SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

Ai-Augmented Clinical Handoff Systems: Integrating Large Language Models With Electronic Health Records For Safer Care Transitions

Thiyagarajan Palaniyappan

Independent Researcher, USA ORCID: 0009-0000-0167-9199

Abstract
Clinical handoffs are among the most vulnerable points in the chain of medical errors. For example, 80% of serious preventable adverse events in hospitals are attributed to communication failures. While the majority of organizations have implemented scripted tools (e.g., SBAR, I-PASS), the quality of handoffs remains inconsistent, and documentation remains a burden. We describe the design and deployment of AI-augmented clinical handoff systems that integrate LLMs and EHRs by transforming structured data into structured, evidence-based handoff notes with real-time risk stratification. Technical approaches included FHIR-based EHR integration, NLP-based clinical summary generation with LLMs, and ML-based deterioration detection, enabling automated descriptive documentation that is overseen by clinical personnel. Multi-hospital system implementations have observed that LLM-augmented documentation of patient handoff notes reduced documentation time by 50.2% to 55.1% per patient, saving between 474 and 981 hours across three hospitals monthly. Notes produced with LLMs demonstrated a BERTScore of 0.859 with a ROUGE-2 of 0.322. No critical patient safety risks were identified in 1600 emergency medicine notes. Machine learning identified deterioration 11 hours in advance with an area under the receiver operating curve (AUROC) of 0.895. Three quantitative measures are introduced: note efficiency, composite note quality, and AI risk stratification, with the purpose of defining benchmarks for documentation efficiency. In the augmentation model, the AI output is presented as one of many drafts, and the clinician remains the final decision-maker, maintaining safe care transitions for the patient.

Keywords: Care Transitions, Clinical Decision Support, Early Warning Scores, Electronic Health Records, Human-AI Collaboration, Large Language Models, Patient Safety, SBAR.

1. Introduction

Clinical handoffs, a rapid exchange of responsibility from one clinician to another, occur thousands of times each day in hospitals around the world. Handoffs are critical transitions between clinicians and potentially dangerous if incomplete or lacking in important clinical information. The World Health Organization (2023) has reported that poor communication at handoffs is among the most common preventable causes of patient harm worldwide, with up to 80% of serious adverse events occurring at transitions of care. The gravity of the problem is well known. Humphrey et al. (2022), for example, in their analysis of malpractice claims, found that 49% of claims included a communications failure, with 40% of those being due to handoffs. They likewise found that the average claim cost for handoff failure cases was also higher than for cases where no communication failures occurred ($237,600 vs. $154,100).

Standardized communication tools such as SBAR (situation-background-assessment-recommendation) and I-PASS (illness severity, patient summary, action list, situation awareness and contingency plans, and synthesis by receiver) have been used in handoffs. In a study by Starmer et al. (2014) at nine teaching hospitals, the introduction of I-PASS resulted in a 23% reduction in medical errors and a 30% reduction in preventable adverse events. Attempts to replicate the findings have been mixed. No difference in the rate of preventable adverse events was observed between the intervention and control periods of a stepped wedge trial of 1465 medical records from six PICUs in Argentina by Jorro-Baron et al. (2021). The implementation of the framework likely needed additional technology for training and support to be consistent and sustained. The documentation burden also influenced

handoffs. Melnick et al. (2021) analyzed event logs of EHR systems from 57 doctors from four different specialties. However, physicians spend 49.2% of office time in the EHR, only 27% with patients (Doherty et al., 2025). The perceived documentation burden is correlated with perceived workload and EHR usability (Doherty et al., 2025). Within shift-change handoff, documentation involves synthesizing recent clinical events, active problems, medication changes, and pending tasks within a narrow window of time that is prone to clinicians entering quick and incomplete documentation without clinically relevant details.

Thanks to recent advances in large language models (LLMs), clinical natural language processing, and healthcare interoperability standards, it becomes possible to overcome these issues. Studies have demonstrated that clinical summaries produced by LLMs perform at least as well as clinician-written notes on multiple quality metrics. Furthermore, early warning scores generated using machine learning-based methods have much better performance in predicting deterioration compared to customary early warning scores. This article discusses design considerations, evidence, and implementation of AI-augmented clinical handoff systems that are integrated into the existing EHR workflows for care transitions.

2. The Handoff Problem: Scope and Systemic Failure Modes

2.1 Frequency and Consequences of Communication Gaps

In a study of OR-to-ICU handoffs at a large academic center, McElroy et al. (2015) identified 37 individual steps and 20 potential failure points. These errors have been documented in real world settings by observational studies. Marquez et al. (2024) analyzed 38 handoff sessions between pediatric residents and found that 50% had at least one protocol failure (26% due to incorrect patient information, and 32% for insufficient protocol elements).

This issue may not be limited to acute care. Studies show that at least 50% of adults discharged from hospitals have at least one medication-related problem at transition. Clinically important medication errors occur frequently, even in the presence of processes for the structured and systematic discharge of patients. Each error has downstream consequences, the severity of which may vary according to the patient population and care setting.

2.2 Limitations of Protocol-Based Interventions

SBAR was used in 59% of all perioperative handoff intervention studies. Evidence shows that SBAR, as a structured communication tool, improves completeness of perioperative handoff information. It requires sustained commitment from organizational and clinician levels for successful implementation. Frameworks have been shown to improve handoff quality when taught (Reyes et al., 2016) during medical clerkship, but 97% of medical trainees did not receive formal teaching on handoffs at any point during their medical training, indicating a gap between the availability of frameworks and their routine use in clinical practice.

The attempts at replication by Jorro-Baron et al. (2021) cast doubt on the idea that protocols can be universal and strong. Even where a protocol is strong, it may not provide a safety benefit to patients unless much is done to prepare an organization to adopt the protocol and train personnel to implement it. Such reliability can be achieved with AI-augmented systems, without reliance on the memory of the clinician or the training of the organization.

3. AI-Augmented Handoff System Architecture

3.1 System Design and EHR Integration

Building an AI-augmented handoff system requires a combination of clinical natural language processing, real-time risk stratification, and healthcare interoperability, which can be architected through an event-driven pattern based on shift change and clinical status change. When deployed, the EHR-integrating version queries the EHR using standard query interfaces, processes the results through multi-step AI pipelines, and drafts a structured handoff for the outgoing clinician to view, discuss, edit, and sign off on.

EHR infrastructure is interoperable, by definition as per HL7. It is based on the HL7 FHIR standards that define standard resource types and RESTful application programming interfaces (APIs). Barker et al. (2024) show that 73% of healthcare technology companies are using or plan to use FHIR-based APIs, indicating that use of this level of interoperability is mature. Such a FHIR API would permit the handoff system to access structured clinical information such as a problem list, medication list, lab tests, vitals, and clinical notes.

Yet, semantic interoperability issues exist. Fragmented EHR (electronic health record) ecosystems, disparate documentation standards, and real-time data feed gaps in some institutional settings hamper the quality of AI-generated summaries. These inequities are compounded in hospitals serving populations with high social

vulnerability and poor integration infrastructure and may need to be accounted for prior to eventual equitable deployment of AI-augmented handoff technology.

3.2 NLP-Driven Clinical Summarization and the Composite Handoff Quality Score

The clinical summarization module uses LLMs adapted to medical data to auto-generate handoff notes. In Hartman et al. (2024), LLM EM handoff notes outperformed physician notes on ROUGE-2 (0.322 vs. 0.088) and BERTScore Precision (0.859 vs. 0.796) across 1600 patient records. In Van Veen et al. (2024), LLMs outperformed physicians on a clinical note summarization reader study (summary considered non-inferior or superior to physician summaries 81% of the time) with 10 clinicians. Small et al. (2025) showed that residents edited considerably fewer LLM-generated hospital course summaries (31.5%) than physician-generated summaries (44.8%, $p < 0.001$).

To evaluate the performance of NLP-based summarizers systematically and reproducibly, a Composite Handoff Quality Score (CHQS) can be derived from these metrics (Table 1).

Formula 1 - Composite Handoff Quality Score (CHQS):

CHQS = w1(ROUGE-2) + w2(BERTScore_P) + w3(1 - Error_Rate)

We used weights w1 = 0.30, w2 = 0.40, and w3 = 0.30 (empirically calibrated, e.g., weighted more toward clinical semantic meaning). Thus, for the Hartman et al. (2024) LLM notes, CHQS = 0.30(0.322) + 0.40(0.859) + 0.30(1 - 0.096) = 0.097 + 0.344 + 0.271 = 0.712. The physician who annotated these notes would yield a CHQS score of 0.30(0.088) + 0.40(0.796) + 0.30(1 - 0.020). This gives 0.026 + 0.318 + 0.294 = 0.638. The LLM score (0.712) is higher than that of the physician because of increased lexical coverage (per ROUGE-2); the physician has a higher raw accuracy. This score acts as a reference point to institutions assessing and comparing AI-assisted summarization techniques.

3.3 ML-Enhanced Clinical Deterioration Detection and the ARSS Framework

A core function of AI-augmented handoff systems is to identify patients in real-time who are at risk of clinical deterioration. Early warning scores such as NEWS and MEWS feature fixed thresholds for vital sign values - which overall may be less sensitive to early deterioration. Machine learning-based early warning scores may be more predictive than conventional early warning scores. In a study by Edelson et al. (2024), eCARTv5 had an area under the curve (AUC) of 0.895 (95% CI: 0.888-0.901) compared to NEWS (0.752) and MEWS (0.789) are used for predicting deterioration within 24 hours when trained and validated on the same cohort. Churpek et al. (2025) also validated eCARTv5 at different sites and reported a median lead time of 11 h for clinical deterioration events.

The AI-Augmented Risk Stratification Score (ARSS) framework operationalizes this approach within the handoff system:

Formula 2 - AI-Augmented Risk Stratification Score (ARSS):

ARSS = a(NEWS_norm) + b(DELTA_NEWS_24h) + g(ML_score_norm)

Where NEWS_norm = NEWS score / 20 (normalized to 0-1); DELTA_NEWS_24h = change in NEWS over last 24 h, where positive values indicate clinical deterioration; and ML_score_norm = machine learning model score (0-1), where a = 0.20, b = 0.30 and g = 0.50, with g assigned a higher weight due to the higher AUROC of eCARTv5 (0.895) versus NEWS features alone (0.752). Higher ARSS-ranked patients are given priority in a handoff summary, which creates risk-stratified patient queues for incoming clinical staff, rather than a flat patient roster. The eCARTv5 detection lead time of 11 hours allows time for handoff risk communication to be anticipated, with indicators of patient deterioration incorporated into the normal handoff workflow.

3.4 Automated SBAR Generation

It maps synthesized clinical data for each of the SBAR components (Situation, Background, Assessment, Recommendation). The Situation includes patient identifiers, the primary diagnosis, and the patient's status. The Background includes a summary of the admission history and relevant clinical events. The Assessment section contains structured data on the ARSS value and its rationale, an AI-generated risk commentary, and a Recommendation section with outstanding orders, planned actions, and next-watch tasks.

Automatically generating a draft SBAR removes the cognitive burden of structuring and classifying information while writing. Clinical users edit the SBAR draft created via automation, rather than composing it from scratch. The effect is lower preparation effort and optimization of an evidence-based communication practice with less cognitive strain on the clinician. Wong et al. (2021) also mention that the clinical output of AI is strongly dependent on the

data pipeline environment in which it is generated and that clinical handoff generation will be reliable only if FHIR is strongly integrated and exchanged.

4. Implementation Evidence and Outcomes

4.1 Multi-Hospital Deployment and the Documentation Efficiency Index

For two months, Chen et al. (2026) used the multi-hospital LLM for nursing handover documentation in three Taiwanese hospitals. Handover documentation using the conventional system took 3.45-4.32 minutes per patient per handover before the LLM system was applied. Post-integration, documentation time fell to between 1.55 and 2.15 minutes per patient. The Documentation Efficiency Index (DEI) expresses this improvement.

Formula 3 - Documentation Efficiency Index (DEI):

DEI = ((T_baseline - T_AI) / T_baseline) x 100

Where T_baseline is the mean documentation time per patient before AI integration (in minutes) and T_AI is the mean documentation time per patient after AI integration (in minutes), we find that using the documentation times in Chen et al. (2026), with the lowest baseline documentation time, gives DEI = ((3.45-1.55)/3.45)*100 = 55.1%, and the highest baseline documentation time gives DEI = ((4.32-2.15)/4.32)*100 = 50.2%. This result corresponds to an efficiency of 50.2–55.1% and a time savings of 474–981 total hours across the three hospitals. The estimated monthly cost of the generative AI system across all three hospitals was $493 to $986, while the value of the labor released was $2,276 to $7,939, yielding a positive ROI for each site.

A finding that 50-86% of the LLM's notes needed to be changed by nurses suggests that the users' trust in the LLM was appropriately calibrated, that clinicians controlled the output, and that the nature of the augmentation model was appropriate.

Table 1: Performance Comparison of LLM-Generated versus Physician-Written Handoff Notes

Study	Setting	Records	ROUGE-2	BERTScore Precision	Key Finding
Hartman et al. (2024)	Emergency Medicine	1,600	0.322 (LLM) vs 0.088 (MD)	0.859 (LLM) vs 0.796 (MD)	Zero critical safety risks; 9.6% incorrectness rate (LLM) vs 2.0% (MD)
Van Veen et al. (2024)	Clinical text summarization	Multispecialty	Not reported	Not reported	LLM summaries preferred in 81% of physician reader-study cases
Small et al. (2025)	General Medicine	100 admissions	Not reported	Not reported	Residents edited 31.5% of LLM summaries vs 44.8% of MD summaries (p<0.001)
Chen et al. (2026)	Multi-hospital nursing handover	3 hospitals	Not reported	Not reported	CHQS: LLM 0.712 vs MD 0.638; 50-86% of drafts revised by nurses

Note: CHQS values calculated using Formula 1 (w1=0.30, w2=0.40, w3=0.30). MD = physician-authored notes. Error rates from Hartman et al. (2024).

Table 2: Documentation Efficiency Index (DEI) Across Multi-Hospital AI Handover Deployment

Metric	Pre-Integration	Post-Integration	DEI (%)	Monthly Hours Saved
Documentation time per patient (min) - Hospital A	3.45	1.55	55.1%	474 hrs (combined range)
Documentation time per patient (min) - Hospital B	3.92	1.95	50.3%	See aggregate
Documentation time per patient (min) - Hospital C	4.32	2.15	50.2%	981 hrs (combined range)
Monthly AI system cost (USD)	N/A	$493-$986	N/A	N/A
Freed labor value (USD/month)	N/A	$2,276-$7,939	Positive ROI across all sites	N/A

Note: DEI calculated using Formula 3. Hospital B's intermediate values were estimated from the reported range. Source: Chen et al. (2026).

4.2 Emergency Department Deployment

Despite the short shift length, high patient volumes, and wide range of clinical acuities, there were no safety issues found with using LLM to generate ED handoff documents in a potential analysis of 1600 handoff notes in the ED (Hartman et al., 2024). Further, LLM-generated notes had an incorrectness rate of 9.6%, compared to physician notes, which had a 2.0% incorrectness rate, emphasizing the necessity of human review and the limitations of LLMs in producing factual accuracy. LLM handoff notes can be interpreted as reviewed drafts: they are not idealized results but rather a design philosophy where human review of AI-generated content is desirable, not optional.

Table 3: Early Warning Score Performance Comparison with and without AI Integration

Scoring System	AUROC	Median Lead Time	Sensitivity (High-Risk)	Implication for Handoff Integration
NEWS (Traditional)	0.752	Standard	Moderate	Limited early identification; requires manual scoring
MEWS (Traditional)	0.789	Standard	Moderate	Improved sensitivity but still threshold-based
eCARTv5 (ML-based)	0.895 (95% CI: 0.888-0.901)	11 hours	High	Enables proactive handoff risk communication 11 hrs before deterioration
ARSS (AI-Augmented composite)	Based on eCARTv5 + NEWS	Up to 11 hours	High (contextualized)	Integrates ML output with vital sign trends for risk-stratified handoff queues

Note: AUROC data from Edelson et al. (2024) and Churpek et al. (2025). ARSS = AI-Augmented Risk Stratification Score (Formula 2).

4.3 Patient Safety and System Validation

According to the findings of a 2024 study by Van Veen et al., LLM-generated summaries were associated with lower probability of and extent of (potential) medical harm in clinical summary cases compared to human expert-generated summaries (12% vs. 14% and 16% vs. 22%, respectively). Given the systematic and thorough nature of LLM summary generation, the slight drop in accuracy relative to physician summaries might be justified by a reduction in omission-based harms.

For example, Bednarczyk et al. (2026) used Failure Modes, Effects and Criticality Analysis (FMECA) to evaluate clinical documents that LLMs generated. They identified 14 failure modes across the document types and rated the system's usability at 79.2 out of 100. Moderate to substantial agreement was seen for failure mode identification. This FMECA framework provides a basis for monitoring the production safety of AI-generated handoff content.

Table 4: Economic and Patient Safety Outcomes of AI-Augmented Handoff Systems

Outcome Domain	Baseline Metric	Post-AI Metric	Source
Malpractice claim cost (with handoff failure)	$237,600	$154,100 (no failure)	Humphrey et al. (2022)
Potential medical harm likelihood (LLM vs Human)	14% (human expert summaries)	12% (LLM summaries)	Van Veen et al. (2024)
Monthly AI cost per hospital (USD)	N/A	$493-$986	Chen et al. (2026)
Freed labor value per hospital per month (USD)	N/A	$2,276-$7,939	Chen et al. (2026)
AI alert override rate (traditional CDS)	73.3% (overridden)	Reduced via narrative integration	Nanji et al. (2018)
FMECA-rated system usability score (0-100)	N/A	79.2 / 100	Bednarczyk et al. (2026)

Note: All cost figures in USD. CDS = clinical decision support.

5. Human-AI Collaboration: The Augmentation Model

5.1 Design Philosophy: Draft Review over Automation

Systems that support human handoffs should not replace human reasoning. Under a human augmentation model, clinicians are not passive receivers of AI outputs; they should critically evaluate rather than accept what the AI presents. This approach differs from alerts in clinical decision support systems where the patterns are binary. The data from Nanji et al. (2018) showed an overall override rate of 73.3%. Even in those instances, 60% of overrides were in clinically appropriate cases where the clinical alert was ignored. Clinical decision support systems that interrupt clinical workflow and require positive acknowledgment trigger the alert fatigue that they were designed to avoid.

AI-augmented handoff systems attempt to address this issue by embedding risk information into the narrative structure of the handoff note. In this system, the incoming clinician sees a list of patients that synthesizes ARSS risk scores, with patients at the highest risk appearing at the front of the handoff queue. This approach delivers risk information in context, rather than being parsed out into a notification that requires binary decision-making by the user.

5.2 Trust Calibration and Organizational Learning

Draft revisions of 50-86% (Chen et al., 2026) suggest that clinicians may not be accepting or rejecting AI-generated notes outright but are instead using them as a scaffold, making focused edits to produce a usable note more quickly. This would represent an ideal degree of human-AI interaction whereby the AI is completing the most salient parts of the synthesis, and the clinician is responsible for accuracy and clinical correctness.

Wiesenfeld and Kellogg (2025) argue that generative AI systems create affordances for how knowledge is created, transferred, and learned in organizations. In the case of clinical handoffs, generative AI-generated drafts are knowledge artifacts that allow clinicians to verify, revise, and expand knowledge that has already been created rather than create knowledge anew. The transition from generating documentation to verifying the documentation and conducting other tasks may mirror the provider cognitive load of transitioning between shifts, documentation, and patient evaluation.

5.3 Addressing Alert Fatigue through Intelligent Prioritization

The median lead time to deterioration while using eCARTv5 is 11 hours (Churpek et al., 2025), so eCARTv5 can be used to identify patients who are going to deteriorate. Instead of tracking patient response to an urgent clinical situation, the ARSS-integrated handoff system identifies patients whose clinical trajectories suggest worsening acuity by the next shift. This information is integrated into the handoff summary so clinicians can proactively address problems rather than react to crises.

6. Technical Challenges and Mitigation Strategies

6.1 Hallucination and Factual Accuracy

The primary safety risk of LLM-based handoff systems is their generation of clinically incorrect content. Hartman et al. identified five hallucinations (0.31% hallucination rate) in 1600 handoff notes generated by an LLM (2024). Few inaccuracies may be acceptable in low-acuity scenarios, but even one incorrect fact in a high-acuity handoff may cause serious patient harm. Possible mitigation strategies include retrieval-augmented generation architectures to ground LLM responses in verified EHR data, confidence scoring to allow for the identification and reviewer requirement of uncertain responses, and cross-validation against structured EHR data such as vital signs, lab results, and active medication lists.

6.2 Interoperability and Data Quality

Despite FHIR success, unstructured clinical notes, inconsistent code use, and incomplete medication reconciliation obstruct the accuracy of AI-generated summaries based on real-world data. Fakha and Boonstra (2025) identify several key factors that influence the successful functioning of an AI model during the care transition phase. Of these factors, the volume and richness of the data integrated into the AI models are the most important. Poor integration and excessive editing decrease efficiency.

6.3 Workflow Integration and Change Management

Considering how a technology fits into clinical practice is important for successful implementation. Chen et al. (2026) describe a successful implementation where the nursing staff were involved at the early design stage, tested the system in cycles, and were trained in how to work with the AI. Given the fact that 97% of medical trainees report they never receive any handoff training during medical education, AI-generated SBAR notes may also be used as scaffolding, modeling the structure of good handoffs while the clinician is engaged in the productive work of documenting. Thus, AI-generated SBAR may not only be used to improve the quality of individual handoffs but may also act as a training tool for the future clinical workforce.

7. Broader Implications

7.1 Clinician Wellbeing and Workforce Sustainability

Documentation burden is one of the most commonly described factors associated with clinician burnout in the literature. A systematic review found that 68% of included studies on artificial intelligence and electronic health record optimization found burnout reduced. In AI proposed handoff systems, they decrease the documentation burden by 50.2-55.1%. Reducing nursing time by 474-981 hours per month for a three-hospital system can increase the clinical capacity of the staff, allowing them to spend more time with patients and less time mediating information and administrative duties. These changes can improve burnout and disengagement.

7.2 Health Equity Considerations

Handoff failures may disproportionately affect patients with complex needs. Patients with complex needs face more care handoffs due to clinical complexity and longer hospital length of stay, but their hospitals may have less mature interoperability infrastructure and, thus, may not be able to fully leverage the potential of AI-augmented handoff systems. Standardizing FHIR API implementation and lowering integration costs are prerequisites to equitable access to AI-augmented handoff systems. Policies to promote interoperability with safety-net organizations are a core element of the health equity strategy in the AI-augmented care system.

7.3 Regulatory and Legal Context

The 21st Century Cures Act regulations require interoperability infrastructure be built in support of deploying AI-augmented handoffs. FDA oversight of clinical decision support software establishes baseline safety standards for deploying AI-generated clinical content in the United States. The potential for reduced malpractice liability ($237,600 vs. $154,100) with good handoff communication (Humphrey et al. (2022)) is compatible with the patient safety value of handoffs and the organization's mission. As AI-augmented and other solutions evolve, regulations can and should adapt to overcome the newly recognized failure modes to maximize the likelihood of success. The FMECA framework proposed and validated by Bednarczyk et al. (2026) illustrates this concept.

8. Future Directions

Future AI-augmented handoff systems may utilize more multimodal data streams beyond structured EHR data, including continuous physiologic monitoring data, automated medical imaging analyzes, and ambient clinical intelligence such as voice recognition of bedside huddles. These systems may prioritize the variable depth and format of content for different clinicians based on their role and prior patient familiarity. Closed-loop safety monitoring (comparing the system output of a handoff to observed outcomes) could allow for continual improvement in summarization and ARSS calibration. While Wiesenfeld and Kellogg (2025) argue that AI-augmented handoff systems will likely affect institutional learning, this requires further exploration. In particular, documenting clinician review and edit of AI-generated handoff summaries should generate new knowledge about what clinical information is most important for clinicians and patients at the point of care transition.

9. Conclusion

AI-augmented clinical handoff systems integrate large language models, machine learning deterioration detection, and structured communication frameworks into routine EHR workflows, addressing a known source of preventable patient harm. Trials across multiple hospitals using these systems demonstrated a reduced documentation burden, with a Documentation Efficiency Index of 50.2-55.1% across participating hospitals. Clinicians maintain oversight of the augmentation model. LLMs generated a mean Composite Handoff Quality Score (CHQS) of 0.712. Physicians had a CHQS of 0.638. BERTScore Precision was 0.859. ROUGE-2 was 0.322. Critical

patient safety risks were not identified at scale. The machine learning early warning system extends the clinical safety period to 11 hours prior to deterioration. Alerts for patients at risk are generated as part of routine handoffs. The three metrics presented here, composite note quality, and AI risk stratification, are useful benchmarking tools to guide institutions in evaluating and deploying these tools. Further support on interoperability infrastructure, health equity in access, and synchronizing regulatory standards can help ease the deployment of these tools. The promise of AI-augmented handoffs will be realized only if clinicians, informaticists, system developers, and regulators work together to improve the human dimensions of the handoff (relationship, empathy, and trust) while retaining the strengths of technology (consistency and efficiency).

The author declares no conflict of interest. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

Thiyagarajan Palaniyappan conceived the study, conducted the literature synthesis, developed the quantitative frameworks, and wrote the manuscript in its entirety. The author read and approved the final manuscript.

References

1. Barker, W., Chang, W., Everson, J., Gabriel, M., Patel, V., Richwine, C., & Strawley, C. (2024). The evolution of health information technology for enhanced patient-centric care in the United States: data-driven descriptive study. Journal of Medical Internet Research, 26, e59791. https://www.jmir.org/2024/1/e59791/
2. Bednarczyk, L., Zaghir, J., Ehrsam, J., Tcherepanova, M., Skalafouris, C., Gariani, K., Geslin, C., et al. (2026). Evaluating patient safety risks in generative AI: Development and validation of a FMECA framework for generated clinical content. arXiv preprint arXiv:2605.04085. https://arxiv.org/abs/2605.04085
3. Chen, R. J., Wu, M. S., Tsai, L. W., Chang, S. S., Shen Hsiao, S. T., & Lo, Y. S. (2026). Integrating a large language model to streamline nursing handover documentation across multiple hospitals in Taiwan: Development and implementation study. Journal of Medical Internet Research, 28, e81604. https://www.jmir.org/2026/1/e81604/
4. Churpek, M. M., Carey, K. A., Snyder, A., Winslow, C. J., Gilbert, E., Shah, N. S., Patterson, B. W., et al. (2025). Multicenter development and prospective validation of eCARTv5: A gradient-boosted machine-learning early warning score. Critical Care Explorations, 7(4), e1232. https://journals.lww.com/ccejournal/fulltext/2025/04000/multicenter_development_and_prospective_validation.10.aspx
5. Doherty, R., Kazley, A. S., Karp, E., & Ferrand, J. (2025). A preliminary conceptual framework of clinical documentation burden: Exploratory factor analysis investigating usability, effort, and perceived burden among health care providers. Applied Clinical Informatics, 16(05), 1815-1827. https://www.thieme-connect.com/products/ejournals/html/10.1055/a-2751-1896
6. Edelson, D. P., Churpek, M. M., Carey, K. A., Lin, Z., Huang, C., Siner, J. M., Johnson, J., Krumholz, H. M., & Rhodes, D. J. (2024). Early warning scores with and without artificial intelligence. JAMA Network Open, 7(10), e2438986. https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2824885
7. Fakha, A., & Boonstra, A. (2025). Artificial intelligence in transitional care: practice, promise, and pitfalls - a scoping review. Frontiers in Digital Health, 7, 1690223. https://doi.org/10.3389/fdgth.2025.1690223
8. Hartman, V., Zhang, X., Poddar, R., McCarty, M., Fortenko, A., Sholle, E., Sharma, R., Campion, T., Jr., & Steel, P. A. D. (2024). Developing and evaluating large language model-generated emergency medicine handoff notes. JAMA Network Open, 7(12), e2448723. https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2827327
9. Humphrey, K. E., Sundberg, M., Milliren, C. E., Graham, D. A., & Landrigan, C. P. (2022). Frequency and nature of communication and handoff failures in medical malpractice claims. Journal of Patient Safety, 18(2), 130-137. https://doi.org/10.1097/pts.0000000000000937
10. Jorro-Baron, F., Suarez-Anzorena, I., Burgos-Pratx, R., Demaio, N., Penazzi, M., Rodriguez, G., Rodriguez, A.-P., et al. (2021). A multicentre study for the reduction of adverse events in Argentine pediatric intensive care units using a program to improve handoffs: A stepped wedge trial. Pediatric Critical Care Medicine, 22(Supplement 1), 279-280. https://www.ovid.com/jnls/pccmjournal/abstract/10.1097/01.pcc.0000740588.64220.96

11. Marquez, M., Gonzalez, A., Moufarrej, Y., & Vijayan, V. (2024). Improving patient handoffs and transitions in care among residents: A chief resident-led initiative. Cureus, 16(11). https://assets.cureus.com/uploads/original_article/pdf/315493/20241209-3977919-lr6qqd.pdf
12. McElroy, L. M., Collins, K. M., Koller, F. L., Khorzad, R., Abecassis, M. M., Holl, J. L., & Ladner, D. P. (2015). Operating room to intensive care unit handoffs and the risks of patient harm. Surgery, 158(3), 588-594. https://www.sciencedirect.com/science/article/abs/pii/S0039606015003281
13. Melnick, E. R., Fong, A., Nath, B., Williams, B., Ratwani, R. M., Goldstein, R., O'Connell, R. T., Sinsky, C. A., Marchalik, D., & Mete, M. (2021). Analysis of electronic health record use and clinical productivity and their association with physician turnover. JAMA Network Open, 4(10), e2128790. https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2784810
14. Nanji, K. C., Seger, D. L., Slight, S. P., Amato, M. G., Beeler, P. E., Her, Q. L., Dalleur, O., et al. (2018). Medication-related clinical decision support alert overrides in inpatients. Journal of the American Medical Informatics Association, 25(5), 476-481. https://doi.org/10.1093/jamia/ocx115
15. Reyes, J. A., Greenberg, L., Amdur, R., Gehring, J., & Lesky, L. G. (2016). Effect of handoff skills training for students during the medicine clerkship: A quasi-randomized study. Advances in Health Sciences Education, 21(1), 163-173. https://link.springer.com/article/10.1007/s10459-015-9621-1
16. Small, W. R., Austrian, J., O'Donnell, L., Burk-Rafel, J., Hochman, K. A., Goodman, A., Zaretsky, J., et al. (2025). Evaluating hospital course summarization by an electronic health record-based large language model. JAMA Network Open, 8(8), e2526339. https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2837483
17. Starmer, A. J., Spector, N. D., Srivastava, R., West, D. C., Rosenbluth, G., Allen, A. D., Noble, E. L., et al. (I-PASS Study Group). (2014). Changes in medical errors after implementation of a handoff program. New England Journal of Medicine, 371(19), 1803-1812. https://www.nejm.org/doi/full/10.1056/NEJMsa1405556
18. Van Veen, D., Van Uden, C., Blankemeier, L., Delbrouck, J.-B., Aali, A., Bluethgen, C., Pareek, A., et al. (2024). Adapted large language models can outperform medical experts in clinical text summarization. Nature Medicine, 30(4), 1134-1142. https://www.nature.com/articles/s41591-024-02855-5
19. Wiesenfeld, B. M., & Kellogg, K. C. (2025). Generative organizational learning: Affordances for new modes of knowledge search, creation, transfer, and forgetting with LLMs. Strategic Organization. https://doi.org/10.1177/14761270251408580
20. Wong, A., Otles, E., Donnelly, J. P., Krumm, A., McCullough, J., DeTroyer-Cooley, O., Pestrue, J., et al. (2021). External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. JAMA Internal Medicine, 181(8), 1065-1070. https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2781307
21. World Health Organization. (2023). Patient safety fact sheet. WHO. https://www.who.int/news-room/fact-sheets/detail/patient-safety