language is appropriate for the current reporting period and not copied out of context	Reusing outdated language, normalizing past framing errors, or reducing attention to current risks
Validation and audit trail metadata	Source links, system logs, version histories, reviewer edits,	Claim identifiers, source references, validation status, reviewer identity, edit	Traceability notes, validation flags, reviewer-facing evidence	Review unresolved flags, verify material claims, and confirm final	Loss of defensibility, weak auditability, unclear ownership, or inability to

	approval timestamps	history, approval decision	links, draft provenance record	accountability before publication	reconstruct how the final report was produced
--	------------------------	-------------------------------	--------------------------------------	--------------------------------------	---

Generative AI Engine and Template-Driven Report Assembly

Language model core and prompt engineering

The language model core would operate within a controlled prompt structure that supplies the reporting template, standardized data extracts, approved terminology, and explicit instructions to produce draft text section by section. A locally governed or institutionally controlled deployment would be preferable where sensitive operational data are involved, because privacy and security concerns are central to healthcare generative AI adoption [23, 24]. The model could generate executive summaries, metric commentary, incident narratives, and improvement-plan text while remaining constrained by the provided source material and template. Biomedical LLM research supports the conceptual feasibility of domain-specific generation, but it also reinforces the need for careful prompting, grounding, and human review [13, 15, 16, 25].

Retrieval-augmented generation from historical reports

Retrieval-augmented generation would allow the engine to draw on a curated library of previously approved governance reports, board papers, audit summaries, and organizational terminology. Rather than copying prior text, the retrieval layer would provide examples of acceptable phrasing, escalation language, and section organization so that new drafts remain consistent with institutional style. Healthcare RAG literature suggests that retrieval can help constrain generation by anchoring outputs to relevant documents and data sources, particularly when factual precision is essential [21]. In governance reporting, this approach could reduce variation across reporting cycles while still requiring reviewers to confirm that retrieved language fits the current data and risk context.

Automated commentary and exception flagging

Automated commentary would focus on identifying metrics outside agreed thresholds, trends that appear to require attention, open incidents, overdue audit actions, or compliance indicators needing escalation. The model would be expected to generate plain-language interpretation that explains what has changed, why it matters for governance, and where human reviewers should add operational context. Studies of dashboards and audit-feedback tools show that performance information must be translated into actionable insight for leaders, while patient safety reporting research demonstrates the importance of classifying and contextualizing incidents [4, 6, 10, 11]. The framework should therefore separate mechanical exception detection from governance interpretation, with the latter remaining subject to human confirmation.

Multi-source synthesis and narrative cohesion

A final assembly pass would ensure that sections generated from quality metrics, audit findings, incident logs, and compliance indicators read as one coherent board report rather than disconnected fragments. This pass would check whether the executive summary reflects the most important risks, whether incidents are connected to relevant metrics, and whether compliance concerns are linked to appropriate actions or escalation pathways. Research on clinical summarization, dashboard design, and AI-generated healthcare text indicates that document usefulness depends on coherence, completeness, and alignment with the user’s decision context [12, 17-19]. In this framework, narrative cohesion is not a cosmetic feature but a governance requirement because executives must be able to understand relationships between data sources, risks, and proposed actions.

Ensuring Factual Accuracy, Consistency, and Auditability

Source attribution and fact-checking

Every factual statement in the draft governance dashboard should be linked to a source data cell, audit excerpt, incident record, compliance register, or approved reporting template. A validation module would compare generated text against the ingested evidence and flag discrepancies, unsupported claims, ambiguous summaries, or language that overstates the source material. This is especially important because clinical summarization and biomedical LLM research show that fluent generated text can appear credible even when factual grounding is incomplete [15, 16, 18]. In governance reporting, factual verification should therefore be treated as a core safety control rather than a post-hoc editorial preference.

Consistency checks across sections

The framework should verify that repeated references to the same metric, incident theme, audit action, or compliance issue remain consistent across the executive summary, detailed dashboard sections, and action-oriented commentary. If a performance concern is described as improving in one section and worsening in another, the system should flag the contradiction for human review before the document enters the approval workflow. Prior work on dashboards and patient safety event reporting shows that leaders rely on coherent interpretation across multiple information sources, and inconsistency can weaken the usefulness of governance information [4, 6, 11]. Consistency checks would therefore support both document quality and institutional trust in AI-assisted drafting.

Audit trail and version control

The framework should maintain a complete audit trail from raw data ingestion through draft generation, validation, human editing, approval, and final publication. This record would include source data versions, template versions, retrieved historical material, generated draft text, validation flags, reviewer edits, and approval decisions. Privacy, ethics, and governance analyses of generative AI in healthcare emphasize that organizations must preserve accountability when using automated systems with sensitive data and high-stakes operational implications [22-24]. Version control would make the report defensible by showing how each assertion moved from evidence to draft language to approved board paper.

Table 2 consolidates the verification, oversight, privacy, approval, and evaluation controls required to make generative AI-assisted dashboard drafting safe, auditable, and compatible with clinical governance accountability.

Table 2. Verification, Oversight, and Implementation Controls for Safe Deployment of Generative AI in Clinical Governance Dashboard Drafting

Control layer	Specific control mechanism	What the control prevents	Evidence or signal reviewed by humans	Responsible governance actor	Implementation value
Source grounding	Every substantive sentence is linked to a metric cell, audit excerpt, incident record, compliance indicator, template section, or approved historical report	Unsupported claims, hallucinated explanations, invented trends, and untraceable board statements	Source links, claim-source pairs, unsupported-claim flags, evidence excerpts	Quality reporting lead or governance analyst	Makes the draft defensible and allows reviewers to verify each claim before approval

Factual verification	Automated comparison of generated text against ingested values, dates, thresholds, action statuses, and incident metadata	Incorrect numbers, reversed trends, wrong reporting periods, misquoted audit findings, or inaccurate closure status	Validation report, mismatched values, source conflicts, flagged sentences	Quality intelligence team or data owner	Protects factual integrity and reduces the risk of polished but inaccurate executive reports
Cross-section consistency	Checks repeated references across executive summary, metric sections, audit commentary, incident synthesis, and compliance updates	Contradictory interpretations, inconsistent risk language, duplicated but conflicting action statements	Contradiction flags, repeated-claim map, version comparison	Dashboard editor or committee secretariat	Improves coherence across the full dashboard and strengthens executive readability
Template fidelity	Required headings, mandatory sections, reporting order, escalation categories, and tone rules are enforced before draft release	Missing sections, inconsistent formatting, unapproved wording, and deviation from board-report standards	Template-compliance checklist, missing-section report, style-guide deviations	Governance reporting manager	Ensures the AI output fits established reporting routines and reduces manual formatting burden
Privacy and access governance	Role-based access, de-identification rules, secure storage, controlled retrieval library, and restriction of sensitive incident detail	Exposure of confidential patient, staff, incident, regulatory, or organizational risk information	Access logs, redaction checks, retrieval permissions, privacy review flags	Information governance or compliance officer	Allows generative AI to operate within healthcare confidentiality and regulatory expectations
Human-in-the-loop review	Quality, safety, compliance, and executive reviewers inspect AI text, validation flags, and source evidence before approval	Overreliance on machine-generated interpretation and premature circulation of unapproved language	Reviewer comments, edits, unresolved warnings, contextual additions	Quality director, safety lead, compliance lead, executive sponsor	Preserves professional judgment and keeps accountability with human governance leaders
Approval gating	Final publication requires named sign-off by designated clinical, quality, or executive authority	Autonomous publication, unclear accountability, and unreviewed dissemination to boards or committees	Sign-off record, approval timestamp, final version history, unresolved-flag status	Medical director, nursing executive, quality director, or board-report owner	Makes the system compatible with formal governance accountability structures

Audit trail and version control	The system records source data versions, prompt/template versions, retrieved documents, generated text, edits, validation results, and approvals	Inability to reconstruct report production, weak defensibility, and loss of institutional trust	Full provenance log, version comparison, edit history, source snapshot	Governance office, audit team, or risk committee	Supports internal audit, external review, and learning across reporting cycles
Post-implementation evaluation	Drafts are assessed for accuracy, completeness, readability, actionability, reviewer workload, satisfaction, and governance usefulness	Assuming value without evidence, ignoring workflow burden, or deploying unsafe automation at scale	Expert ratings, reviewer time, correction patterns, stakeholder feedback, implementation logs	Evaluation team with governance committee oversight	Determines whether AI-assisted drafting improves reporting integrity without weakening oversight

Human-in-The-Loop Review and Approval Workflow

Collaborative editing interface

The generated draft should open in a collaborative editing environment where quality specialists, safety leads, compliance staff, and executives can inspect source links, revise narrative emphasis, and add contextual judgment. AI-generated text, validation warnings, and human edits should be visually distinguishable so reviewers understand which material has been machine-drafted and which has been clinically or administratively endorsed. Studies of dashboard acceptability and implementation show that tools are more likely to be useful when they fit existing workflows and support user interpretation rather than imposing a separate technical process [5, 6, 12]. The interface should therefore make review efficient while preserving the authority of human governance teams.

Approval and sign-off gating

The framework should include formal approval gates before any AI-drafted dashboard becomes a final governance document. Designated roles such as quality directors, medical directors, nursing leaders, or executive sponsors would review the draft, confirm that interpretation is appropriate, and sign off before circulation to committees or boards. Clinical governance research emphasizes that accountability for healthcare quality must remain embedded in organizational roles and oversight structures [2, 20]. The AI system should therefore prepare a draft for decision-makers, but it should never become the accountable decision-maker.

Learning from edits and feedback

Reviewer edits should be captured as structured feedback that can improve templates, refine approved terminology, and identify recurring weaknesses in the drafting workflow. For example, repeated corrections to risk language, action descriptions, or compliance terminology could indicate that the template or retrieval library requires revision. Research on generative AI in healthcare suggests that system performance and acceptability depend not only on model capability but also on continuous alignment with clinical workflows, user expectations, and institutional safeguards [15, 16, 25]. This feedback loop should be governed carefully so that learning from edits improves future drafts without embedding unreviewed preferences or unsafe shortcuts.

Integration Into Organizational Governance Rhythms

Alignment with monthly and quarterly reporting cycles

The framework should be scheduled around existing monthly and quarterly governance rhythms, including data lock dates, quality committee meetings, executive review periods, and board submission deadlines. Automated drafting could begin after the reporting period closes, with predefined milestones for source validation, reviewer editing, approval, and publication. Dashboard implementation literature shows that digital reporting tools must be aligned with organizational workflows if they are to support real decision-making rather than become parallel reporting artifacts [3, 5, 6]. In this framework, automation is valuable only if it fits the cadence and accountability structure of the health system.

Embedding into the governance infrastructure

Finalized reports should be published through existing board portals, quality management systems, document repositories, or executive reporting platforms, with access controls appropriate to sensitive governance material. The framework should also archive source links, validation records, and sign-off history so that future committees can understand how each report was produced. Work on healthcare dashboards and clinical governance indicates that reporting tools are most effective when embedded in organizational infrastructure for oversight, escalation, and learning [2, 4, 20, 28]. Embedding the output in established governance systems would help ensure that AI-assisted drafts become part of accountable practice rather than isolated technical demonstrations.

Evaluation Strategy

Intrinsic quality metrics

The intrinsic evaluation strategy should assess factual accuracy, completeness, readability, template adherence, source attribution, and internal consistency without relying on unsupported claims of automated performance. Each draft could be reviewed to determine whether its claims are traceable to the ingested evidence, whether required sections are covered, and whether the generated wording remains faithful to the reporting template. Literature on clinical text summarization and AI-generated healthcare documents supports evaluation approaches that consider correctness, completeness, and usability together rather than treating fluency as sufficient [17-19]. Such evaluation should be framed as a governance assurance process rather than a claim that the system has achieved autonomous reliability.

Human expert review and comparison

Human expert review should compare AI-drafted governance reports with manually prepared reports from equivalent reporting contexts, focusing on clarity, factual faithfulness, actionability, risk interpretation, and suitability for board discussion. Reviewers could include quality specialists, safety officers, compliance leads, clinical executives, and governance committee members who understand the organizational use of these documents. Studies of dashboard effectiveness, clinical governance, and patient safety analytics show that usefulness depends on whether information supports interpretation and action by real decision-makers [4, 9, 12, 29]. The evaluation should therefore prioritize expert judgment about governance value rather than narrow technical scoring.

Operational efficiency and satisfaction

Operational evaluation should examine whether the framework improves the reporting workflow by reducing repetitive assembly tasks, supporting earlier review, and increasing confidence in traceability. It should also assess whether quality teams and governance stakeholders perceive the system as helpful, transparent, and safe to use in recurring reporting cycles. Reviews of healthcare dashboards and generative AI in health care emphasize that adoption depends on workflow fit, user trust, privacy protection, and clear organizational value [6, 23, 25, 28]. Any claims about efficiency or satisfaction should remain prospective until evaluated in real implementation settings with appropriate oversight.

Limitations

Data quality and integration barriers

The proposed framework would depend heavily on the quality, completeness, timeliness, and interoperability of the organization's source data. Heterogeneous systems, inconsistent incident coding, missing audit metadata, ambiguous compliance categories, and unstructured local reporting practices could all undermine draft quality. Patient safety reporting studies show that incident data can be difficult to classify and interpret, while dashboard reviews highlight the importance of robust data pipelines and implementation methods [3, 6, 8, 9]. The framework should therefore be introduced only where data governance, source definitions, and reporting responsibilities are sufficiently mature to support reliable drafting.

Organizational trust and oversight

Organizational trust may be a major limitation because board papers and clinical governance dashboards are accountability documents, not ordinary administrative summaries. Some leaders may resist AI-authored drafts if they fear loss of judgment, reputational risk, privacy exposure, or overreliance on automated language. Ethical analyses of generative AI in healthcare emphasize that these systems should be governed through human responsibility, transparency, privacy safeguards, and careful attention to unintended consequences [22-24]. The framework should therefore be positioned as a draft assistant that strengthens human review, not as a substitute for professional accountability or executive oversight.

Conclusion

A generative AI framework for clinical governance dashboard drafting could transform scattered institutional data into a coherent first draft for board and executive review. By ingesting quality metrics, audit reports, incident logs, compliance indicators, and reporting templates, such a framework could support the recurring work of governance reporting while preserving the need for human judgment.

The key strengths of the framework are automated synthesis, template adherence, source grounding, factual verification, and collaborative human review. These features would help quality teams move from repetitive document assembly toward interpretation, escalation, and improvement planning.

Important challenges remain, including data readiness, integration complexity, privacy protection, cultural acceptance, and the need for rigorous validation in real governance workflows. The framework should not be presented as an autonomous reporting authority, because clinical governance depends on accountable human interpretation and institutional responsibility.

Pilot programs in healthcare organizations with mature governance structures would be a practical next step. Such pilots should demonstrate feasibility, refine review workflows, and evaluate whether AI-assisted drafting can improve timeliness and consistency while maintaining high-integrity governance reporting.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 17 Jul 2025

Revised: 09 Sep 2025

Accepted: 02 Nov 2025

Published online: 25 February 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

- Franklin A, Gantela S, Shifarrow S, Johnson TR, Robinson DJ, King BR, et al. Dashboard visualizations: supporting real-time throughput decision-making. *J Biomed Inform.* 2017;71:211-21.
- Brown A. Understanding corporate governance of healthcare quality: a comparative case study of eight Australian public hospitals. *BMC Health Serv Res.* 2019;19(1):725.
- Rabiei R, Almasi S. Requirements and challenges of hospital dashboards: a systematic literature review. *BMC Med Inform Decis Mak.* 2022;22(1):287.
- Xie CX, Chen Q, Hincapié CA, Hofstetter L, Maher CG, Machado GC. Effectiveness of clinical dashboards as audit and feedback or clinical decision support tools on medication use and test ordering: a systematic review of randomized controlled trials. *J Am Med Inform Assoc.* 2022;29(10):1773-85.
- Helminski D, Kurlander JE, Renji AD, Sussman JB, Pfeiffer PN, Conte ML, et al. Dashboards in health care settings: protocol for a scoping review. *JMIR Res Protoc.* 2022;11(3):e34894.
- Helminski D, Sussman JB, Pfeiffer PN, Kokaly AN, Ranusch A, Renji AD, et al. Development, implementation, and evaluation methods for dashboards in health care: scoping review. *JMIR Med Inform.* 2024;12(1):e59828.
- Fong A, Harriott N, Walters DM, Foley H, Morrissey R, Ratwani RR. Integrating natural language processing expertise with patient safety event review committees to improve the analysis of medication events. *Int J Med Inform.* 2017;104:120-5.
- Wang Y, Coiera E, Runciman W, Magrabi F. Using multiclass classification to automate the identification of patient safety incident reports by type and severity. *BMC Med Inform Decis Mak.* 2017;17(1):84.
- Young IJ, Luz S, Lone N. A systematic review of natural language processing for classification tasks in incident reporting and adverse event analysis. *Int J Med Inform.* 2019;132:103971.

Tabaie A, Sengupta S, Pruitt ZM, Fong A. A natural language processing approach to categorise contributing factors from patient safety event reports. *BMJ Health Care Inform.* 2023;30(1):e100731.

Scott J, Dawson P, Heavey E, De Brún A, Buttery A, Waring J, et al. Content analysis of patient safety incident reports for older adult patient transfers, handovers, and discharges: do they serve organizations, staff, or patients? *J Patient Saf.* 2021;17(8):e1744-58.

Siette J, Dodds L, Sharifi F, Nguyen A, Baysari M, Seaman K, et al. Usability and acceptability of clinical dashboards in aged care: systematic review. *JMIR Aging.* 2023;6:e42274.

Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. *Nature.* 2023;620(7972):172-80.

Yang R, Zeng Q, You K, Qiao Y, Huang L, Hsieh CC, et al. Ascle—a Python natural language processing toolkit for medical text generation: development and evaluation study. *J Med Internet Res.* 2024;26:e60601.

Lu Z, Peng Y, Cohen T, Ghassemi M, Weng C, Tian S. Large language models in biomedicine and health: current research landscape and future directions. *J Am Med Inform Assoc.* 2024;31(9):1801-11.

Tian S, Jin Q, Yeganova L, Lai PT, Zhu Q, Chen X, et al. Opportunities and challenges for ChatGPT and large language models in biomedicine and health. *Brief Bioinform.* 2024;25(1):bbad493.

Aali A, Van Veen D, Arefeen YI, Hom J, Bluethgen C, Reis EP, et al. A dataset and benchmark for hospital course summarization with adapted large language models. *J Am Med Inform Assoc.* 2025;32(3):470-9.

Bednarczyk L, Reichenpfader D, Gaudet-Blavignac C, Ette AK, Zaghir J, Zheng Y, et al. Scientific evidence for clinical text summarization using large language models: scoping review. *J Med Internet Res.* 2025;27:e68998.

Ganzinger M, Kunz N, Fuchs P, Lyu CK, Loos M, Dugas M, et al. Automated generation of discharge summaries: leveraging large language models with clinical data. *Sci Rep.* 2025;15(1):16466.

Thanasas GL, Tatsi K, Koutoupis AG, Davidopoulos LG, Ploumpis IC. Health and Clinical Governance: A systematic literature review. *Rev Eur Stud.* 2023;15:44.

Yang R, Ning Y, Keppo E, Liu M, Hong C, Bitterman DS, et al. Retrieval-augmented generation for generative artificial intelligence in health care. *NPJ Health Syst.* 2025;2(1):2.

Oniani D, Hilsman J, Peng Y, Poropatich RK, Pamplin JC, Legault GL, et al. Adopting and expanding ethical principles for generative artificial intelligence from military to healthcare. *NPJ Digit Med.* 2023;6(1):225.

Chen Y, Esmaeilzadeh P. Generative AI in medical practice: in-depth exploration of privacy and security challenges. *J Med Internet Res.* 2024;26:e53008.

Ning Y, Teixayavong S, Shang Y, Savulescu J, Nagaraj V, Miao D, et al. Generative artificial intelligence and ethical considerations in health care: a scoping review and ethics checklist. *Lancet Digit Health.* 2024;6(11):e848-56.

Moulaei K, Yadegari A, Baharestani M, Farzanbakhsh S, Sabet B, Afrash MR. Generative artificial intelligence in healthcare: a scoping review on benefits, challenges and applications. *Int J Med Inform.* 2024;188:105474.

Lee K, Jung SY, Hwang H, Yoo S, Baek HY, Baek RM, et al. A novel concept for integrating and delivering health information using a comprehensive digital dashboard: analysis of healthcare professionals' intention to adopt a new system and trend of its real usage. *Int J Med Inform.* 2017;97:98-108.

Denecke K, Paula H. Analysis of critical incident reports using natural language processing. In: *IndHealth 2024.* 2024:1-6.

Coiera E, Chan A, Brooke-Cowden K, Rahimi-Ardabili H, Halim N, Tufanaru C. Clinical and economic impact of digital dashboards on hospital inpatient care: a systematic review. *JAMIA Open*. 2025;8(4):ooaf078.

De Micco F, Di Palma G, Ferorelli D, De Benedictis A, Tomassini L, Tambone V, et al. Artificial intelligence in healthcare: transforming patient safety with intelligent systems—a systematic review. *Front Med (Lausanne)*. 2025;11:1522554.