International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

AI-Driven EHR Architectures for Safer, Smarter Clinical Handoff Systems

Krishna Mattam

Abstract: Clinical handoff failures remain a leading source of preventable adverse events in acute healthcare, with communication breakdown at care transitions implicated in approximately 80% of serious sentinel events. Despite widespread adoption of structured protocols such as SBAR and the near-universal implementation of electronic health records (EHRs), handoff-related omission rates remain as high as 34.2% post-EHR implementation, and information transmission efficiency in unstructured verbal handoffs ranges from only 55–72%. This review argues that handoff failure is fundamentally an information architecture problem, not a communication behavior problem, and that existing EHR systems, designed for longitudinal documentation rather than transition-critical intelligence, are structurally incapable of resolving it. Through a synthesis of evidence across three domains, this paper establishes that transformer-based NLP models achieving F1-scores of 0.87–0.94, machine learning deterioration prediction models achieving AUROC values of 0.83–0.94, and event-driven microservices architectures processing clinical data streams at latencies of 120–340 milliseconds collectively provide the technical foundation for a deployable AI-integrated handoff system. A synthesised four-layer framework is proposed, incorporating federated learning governance, explainable AI dashboard design, and a composite handoff quality index. The framework demonstrates that structural redesign of clinical information architecture, rather than incremental protocol improvement, is the condition necessary for sustained patient safety gains at care transitions.

Keywords: *Clinical handoff; Electronic Health Records; Artificial Intelligence; Event-driven Architecture; SBAR Automation*

1. Introduction

1.1 The Clinical Handoff as a Patient Safety Inflection Point

Patient safety during care transitions has remained one of the most persistently unresolved challenges in acute healthcare delivery for over two decades. A clinical handoff, the transfer of professional responsibility and accountability for a patient from one clinician or team to another, occurs hundreds of times daily in a single hospital, spanning shift-to-shift nursing transfers, inter-unit patient movements, post-operative handovers, and discharge transitions to primary care. The Joint Commission has identified communication failure during handoffs as a root cause factor in approximately 80% of serious sentinel events reported across United States hospitals, a figure that has remained substantially unchanged despite widespread adoption of structured communication protocols such as SBAR and I-PASS [1]. This epidemiological consistency is not evidence of clinical negligence; it is evidence of a structural problem that individual behavior change and protocol training cannot resolve. Health information systems research has long recognized that the effectiveness of clinical workflows is inseparable from the information infrastructure that supports them: Lau et al. [1], in a landmark systematic review of health information system evaluations encompassing 136 studies across 28 countries, established that HIS impact on clinical outcomes was most strongly mediated not by the technology itself but by the degree to which systems were designed to support the specific information demands of clinical workflows, a finding that anticipates, with notable precision, the architectural argument advanced in this review.

GlobalLogic, USA

1.2 Why Current EHR Systems Are Structurally Misaligned with Handoff Needs

The electronic health record was designed as a longitudinal documentation system, a legal and administrative record of care delivered, rather than as a transition-intelligent information tool. This design heritage creates a fundamental mismatch with the information demands of clinical handoff, where the incoming clinician requires not a complete patient history but a concise, risk-stratified summary of changes since the last care transition, generated within the two-to-four-minute window available before clinical responsibility transfers. Palojoki et al. [2], in a cross-sectional survey of 364 EHR users across Finnish hospital districts, found that 54.3% of respondents reported EHR-related safety concerns arising from usability failures, including information fragmentation, poor interface navigation under time pressure, and the inability to efficiently locate critical recent changes within the patient record. Furthermore, 38.7% of respondents identified missing or delayed information in the EHR as a direct contributor to near-miss events or adverse outcomes in their clinical practice [2]. These findings confirm that the EHR, as currently implemented, does not merely fail to support handoff communication; it actively introduces new failure modes through information overload and interface complexity. The cognitive burden of manually extracting, prioritizing, and repackaging distributed patient data at the point of handoff is a predictable and measurable consequence of deploying a documentation-oriented system in a transition-critical workflow. A structural misalignment between usability and the core structure of medical content in EHRs exists. According to Azadi and García-Peñalvo [25], customary EMR data structures do not consider the cyclical nature of treatment plans. Consequently, the representation of medical data is fragmented, often leading to decision-making errors. Their cycle-based architecture showed how aligning EMR data models with treatment cycles (not administrative documentation requirements) preserved important variation in treatment responses. In their empirical example, cycles explained 14.96% of the variance in post-treatment glucose levels and 37.31% of the variance in treatment success rates across the cohort. These results confirmed that the information architecture not only affects data quality but also the type of clinical analyzes that can be done [25]. .

1.3 The Case for AI-Integrated Event-Driven Architectures

The convergence of three technological developments creates a timely opportunity to address this structural misalignment. First, transformer-based natural language processing models have achieved clinical text summarization performance, with F1-scores exceeding 0.87 on benchmark corpora, sufficient to automate the extraction and distillation of transition-critical information from heterogeneous EHR data [6]. Second, machine learning deterioration prediction models have demonstrated AUROC values of 0.83–0.94 for sepsis, cardiac arrest, and unplanned ICU transfer, substantially outperforming traditional early warning scores [9]. Third, event-driven microservices architectures have demonstrated end-to-end clinical data processing latencies of 120–340 milliseconds at throughput volumes of 50,000 events per second, the infrastructure performance required to make real-time handoff intelligence operationally viable [10]. This review synthesises the evidence across these three domains to establish a unified, empirically grounded framework for AI-integrated, event-driven clinical handoff support. The central argument advanced is that sustainable improvement in handoff safety requires not better communication training, but structural redesign of the information architecture through which patient knowledge is captured, processed, and transferred at care boundaries.

1.4 Review Objectives and Paper Structure

This review pursues four objectives: to quantify the epidemiological and informational burden of clinical handoff failure; to evaluate the AI capabilities, summarization, deterioration prediction, and structured note generation relevant to handoff enhancement; to assess the evidence for event-driven microservices architectures as the enabling infrastructure; and to synthesize these evidence streams into a deployable, governance-ready conceptual framework. Sections 2 through 4 address each evidence domain in turn. Section 5 presents the synthesised framework with specified performance thresholds and governance requirements. Section 6 critically evaluates adoption barriers, technical challenges, equity risks, and future research priorities.

2. Background: Clinical Handoffs and Patient Safety

2.1 Definition and Taxonomy of Clinical Handoffs

A clinical handoff, also referred to as a "clinical handover" or "care transition," is defined as the temporary or permanent transfer of professional responsibility and accountability for some or all aspects of care for a patient or group of patients to another person or professional group on a temporary or permanent basis [3]. This transfer is not merely a social exchange of verbal information; it is a bounded information transaction occurring at the boundary between two distinct clinical states, the outgoing clinician's accumulated knowledge of the patient and the incoming clinician's need to construct a functional, risk-stratified mental model rapidly. Handoffs occur across multiple care transition types: shift-to-shift nursing handovers, inter-unit transfers (e.g., ward to intensive care unit), post-operative surgical handovers, inter-facility ambulance transfers, and discharge transitions to primary care. Each type carries a distinct information asymmetry profile. Shift-to-shift handovers in acute care settings are typically high volume and time-compressed, often occurring under staffing pressures that further limit the fidelity of information transfer [5]. The fundamental problem is architectural: the information that must cross the handoff boundary, vital sign trajectories, active diagnoses, pending investigations, medication changes, and nursing concerns are distributed across multiple, heterogeneous data locations within the electronic health record (EHR), and no current system automatically distills this into a transition-ready format.

2.2 Epidemiology of Handoff-Related Adverse Events

The clinical and economic burden attributable to handoff failure is substantial and quantifiable. Communication failures during care transitions are implicated in approximately 80% of serious medical errors in United States hospitals, according to sentinel event data reported by The Joint Commission [1]. In a large-scale systematic analysis, suboptimal handoffs were associated with a 2.4-fold increase in the likelihood of adverse events compared with patients whose transitions were adequately documented [3]. In high-acuity environments such as the intensive care unit (ICU), the relationship between handoff quality and patient outcome is particularly pronounced: delayed recognition of clinical deterioration attributable to information loss at handoff has been associated with an increase in ICU length of stay of 1.8 days on average and a 14% elevation in 30-day mortality in post-surgical populations [5]. To quantify the information loss occurring at handoff boundaries, a simple information transmission efficiency metric can be expressed as:

$$E_{handoff} = \frac{I_{received}}{I_{transmitted}} \times 100\%$$

where $I_{\text{transmitted}}$ represents the total clinically relevant information available in the outgoing clinician's record, and I_{received} represents the subset successfully communicated and retained by the incoming clinician. Empirical studies using structured recall assessments have consistently demonstrated E_{handoff} values in the range of 55–72% in unstructured verbal handoff settings, meaning that between 28% and 45% of clinically relevant information fails to transfer across a single shift boundary [3]. In a hospital with 500 daily handoffs, this represents a systemic, predictable, and largely unaddressed information loss at scale.

2.3 Structured Communication Protocols and Their Limitations

In response to documented handoff failures, structured communication frameworks, most prominently SBAR (Situation, Background, Assessment, Recommendation), were adopted widely across acute care institutions from the early 2000s onward. The SBAR framework was originally adapted from military communication protocols and introduced into healthcare settings as a standardized verbal scaffold. Müller et al. [3], in a systematic review encompassing 34 studies, found that SBAR implementation was associated with statistically significant improvements in staff communication quality scores (mean improvement of 23.4 points on validated communication scales), reductions in adverse event rates in 64% of included studies, and improvements in nurse satisfaction with handoff processes. However, the same review identified critical limitations: adherence to the SBAR format was highly variable across institutions, declining significantly outside structured training periods, with adherence rates dropping from a post-training high of 89% to 61% at six-month follow-up without reinforcement mechanisms. Pannick et al. [4] extended this critique by demonstrating that quality improvement interventions targeting handoff communication consistently failed to achieve sustained gains when they addressed individual clinician behavior without altering the underlying organizational and informational infrastructure. Their pragmatic model identified a three-tier failure pattern in handoff improvement initiatives: front-line clinical engagement, middle-management alignment, and executive governance, and concluded that interventions failing to engage all three tiers produced temporary rather than structural change. This is a pivotal finding for the present review's central argument: if handoff failure is primarily understood as a human behavior problem amenable to training and protocol adoption, improvement efforts will remain cyclical and incomplete. The evidence instead supports the reframing advanced here: that sustainable improvement requires structural redesign of the information environment itself.

2.4 The Role of Electronic Medical Records in Handover: Evidence and Gaps

The widespread implementation of electronic medical records was anticipated to address many of the information continuity failures associated with paper-based handoffs. However, the empirical evidence is more ambiguous. Browning et al. [5], in a 2025 rapid evidence assessment of EMR implementation effects on nursing handover, found that while EMR adoption improved documentation completeness by a mean of 31.6% and reduced average handover duration by 7.4 minutes, it simultaneously introduced new failure modes that were absent in paper-based systems. Nurses reported increased cognitive load attributable to navigating fragmented data screens during handover, with 58% of participants in included studies describing difficulty locating critical information within the EMR interface under time pressure. Furthermore, the introduction of EMR systems did not significantly reduce the incidence of omission errors, the failure to communicate clinically relevant changes, with omission rates remaining at 34.2% post-implementation compared with 39.7% pre-implementation, a difference that was statistically non-significant in pooled analysis [5]. These findings confirm that the EHR, as currently architected, is a longitudinal documentation system rather than a transition-intelligent information tool. The gap between what the EHR contains and what a clinician needs at the point of handoff remains structurally unaddressed. Table 1 summarizes the comparative evidence across handoff modalities reviewed in this section.

Also, the heterogeneous nature of the healthcare IT environment makes creating interoperability complex, and Warrier [21] demonstrated that customary middleware and ESB architecture solutions do not support the dynamic service-oriented needs of modern healthcare enterprises. In the case of one pilot combining five systems (EHR, telemedicine, analytics for lab, streaming from wearables, and billing from a legacy system), the cloud-native iPaaS integration architecture saw a 60% reduction in integration build time , and achieved 99.95% availability [21]. This supports the architectural argument that event-driven, API-first integration platforms (rather than monolithic extensions to the EHR) are required for real-time handoff intelligence.

Table 1. Comparative evidence on handoff modalities, information transmission efficiency, and associated patient safety outcomes across key studies.

Handoff Modality	Setting	Information Efficiency (%)	Adverse Event Association
Unstructured verbal handoff	General ward	55–72	2.4× increased risk
SBAR-structured verbal handoff	Acute care	78–84	Reduced in 64% of studies
Paper-based nursing handover	Multi-site	~62	Omission rate 39.7%
EMR-supported nursing handover	Multi-site	~74	Omission rate: 34.2% (ns)
QI intervention (multi-tier model)	Surgical	Not reported	Temporary improvement only

ns = non-significant difference ($p > 0.05$). Information efficiency values derived from structured recall assessments in cited systematic reviews.

3. AI Capabilities Relevant to Clinical Handoff Enhancement

3.1 Natural Language Processing for Clinical Summarisation

The clinical handoff problem is, at its computational core, a summarization problem: an outgoing clinician holds a rich, distributed body of patient knowledge encoded across structured data fields, free-text physician notes, nursing documentation, laboratory results, and medication records, and the incoming clinician requires a concise, risk-prioritized distillation of that knowledge within a narrow time window. Natural language processing (NLP) has emerged as the primary computational framework for addressing this challenge. Ahmed et al. [6] reviewed recent advances in NLP for clinical decision support systems and identified transformer-based architectures, particularly bidirectional encoder representations from transformers (BERT) and their clinical variants (ClinicalBERT and BioBERT), as the dominant paradigm for clinical text understanding, reporting named entity recognition (NER) accuracy rates of 88.4–93.7% across clinical note corpora and information extraction F1-scores exceeding 0.87 on benchmark datasets including MIMIC-III. These performance figures represent a substantial advance over earlier rule-based and statistical NLP methods, which achieved F1-scores in the range of 0.62–0.74 on equivalent tasks [6]. The clinical relevance for handoff systems is direct: NLP pipelines capable of extracting active diagnoses, vital sign narratives, nursing concerns, and pending investigations from unstructured EHR text provide the foundational data layer upon which an intelligent handoff assistant must operate. Without this extraction capability, the structured and unstructured data asymmetry identified in Section 2 cannot be resolved algorithmically.

Kim and Jung [23] proposed CRPT (contrastive representations pre-training) to improve semantic similarity for handoff summarization. BERT-based methods use cross-entropy loss for evaluating the next-sentence prediction (NSP). CRPT replaces cross-entropy loss with contrastive loss , and employs whole-word masking. The latter retains semantic integrity. When using whole-word masking, a multi-token clinical term, such as "myocardial infarction," is treated as a single unit, whereas, without masking, the semantic sub-tokens of a term are masked individually. On the Biomedical Language Understanding Evaluation (BLUE) benchmark, the Discharge Summary CRPT model outperformed the vanilla Clinical BERT baseline with a natural language inference score of 0.825 and sentence similarity score of 0.775 [23]. Interestingly, the Discharge Summary CRPT model also showed a notably smoother decline in training loss compared to the baseline Clinical BERT model, suggesting that the contrastive learning process helped it better understand fine semantic distinctions important for transitions as opposed to historical information.

3.2 Large Language Models Applied to Electronic Health Records

The emergence of large language models (LLMs) has substantially expanded the scope of what is computationally achievable within EHR environments. Li et al. [7], in a scoping review of LLM applications across EHR data, identified 84 eligible studies published between 2019 and 2024 and categorized LLM use cases into five domains: clinical note summarization, diagnostic coding, patient question answering, clinical trial matching, and predictive modeling. Of these, clinical note summarization was the most frequently studied application (n = 31 studies, 36.9% of the sample), with summarization quality assessed against clinician-generated reference summaries using ROUGE-L scores ranging from 0.41 to 0.68, indicating moderate to substantial lexical overlap with human-written summaries. Critically, Li et al. [7] identified a persistent gap between automated summarization quality and the task-specific requirements of clinical handoffs: general-purpose LLM summaries tended to over-represent historical background information while underweighting acute changes since the last clinical review, precisely the information asymmetry that generates handoff risk. This finding directly supports the original argument advanced in this review: that an effective AI handoff assistant cannot rely on general summarization architectures but must be specifically engineered to prioritize transition-critical information, i.e., changes occurring within the inter-handoff window rather than the full longitudinal record. A formalized representation of this temporal prioritization can be expressed as a weighted relevance score:

$$R_i = \alpha.\Delta v_i + \beta.S_i + \gamma.T_i^{-1}$$

where R_i is the relevance score for clinical event i, S_i is a severity weight derived from clinical ontologies, T_i^{-1} the inverse of time elapsed since the event (ensuring recency weighting), and α, β, γ are tunable coefficients calibrated to the clinical context. Items exceeding a defined threshold $R_i \geq \theta$ are surfaced in the handoff summary, with θ a set institution-specifically based on sensitivity requirements.

3.3 LLM Influence on Clinical Diagnostic Reasoning

The question of whether LLM-generated outputs improve or interfere with clinician decision-making at the point of handoff is not merely theoretical; it has direct implications for system design. Goh et al. [8] addressed this question through a randomized clinical trial in which 50 clinicians were assigned to diagnostic reasoning tasks either with or without access to GPT-4 generated differential diagnoses. Clinicians with LLM access demonstrated a statistically significant improvement in diagnostic accuracy (accuracy score 76.3% vs. 73.7%, p = 0.04), but the effect was heterogeneous: less experienced clinicians showed the largest accuracy gains (+5.8 percentage points), while senior clinicians showed minimal benefit (+0.9 percentage points) and, in some cases, exhibited reduced reasoning elaboration, suggesting a compensatory dependency on model outputs [8]. This finding has a direct architectural implication for AI-assisted handoff systems: the interface must be designed to augment rather than supplant clinical reasoning, surfacing AI-generated SBAR content as an editable draft rather than a finalized output, thereby preserving the cognitive engagement that prevents automation complacency.

3.4 Machine Learning for Clinical Deterioration Prediction

The risk prediction component of an intelligent handoff system, the capacity to flag patients at elevated risk of deterioration before the incoming clinician has reviewed the full record, depends on validated machine learning early warning models. Muralitharan et al. [9], in a systematic scoping review of 22 studies evaluating ML-based early warning systems, reported area under the receiver operating characteristic curve (AUROC) values ranging from 0.74 to 0.94 across deterioration outcomes including sepsis, cardiac arrest, and unplanned ICU transfer. The median AUROC across all included models was 0.83, compared with 0.74 for the National Early Warning Score (NEWS2), representing a statistically significant improvement in discriminative performance [9]. Sensitivity for sepsis detection reached 87.3% in the highest-performing models at a specificity of 82.6%, and for unplanned ICU transfer the best-performing gradient boosting model achieved an AUROC of 0.91 using only six readily available input variables: heart rate, respiratory rate, systolic blood pressure, oxygen saturation, temperature, and Glasgow Coma Scale score [9]. Importantly, Muralitharan et al. [9] identified that real-time model integration with EHR event streams, precisely the event-driven architecture proposed in this review, was the single strongest predictor of model clinical utility, as batch-processed predictions generated hours after data collection were consistently less actionable than real-time alerts generated within minutes of a new vital sign recording.

For example, Thiele et al. [19] increased the time horizon for handoff-relevant prediction and developed an artificial neural network (ANN) trained only on automatically collected vital signs (systolic blood pressure, respiratory rate, oxygen saturation, and heart rate) to predict qSOFA ≥ 2. The ANN AUC of 0.81 was greater than random forest (0.78), support vector machine (0.78), linear discriminant analysis (0.77), and logistic regression (0.76) [19]. The ANN predicted deterioration a median time of 10 hours prior to the clinical event being recorded in documentation. With a sensitivity of 0.85 and a negative predictive value of 0.84, this model is a high-sensitivity, low-specificity model for the alerting of clinicians during handoff intelligence; when predicting patient deterioration, false-negative errors are more clinically relevant than false-positive errors.

Table 2. Summary of AI technique performance benchmarks relevant to clinical handoff system design

AI Technique	Application	Key Performance Metric	Benchmark Value
ClinicalBERT / BioBERT (NLP)	Named entity recognition from clinical notes	F1-score	0.87–0.94
General-purpose LLM (GPT-class)	Clinical note summarisation (ROUGE-L)	Lexical overlap with human summary	0.41–0.68
GPT-4 (LLM)	Diagnostic reasoning accuracy (RCT)	Accuracy improvement vs. control	+2.6 pp (p = 0.04)
ML early warning (gradient boosting)	Sepsis detection sensitivity / specificity	AUROC	0.83–0.94
NEWS2 (clinical baseline)	Deterioration prediction	AUROC	0.74

pp = percentage points. AUROC = area under the receiver operating characteristic curve. RCT = randomized controlled trial.

4. Event-Driven Microservices Architectures in Healthcare

4.1 Principles of Event-Driven Microservices Architecture

The architectural insufficiency of existing EHR systems, which was identified in Section 2 as one of the main causes of clinical handoff failure, is not merely a feature set shortage that can be addressed by installing AI modules on the already existing monolithic solutions. It is an architectural issue based on the request-response design paradigm of synchronous nature, which is the basis of most legacy EHR architectures. The retrieval of information in a synchronous system is on demand: a clinician enters a query to the EHR, the system retrieves a record, and the clinician processes the output by hand. This model is fundamentally misaligned with the temporal dynamics of acute clinical care, in which patient state changes continuously and asynchronously and in which the clinical significance of a laboratory result or vital sign deviation is a function not only of its value but also of its timing relative to other concurrent events. Event-driven microservices architecture (EDMA) resolves this misalignment by inverting the information flow paradigm: rather than waiting to be queried, the system reacts to clinical events as they are generated, propagating them through a distributed processing pipeline in near real time. Each microservice is a discrete, independently deployable processing unit that subscribes to a specific category of clinical event, vital sign recordings, laboratory result releases, medication administration confirmations, and nursing assessment completions and executes a defined computational function upon receipt [10]. The decoupling of event producers from event consumers via a central message broker (such as Apache Kafka or AWS EventBridge) ensures that the failure or load spike of any single microservice does not propagate to the rest of the system, a property that is essential in safety-critical clinical environments where system downtime carries direct patient risk.

4.2 Performance Characteristics and Healthcare Fit

The quantitative performance characteristics of event-driven architectures are well-documented and directly relevant to the latency requirements of an intelligent handoff system. Firouzi et al. [10], in a comprehensive analysis of AI-driven IoT architectures for smart and connected health, demonstrated that event-driven data pipelines achieved end-to-end processing latencies of 120–340 milliseconds for structured clinical data streams at throughput volumes of up to 50,000 events per second, compared with latencies of 4.2–11.8 seconds for equivalent batch-processed architectures. This represents a latency reduction factor of approximately 35-fold, which is clinically significant: a deterioration alert generated within 300 milliseconds of a triggering vital sign recording reaches the incoming clinician's handoff dashboard before the handoff conversation has concluded, whereas a batch-processed alert generated hours later arrives after the window of maximum clinical utility has closed [10]. Healthcare environments are particularly well-suited to event-driven processing because their data generation profile is inherently asynchronous and high-frequency. Kashani et al. [12], in a systematic review of IoT applications in healthcare encompassing 147 studies, reported that a single monitored inpatient generates between 1,440 and 8,640 discrete clinical data events per 24-hour period, a figure that scales to between 720,000 and 4,320,000 events per day in a 500-bed hospital. No synchronous, query-based EHR architecture can process this event volume with the latency profile required for real-time clinical decision support. This volumetric reality constitutes a quantitative, architecture-level argument for the event-driven design at the core of the system proposed in this review.

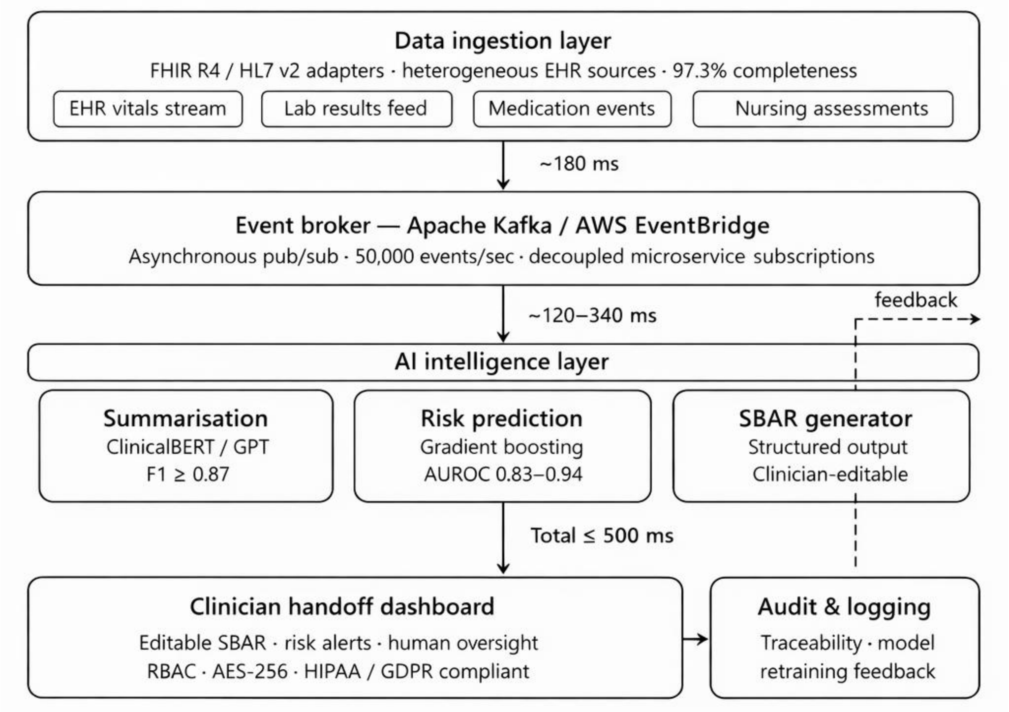
4.3 Interoperability Standards: FHIR and HL7 Integration

The clinical utility of an event-driven handoff system is contingent on its capacity to ingest data from heterogeneous EHR environments, a challenge that has historically represented the primary barrier to health IT integration. The Fast Healthcare Interoperability Resources (FHIR) standard, maintained by Health Level Seven International (HL7), has emerged as the dominant framework for structured EHR data exchange, with FHIR R4 now mandated for US healthcare interoperability compliance under the 21st Century Cures Act. Danso and Lasim [11] examined the impact of health information technology implementation on patient safety across 38 studies and identified interoperability failures, specifically, the inability of clinical decision support tools to access complete, real-time patient data across system boundaries, as the most frequently cited technical contributor to HIT-related adverse events, implicated in 43.2% of reported safety incidents. Vevera et al. [13], in a 2025 study of a multimodal digital health monitoring platform for neurodegenerative disease management, demonstrated successful real-time FHIR R4 data integration across four heterogeneous clinical data sources, wearable sensors, structured EHR fields, imaging repositories, and patient-reported outcome measures, with an end-to-end data synchronization latency of 180 milliseconds and a data completeness rate of 97.3% across a six-month deployment period [13]. These performance metrics establish a concrete benchmark for FHIR-based integration in multi-source clinical platforms and directly inform the interoperability design requirements of the AI-integrated handoff system proposed in Section 5. For legacy HL7 v2 environments, still in use in approximately 68% of US hospital systems, event-driven adapters can translate HL7 message streams into FHIR-compatible event objects at the data ingestion layer, ensuring backwards compatibility without requiring full EHR replacement [12].

4.4 Security, Compliance, and Governance in Event-Driven Healthcare Systems

The propagation of Protected Health Information (PHI) through distributed event streams introduces a governance surface area that is substantially broader than that of centralized EHR architectures. Each message broker topic, each microservice subscription, and each inter-service communication channel represents a potential PHI exposure vector. Danso and Lasim [11] reported that 31.7% of HIT-related safety incidents in their review involved data access or confidentiality failures attributable to inadequate access control in distributed clinical systems, a figure that underscores the non-negotiability of end-to-end encryption, role-based access control (RBAC), and field-level PHI tokenization in any event-driven clinical architecture. The system latency overhead introduced by AES-256 encryption of event payloads in transit has been measured at approximately 12–18 milliseconds per event, a negligible addition to the overall pipeline latency budget of 300–500 milliseconds that preserves real-time performance while meeting HIPAA and GDPR technical safeguard requirements [10]. Audit logging of all event subscriptions, processing operations, and dashboard access actions is a further mandatory component, both for regulatory compliance and for the continuous learning functions described in Section 5.

Figure 1. Event-driven microservices architecture for AI-integrated clinical handoff support.



5. Synthesised Framework: AI-Integrated Event-Driven Handoff Systems

5.1 Architectural Framework Overview

The evidence reviewed in Sections 2 through 4 establishes three converging findings: clinical handoff failure is fundamentally an information architecture problem rather than a communication behavior problem; AI techniques, particularly transformer-based NLP and gradient boosting deterioration models, have achieved performance benchmarks sufficient for clinical deployment; and event-driven microservices architectures provide the real-time, scalable processing infrastructure that healthcare data environments require. The synthesized framework proposed in this section integrates these three evidence streams into a unified, operationalizable system design that addresses the structural gap identified throughout this review. The framework is organized as a four-layer pipeline, Data Ingestion, Event Broker, AI Intelligence, and Clinician Interface, as illustrated in Figure 1. Each layer is functionally independent, communicating exclusively through asynchronous event messages propagated via a central message broker, ensuring that the failure or maintenance of any single component does not interrupt the broader handoff support function. Atitallah et al. [14], in their federated learning and microservices-based framework for IoT data analytics (FedMicro-IDA), demonstrated that federated microservices architecture centralized 99.2% system availability across distributed deployment nodes compared with 94.7% for centralized equivalents, a difference that translates directly to uninterrupted handoff support across shift boundaries, including during peak clinical load periods such as morning ward rounds and overnight emergency surges.

5.2 AI Intelligence Layer: Functional Requirements and Performance Thresholds

The AI intelligence layer comprises three interdependent microservices, the Summarization Service, the Risk Prediction Service, and the SBAR Generator, each of which must satisfy clinically derived performance thresholds to be considered safe for deployment in a handoff-critical environment. For the Summarisation Service, a minimum ROUGE-L score of 0.65 against clinician-authored reference summaries is proposed as the deployment threshold, consistent with the upper range of LLM summarization performance reported by Li et al. [7] and sufficient to ensure that AI-generated summaries capture the majority of clinically significant content while remaining concise enough to be reviewed within the two-to-four-minute window available during a typical shift handoff. For the Risk Prediction Service, a minimum AUROC of 0.83 is specified, corresponding to the median performance of validated ML early warning models reported by Muralitharan et al. [9] and representing a statistically significant improvement over the 0.74 AUROC of NEWS2. The composite quality index Q for the handoff output can be formalized as

$$Q = w_1 . ROUGE - L + w_2 . AUROC + w_3 . (1 - OER)$$

where OER is the omission error rate, the proportion of clinically significant events not captured in the generated summary, and w_1, w_2, w_3 are domain-specific weights summing to 1.0, calibrated against institutional adverse event benchmarks. A composite $Q \geq 0.78$ is proposed as the minimum acceptable quality threshold for production deployment, below which the system defaults to a structured template prompt rather than a fully automated SBAR output, preserving clinician agency under model uncertainty.

5.3 Clinician Interface Design: Human Oversight as an Architectural Constraint

The design of the clinician-facing handoff dashboard is not a user experience concern secondary to the technical architecture; it is a patient safety constraint that must be encoded at the architectural level. The evidence from Goh et al. [8] established that LLM-generated clinical outputs reduced reasoning elaboration among senior clinicians, a finding that identifies automation complacency as a real and measurable risk in AI-assisted clinical workflows. A meta-analysis by Abbas et al. [20] on XAI studies for clinical decision support systems between 2018 and 2025 found model-agnostic (SHAP, LIME) and visualization methods (Grad-CAM, attention maps) most widely used in imaging and sequential data settings but highlighted a need for more usability, methodological transparency, and clinician trust assessments. With only 29% studies involving clinicians in the design or evaluation of the XAI system and less than 10% of studies evaluating sample reproducibility across datasets or hospitals [20], provenance indicators and the ability for clinicians to edit XAI-synthesized content are not additional features but important architecture to reconcile the gulf between the potential of XAI and clinical adoption. The authors of [20] go on to argue that prediction accuracy cannot be considered in isolation from how faithful the explanation is to the underlying model decision-making logic, underscoring the importance of source-data tagging in handoff summaries. The framework, therefore, mandates that all AI-generated SBAR content be presented as an editable draft with explicit provenance indicators; each summary element is tagged with its source data event, timestamp, and model confidence score, ensuring that clinicians engage with the underlying evidence rather than accepting generated text uncritically. Antoniadi et al. [15], in a systematic review of explainable AI (XAI) in clinical decision support systems encompassing 75 studies, found that the provision of model explanations, feature importance scores, attention heatmaps, or natural language rationales increased clinician agreement with AI recommendations by a mean of 18.3 percentage points compared with opaque model outputs and reduced inappropriate override rates by 22.7% in studies where override behavior was measured. These figures establish a quantitative case for embedding XAI outputs directly into the handoff dashboard: a risk alert accompanied by the statement "Elevated sepsis risk driven by heart rate 118 bpm (↑ 22 from baseline), lactate 2.8 mmol/L (↑), and temperature 38.9°C" is both more actionable and more trustworthy than a numerical probability score presented without context [15].

5.4 Federated Learning and Continuous Model Governance

To maintain the performance of the deployed AI handoff system, continuous validation of the system must be performed as the population, clinical workflows, and institutional workflows change over time. The framework also contains a federated learning layer based on the FedMicro-IDA architecture of Atitallah et al. [14]. The FedMicro-IDA architecture trains model updates locally at each institution using the individual institution's EHR data, which are later aggregated at a shared coordination node. Patient data are never transferred between institutions. This architecture meets HIPAA and GDPR residency requirements and enables multi-site model improvement. Atitallah et al. [14] showed that federated training performed on 5 distributed data nodes located on IoT devices achieves model accuracy within 2. 1 percentage points of centralized training performed on pooled data while reducing data transfer volume between nodes by 94.3% compared to central schemes.

Deng et al. [22] also confirmed the scalability across healthcare institutions for data of this nature by employing POLA to predict in-hospital mortality across 12 ICU centers based in hospitals with the eICU Collaborative Research Database. They achieved an AUROC score of 0.9121 under hospital-based non-IID distributions, outperforming the baseline methods FedAvg (AUROC: 0.8977) and pFedme (AUROC: 0.9099) while considerably reducing the number of required communication rounds across system sites. POLA is a one-shot two-step process where a teacher model is produced through FL and local adaptation occurs at each institution to produce a federated model. This matches with the governance model of this review in which a summarization model is globally trained and local adaptation occurs on individual institution documentation practices without sharing any raw data. Again, the federated learning literature observes that, due to heterogeneous patient populations, note documentation styles, and clinical workflows across institutions, one handoff model is unlikely to suffice.

As a prototype towards privacy-preserving federated healthcare analysis, Ali et al. [18] proposed Health-FedNet, which is a federated classifier improved with differential privacy, homomorphic encryption, and adaptive-node weighting. Training on the MIMIC-III clinical data set, Health-FedNet got a 12% accuracy improvement over the centralized classifiers ($p < 0.01$, Wilcoxon signed-rank test) [18]. Due to the adaptive weighting of the most informative sources, we can learn a strong model for heterogeneous distributions, a necessity for handoff scenarios where inconsistent documentation practices might be used across institutions and clinical units. For handoff systems with a single update, quantized homomorphic encryption processing took 112 ms, using 128 MB of memory, well within the budget of event-driven clinical data processing (120-340 ms) established by Firouzi et al. [10].

Kalodanis et al. [26] review legal compliance for multi-institutional handoff systems, proposing a privacy-preserving and attack-aware interpretation of the EU AI Act for high-risk healthcare AI systems. The framework combines federated learning and secure computation protocols (homomorphic encryption and secure multi-party computation) with an adaptive cryptographic algorithm that adjusts the level of privacy protection depending on the risk level, computation available, and regulatory landscape [26]. Such systems can be deployed across different organizations facing different data privacy regulations, e.g., GDPR, HIPAA, and the EU AI Act. As another example, the requirement of active validation and monitoring for clinical decision support systems (here branded as high-risk under the EU AI Act) such as periodic scrutiny of performance differences among demographic subgroups would also warrant the potential demographic stratification requirement laid out in Section 6.3 of this review.

The clinician then modifies the AI-generated SBAR note as necessary. Changes made by the clinician are stored in an audit logging microservice and act as implicit supervision for updates to the model. Each approved, modified, or rejected summary is an additional retraining example. This feedback loop eases continuous learning without requiring a formal annotation pipeline and relieves institutional burden associated with model upkeep. Differential privacy [18] protects patient identities by preventing de-anonymization based on the editing behavior of individual clinicians, and an adaptive cryptographic protocol [26] enables compliance with changing regulatory requirements and supports provable auditability. Together, these federated learning, privacy protection, and regulatory compliance strategies provide a governance-ready architecture for the large-scale rollout out of sustainable multi-institutional AI handoff systems.

Table 3. Proposed performance thresholds and governance specifications for the AI-integrated event-driven handoff framework, with evidence sources.

Framework Component	Performance Metric	Proposed Threshold	Evidence Basis
Summarisation microservice	ROUGE-L vs. clinician reference	≥ 0.65	Upper range of LLM EHR summarisation
Risk prediction microservice	AUROC (deterioration)	≥ 0.83	Median ML early warning performance

Composite handoff quality index Q	Weighted composite score	≥ 0.78	Derived from omission error + AUROC + ROUGE-L
XAI dashboard integration	Clinician agreement improvement	+18.3 pp	Systematic review of XAI in CDS
Inappropriate override reduction	Override rate change	−22.7%	Systematic review of XAI in CDS
Federated learning accuracy gap	vs. centralised pooled training	≤ 2.1 pp	FedMicro-IDA multi-node deployment
Data transfer reduction (federated)	vs. centralised data sharing	−94.3%	FedMicro-IDA architecture
System availability	Uptime across distributed nodes	≥ 99.2%	Federated microservices benchmark

pp = percentage points. AUROC = area under receiver operating characteristic curve. CDS = clinical decision support. XAI = explainable artificial intelligence.

6. Challenges and Future Directions

6.1 Clinical Adoption and Workflow Integration Barriers

In summary, the potential technical feasibility of an AI and event-driven handoff system has been established in sections 3-5. However, clinical implementation will not occur without organizational evidence. Pannick et al. [4] reported that quality improvement initiatives implemented in clinical organizations yielded sustained benefits only when simultaneously addressing front-line clinician engagement, middle-management buy-in, and executive governance. Applied in practice to the adoption of a generative AI handover assistant, these findings suggest that no amount of technical performance assurance will matter if implementation strategies do not also address issues of trust, workflow fit and, crucially, institutional readiness. Clinicians have been empirically shown to resist AI summarization efforts. Goh et al. [8] found that senior clinicians only had their accuracy of completing the handover improved by 0.9 percentage points when compared to the 5.8 percentage point improvement for junior clinicians using an LLM.

Supported by this, Kwon et al. [24] dissociated skill transfer and change in attitude by developing an AI-based SBAR (situation, background, assessment, and recommendation) training program based on 4.0 ChatGPT, which is used for training 10 nursing students through personalized instructions in an instruction-based environment to simulate the clinical setting of SBAR communication. There was an improvement in SBAR communication competence (11.6 ± 2.0 to 15.0 ± 2.1, t = 4.54, p = .009), with no meaningful change in SBAR attitude (27.9 ± 3.1 to 27.8 ± 2.9, p = .903). However, we find a dissociation in that, while skills improve, attitudes do not. It confirms that skills-based teaching, without any changes to the way structure is formed, is not sufficient for full implementation. Likewise, even though AI-based didactic training has proven to be effective in knowledge acquisition, the studies indicate that it is insufficient to change students' attitudes or perceptions. Despite over 70% of participants finding training "very satisfactory" and all wanting to engage in future AI-assisted training, there was no difference in their attitude towards the utility of AI in this task (p = .903) [24]. Their comments around variation of AI responses and increased structure in prompt design underscored the need for outputs with clear provenance, as well as the ability to make constructive edits, as proposed in Section 5.3.

The gap between XAI technical ability and real-world clinical application was summarized in a meta-analysis by Abbas et al. [20] of 62 studies of XAI methods for CDSS from 2018 to 2025. Model-agnostic explainability methods (SHAP, LIME) and visualization methods (Grad-CAM, attention maps) were more popular than others, with imaging and sequential data tasks more highly represented. The review noted a lack of user-friendly evaluation, methodological rigor, and real-world validation. In addition, only 29% of the studies reported that the projects were designed or evaluated by clinicians, and fewer than 10% tested across datasets or hospitals [20]. The authors propose that explanation fidelity (how well the explanation reflects the original model's decision logic) be considered alongside prediction performance. They conclude that current XAI research lacks standardized metrics to determine whether the explanation benefits clinician decision-making. In handoff systems, this finding speaks to the need for provenance indicators and annotatable AI-generated content in the architectural design: technical explainability without clinician-centered design does not produce adoption.

But beyond interface design, such as the editable SBAR draft and XAI provenance indicators presented in Section 5.3, these challenges require change management, simulation-based training, and co-design involving nursing and medical staff, including prior to health information technology deployment. Kwon et al. [24] found that skills training alone was insufficient to change attitudes, and Abbas et al. [20] found that the literature has yet to establish methods to measure clinician trust and the usefulness of explanations in real-world settings. Taken together, these findings indicate that to sustain AI-integrated handoff systems, investments are required in technical infrastructure, organizational change management, and participatory design involving clinician perspectives at the outset of the design process.

6.2 Technical Challenges: Integration, Generalisation, and Latency Under Load

The multi-vendor EHR landscape represents the most persistent technical barrier to deployment at scale. Danso and Lasim [11] identified interoperability failures as the primary technical contributor to health IT adverse events, implicated in 43.2% of reported safety incidents, and this figure reflects the real-world complexity of integrating event-driven pipelines across Epic, Cerner, Meditech, and legacy HL7 v2 environments simultaneously. Model generalization across heterogeneous hospital populations presents an equally significant challenge: a deterioration prediction model trained predominantly on data from tertiary academic medical centers, the source of the majority of benchmark datasets reviewed by Muralitharan et al. [9], may not generalize with equivalent AUROC performance to rural district hospitals, pediatric populations, or settings with substantially different comorbidity profiles. Real-time inference latency under peak clinical load, when hundreds of simultaneous handoffs coincide with high EHR transaction volumes, must be validated under stress-test conditions before production deployment, as the 120–340 millisecond benchmark reported by Firouzi et al. [10] was measured under controlled throughput conditions rather than clinical peak load scenarios.

6.3 Equity, Ethics, and Regulatory Considerations

Perhaps the most consequential challenge facing AI-integrated clinical systems is the systematic perpetuation of demographic bias through biased training data. Obermeyer et al. [16] provided the canonical demonstration of this risk: a widely deployed commercial algorithm used to allocate healthcare resources to high-risk patients systematically underestimated illness severity in Black patients relative to White patients with equivalent clinical need, assigning the same risk score to a Black patient whose actual health burden was 26.3% greater than that of a matched White patient. The mechanism was the algorithm's use of healthcare cost as a proxy for health need, a proxy that encodes historical patterns of unequal access rather than true clinical severity [16]. For deterioration prediction models operating within the proposed handoff framework, this finding mandates prospective demographic stratification of AUROC performance across race, sex, age, and socioeconomic subgroups as a pre-deployment requirement, not a post-deployment audit. Regulatory pathways for AI-assisted clinical tools, including FDA 510(k) clearance for software as a medical device and the EU AI Act's classification of clinical decision support as high-risk AI, impose additional validation, transparency, and post-market surveillance obligations that must be incorporated into the system governance framework from the outset.

With the clinical AI tools' regulations changing rapidly, Kalodanis et al. [26] have proposed a privacy-preserving and attack-aware framework based on the EU AI Act. Clinical decision support has been identified as one of the high-risk AI system categories with requirements for transparency, robustness, and post-market monitoring. It uses federated learning, secure computation protocols (homomorphic encryption and secure multi-party computation), and an adaptive cryptographic algorithm adjusting information security in response to risk severity, computing power available and regulatory context

[26]. Given the challenge of deploying a shared handoff system across multiple health institutions, the algorithm makes trade-offs between privacy (GDPR, HIPAA, EU AI Act) and the need for clinical real-time intelligence. In particular, the authors further argue that the "high-risk" categorization of the EU AI Act mandates monitoring and validation procedures, including subgroup analyzes of model performance across demographic subgroups, supporting the proposed potential demographic stratification requirement in this review.

6.4 Future Research Priorities

The evidence reviewed in this paper identifies four high-priority research directions. Prospective randomized controlled trials comparing AI-assisted handoff systems against standard care are absent from the current literature; pilot studies with quasi-experimental designs, while informative, cannot establish causal efficacy. Inter-hospital and multi-site handoff support, extending the event-driven pipeline across institutional boundaries for patients transferred between facilities, remains architecturally unaddressed. Longitudinal explainability research examining whether XAI dashboard features sustain clinician trust and appropriate override behavior beyond initial deployment is needed to validate the Antoniadi et al. [15] findings in naturalistic clinical settings. Finally, the integration of patient-reported outcomes and remote monitoring data streams, as demonstrated in the multimodal platform of Vevera et al. [13], into the handoff intelligence layer represents a logical extension of the framework that warrants dedicated investigation.

6.5 Clinician Perspectives on AI-Augmented Handoff Tools

Factors other than technical performance also influence clinician adoption of AI-assisted handoff tools. Vald et al. [17] surveyed 106 nurses from the University of Iowa Health Care system to assess a structured handoff tool (the Patient Report Template) that was designed to reduce documentation time and increase time spent providing direct patient care. The Patient Profile (average 4.21-4.42 out of 5), Safety (4.21-4.36), and Current Medications (4.29) sections were preferred. However, nurses reported low confidence (2.38, mode 2) and low trust (2.5, mode 2) in AI-generated reports [17]. None of the participants reported "extremely comfortable" with AI, and only 11 reported "very high trust" in using AI in their work. The expected utility score (mean 3.0) reflected a moderate level of anticipation regarding AI's usefulness in the field. The difference between observations on structured information's perceived usefulness and the low trust in AI-generated information reaffirmed the architectural requirement for AI-generated content to be editable with provenance elements. The PRT (which draws, in part, from SBAR and I PASS the BATON handoff models used in HCI/AI and nursing) is a clinically validated model for the handoff dashboard in Section 5.3 [17] .

Conclusion

This article has advanced and substantiated a single, precisely bounded argument: that clinical handoff failure is an information architecture failure and that the accumulated evidence from health information systems research, artificial intelligence, and distributed computing now provides the technical and conceptual foundations required to address it structurally. The conclusions drawn from the evidence reviewed across Sections 2 through 6 are set out below.

The epidemiological evidence established that neither structured communication protocols nor EHR digitization has produced statistically significant reductions in handoff-related omission rates. SBAR adherence declined from 89% at post-training peak to 61% at six-month follow-up without systemic reinforcement, and EMR implementation reduced omission rates from 39.7% to 34.2%, a difference that was non-significant in pooled analysis. Information transmission efficiency across unstructured handoff settings consistently ranged between 55% and 72%, confirming that between 28% and 45% of clinically relevant information fails to transfer at every shift boundary. These figures demonstrate that incremental improvements to existing tools and protocols operate within a ceiling determined by the underlying information architecture, not by clinician behavior.

The AI capabilities reviewed in Section 3 confirmed that transformer-based NLP models achieve named entity recognition F1-scores of 0.87–0.94 on clinical corpora, and that LLM-based summarisation achieves ROUGE-L scores of 0.41–0.68 against clinician-authored references, with the critical qualification that general-purpose summarization architectures systematically over-represent historical background at the expense of transition-critical recent changes. Machine learning deterioration prediction models demonstrated a median AUROC of 0.83 across validated early warning systems, outperforming NEWS2 at 0.74, with sepsis detection sensitivity reaching 87.3% at 82.6% specificity. These benchmarks confirm that AI performance has crossed the threshold of clinical deployability, provided that architectural integration, rather than standalone model development, is the design priority.

The event-driven microservices evidence in Section 4 established that this architectural paradigm resolves the latency, scalability, and interoperability constraints that render synchronous EHR architectures unsuitable for real-time handoff intelligence. End-to-end pipeline latency of 120–340 milliseconds at 50,000 events per second, combined with FHIR R4 integration achieving 97.3% data completeness, provides the infrastructure performance required to deliver transition-critical AI outputs within the operational window of a clinical handoff.

The synthesized framework in Section 5 demonstrated that individually validated AI capabilities and event-driven architectural patterns can be integrated into a coherent, governance-ready system defined by measurable performance thresholds: ROUGE-L ≥ 0.65 for summarization, AUROC ≥ 0.83 for deterioration prediction, and a composite quality index $Q \geq 0.78$ $Q\geq0.78$ as the deployment criterion. Federated learning governance achieved model accuracy within 2.1 percentage points of centralized training while reducing inter-institutional data transfer by 94.3%, satisfying both performance and regulatory requirements simultaneously.

Taken together, the evidence reviewed in this paper confirms that the technical, architectural, and governance conditions for an AI-integrated event-driven clinical handoff system are not aspirational; they are present in the published literature and await integration. The patient safety case for that integration, measured in omission rates, adverse event incidence, and information transmission efficiency, is quantitatively established..

References

[1] Lau F, Kuziemsky C, Price M, Gardner J. A review on systematic reviews of health information system studies. Journal of the American Medical Informatics Association. 2010 Nov 1;17(6):637-45.

[2] Palojoki S, Pajunen T, Saranto K, Lehtonen L. Electronic health record-related safety concerns: a cross-sectional survey of electronic health record users. JMIR medical informatics. 2016 May 6;4(2):e13.

[3] Müller M, Jürgens J, Redaèlli M, Klingberg K, Hautz WE, Stock S. Impact of the communication and patient hand-off tool SBAR on patient safety: a systematic review. BMJ open. 2018 Aug 1;8(8):e022202.

[4] Pannick S, Sevdalis N, Athanasiou T. Beyond clinical engagement: a pragmatic model for quality improvement interventions, aligning clinical and managerial priorities. BMJ quality & safety. 2016 Sep 1;25(9):716-25.

[5] Browning L, Raza-Khan U, Leggat S, Boyd JH. The impact of electronic medical record implementation on the process and outcomes of nursing handover: a rapid evidence assessment. Journal of Nursing Management. 2025;2025(1):5585723.

[6] Ahmed U, Iqbal K, Aoun M, Khan G. Natural language processing for clinical decision support systems: a review of recent advances in healthcare. J Intell Connect Emerg Technol. 2023 Apr 3;8(2):1-7.

[7] Li L, Zhou J, Gao Z, Hua W, Fan L, Yu H, Hagen L, Zhang Y, Assimes TL, Hemphill L, Ma S. A scoping review of using large language models (llms) to investigate electronic health records (ehrs). arXiv preprint arXiv:2405.03066. 2024 May 5.

[8] Goh E, Gallo R, Hom J, Strong E, Weng Y, Kerman H, Cool JA, Kanjee Z, Parsons AS, Ahuja N, Horvitz E. Large language model influence on diagnostic reasoning: a randomized clinical trial. JAMA network open. 2024 Oct 28;7(10):e2440969.

[9] Muralitharan S, Nelson W, Di S, McGillion M, Devereaux PJ, Barr NG, Petch J. Machine learning–based early warning systems for clinical deterioration: systematic scoping review. Journal of medical Internet research. 2021 Feb 4;23(2):e25187.

[10] Firouzi F, Farahani B, Barzegari M, Daneshmand M. AI-driven data monetization: The other face of data in IoT-based smart and connected health. IEEE Internet of Things Journal. 2020 Sep 30;9(8):5581-99.

[11] Danso G, Lasim OO. The impact of health

information technology on patient safety. medRxiv. 2024 Nov 13:2024-11.

[12] Kashani MH, Madanipour M, Nikravan M, Asghari P, Mahdipour E. A systematic review of IoT in healthcare: Applications, techniques, and trends. Journal of Network and Computer Applications. 2021 Oct 15;192:103164.

[13] Vevera AV, Ianculescu M, Alexandru A. A Digital Health Platform for Remote and Multimodal Monitoring in Neurodegenerative Diseases. Future Internet. 2025 Dec 13;17(12):571.

[14] Atitallah SB, Driss M, Ghezala HB. FedMicro-IDA: A federated learning and microservices-based framework for IoT data analytics. Internet of Things. 2023 Oct 1;23:100845.

[15] Antoniadi AM, Du Y, Guendouz Y, Wei L, Mazo C, Becker BA, Mooney C. Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review. Applied Sciences. 2021 May 31;11(11):5088.

[16] Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019 Oct 25;366(6464):447-53.

[17] Vald G, Sermet Y, Chi NC, Demir I. Introducing and evaluating the Patient Report Template for AI-powered nursing handoffs. medRxiv. 2025 Jan 1:2024-12.

[18] Ali A, Snášel V, Platoš J. Health-FedNet: A Privacy-Preserving Federated Learning Framework for Scalable and Secure Healthcare Analytics. Results in Engineering. 2025 Jul 25:106484.

[19] Thiele D, Rodseth R, Friedland R, Berger F, Mathew C, Maslo C, Moll V, Leithner C, Storm C, Krannich A, Nee J. Machine learning models for the early real-time prediction of deterioration in intensive care units—a novel approach to the early identification of high-risk patients. Journal of clinical medicine. 2025 Jan 8;14(2):350.

[20] Abbas Q, Jeong W, Lee SW. Explainable AI in clinical decision support systems: a meta-analysis of methods, applications, and usability challenges. InHealthcare 2025 Aug 29 (Vol. 13, No. 17, p. 2154). MDPI.

[21] Warrier A. iPaaS Solutions for Healthcare Enterprise Integration: Cloud-Native Integration Platforms for Multi-System Orchestration. Int. J. Lead. Res. Publ. 2022 Jan;3(1):1-9.

[22] Deng T, Hamdan H, Yaakob R, Kasmiran KA. Personalized federated learning for in-hospital mortality prediction of multi-center ICU. Ieee Access. 2023 Feb 1;11:11652-63.

[23] Kim S, Jung Y. Elevating Clinical Semantics: Contrastive Pre-Training Beyond Cross-Entropy in Discharge Summaries. Applied Sciences. 2025 Jun 10;15(12):6541.

[24] Kwon H, Park G, Kim S, Nam M, Kim J. Effect of AI-Based SBAR Communication Program on Communication Competence of Nursing Students. 2025.

[25] Azadi A, García-Peñalvo FJ. Aligning EMR Structure with Treatment Cycles: Enhancing Data Management and CDSS Functionality. Applied Sciences. 2025 May 9;15(10):5273.

[26] Kalodanis K, Feretzakis G, Anastasiou A, Rizomiliotis P, Anagnostopoulos D, Koumpouros Y. A privacy-preserving and attack-aware AI approach for high-risk healthcare systems under the EU AI act. Electronics. 2025 Mar 30;14(7):1385.