REVIEW **Open access**

Patient Safety Event Analytics: Narrative Mining, Taxonomy Development, and Learning Health System Integration

Andrei Popescu[1*], Mihai Ionescu[1], Elena Stan[2]

Abstract

The rapid expansion of unstructured narrative data within patient safety event (PSE) reporting systems presents both a valuable source of safety intelligence and a major analytical challenge for healthcare organizations. Traditional manual review processes are labor-intensive, subjective, and incapable of scaling to the vast volumes of incident reports generated across modern health systems. Artificial intelligence techniques, particularly natural language processing and machine learning, provide scalable approaches for extracting meaningful insights from these narratives. This narrative review synthesizes advances in AI-enabled PSE analytics across three interconnected domains: automated narrative mining, data-driven taxonomy development, and integration within learning health systems that transform safety data into continuous improvement cycles. Evidence indicates that AI methods can improve event classification, accelerate detection of emerging safety signals, and reduce the analytical burden on safety teams. However, challenges remain regarding model generalisability, interpretability, and governance. AI-driven narrative analytics is emerging as a foundational component of next-generation safety intelligence infrastructures.

Keywords Natural language processing, Artificial intelligence, Patient safety event analytics, Narrative mining, Taxonomy development, Learning health systems

*Correspondence:
Andrei Popescu
andrei.popescu@gmail.com
[1] Department of Health Informatics, Faculty of Medicine, University of Bucharest, Bucharest, Romania
[2] Department of Digital Systems Engineering, Politehnica University of Bucharest, Bucharest, Romania

Introduction

The persistent challenge of unstructured safety narratives in modern healthcare

Patient safety remains a central concern of contemporary healthcare policy and research more than two decades after the landmark Institute of Medicine report *To Err Is Human* revealed the magnitude of preventable harm within healthcare systems. Despite substantial progress in safety science, preventable adverse events continue to affect millions of patients worldwide each year, generating considerable clinical, ethical, and economic consequences. A substantial proportion of the knowledge required to understand these events resides within patient safety event (PSE) reporting systems, which are widely implemented across hospitals, national health services, and regulatory surveillance infrastructures. These systems collect detailed descriptions of adverse events, near misses, and hazardous conditions reported by clinicians, administrators, and occasionally patients themselves [1-8].

While structured reporting fields capture standardized categories such as event severity, location, or clinical domain, the richest contextual information typically appears within the accompanying narrative descriptions. These free-text narratives document the sequence of clinical actions, environmental circumstances, communication failures, and latent system

factors that contributed to the event. Such narratives often contain subtle causal cues that cannot easily be represented through predefined checkboxes or categorical codes. Consequently, unstructured text has become one of the most valuable yet underutilized sources of safety intelligence within healthcare systems.

However, extracting actionable knowledge from these narratives presents significant operational challenges. Manual analysis of safety narratives requires extensive expert review, often performed by dedicated patient safety officers or multidisciplinary review committees. The scale of modern reporting systems renders this process inherently constrained. Large academic medical centers frequently generate tens of thousands of reports annually, and national reporting infrastructures may accumulate millions of records over time [4, 9-15]. The cognitive workload required to interpret such volumes of narrative text far exceeds the capacity of traditional manual review processes. Furthermore, human interpretation introduces additional variability due to differences in expertise, attention, and subjective judgment among reviewers. These limitations contribute to delayed recognition of emerging hazards, incomplete pattern detection, and inconsistent categorization of events across institutions [5].

As healthcare systems increasingly emphasize proactive risk management and continuous quality improvement, the inability to analyze narrative safety data systematically has become a critical barrier. Without scalable analytical approaches, valuable insights embedded in safety reports remain fragmented and inaccessible, limiting the capacity of health systems to identify systemic vulnerabilities and implement timely interventions.

Emergence of AI as a scalable analytical layer

Advances in artificial intelligence—particularly natural language processing (NLP)—have created new opportunities to address the analytical challenges associated with large-scale safety narratives. Early computational approaches to narrative analysis relied primarily on rule-based keyword detection and manually curated lexicons. Although these techniques enabled basic classification of safety events, they were limited in their ability to capture semantic nuance or contextual meaning. The development of modern machine-learning approaches has substantially expanded the analytical capabilities available for interpreting clinical narratives.

Supervised learning models trained on annotated safety reports have demonstrated the capacity to automatically classify event types, detect contributing factors, and identify latent safety themes embedded within free-text descriptions. Studies applying traditional machine-learning algorithms—including support vector machines, random forests, and gradient-boosting models—have shown strong performance in categorizing incident reports when trained on sufficiently curated datasets [1, 6]. More recent work has incorporated deep learning architectures and contextual language embeddings, enabling models to capture complex linguistic relationships within clinical narratives. These models can detect subtle patterns of meaning that extend beyond simple keyword matching, improving both sensitivity and specificity in event classification tasks.

Comparative analyses indicate that automated NLP systems can process narrative safety reports at a scale and speed unattainable through manual review alone. In many cases, computational pipelines reduce annotation workloads by several orders of magnitude while maintaining or improving classification accuracy [1, 2]. Hybrid analytical pipelines combining rule-based preprocessing with neural language models have proven particularly effective, as structured heuristics can guide model attention toward clinically relevant features while deep representations capture contextual semantics. Such approaches facilitate rapid triage of incoming reports, enabling safety teams to focus their attention on high-risk or novel event patterns rather than routine incidents.

The scalability of AI-driven narrative analysis also introduces new possibilities for cross-institutional learning. When applied to aggregated safety datasets spanning multiple healthcare organizations, NLP models can detect recurrent patterns that may remain invisible within isolated local reporting systems. This capacity for large-scale pattern recognition positions AI as a powerful analytical layer capable of transforming safety reporting infrastructures from passive repositories into active knowledge-generation systems.

Taxonomy development: from static hierarchies to dynamic ontologies

The effectiveness of automated safety analytics depends not only on the ability to process narrative text but also on the conceptual frameworks used to categorize events. Traditional patient safety taxonomies were designed primarily for standardization rather than analytical discovery. Many widely adopted classification schemes rely on hierarchical category structures that define a fixed set of event types and contributing factors. While these frameworks provide valuable consistency for reporting and regulatory oversight, they often struggle to represent the evolving and context-dependent nature of clinical safety events.

Healthcare delivery environments are highly dynamic systems in which new technologies, treatment protocols, and organizational practices continually reshape patterns of risk. Static taxonomies may therefore fail to capture emerging categories of safety events or subtle variations within established event types. In practice, safety officers frequently encounter narrative reports that do not fit neatly within predefined classification categories, leading to forced categorization or inconsistent coding practices. These limitations reduce the analytical precision of safety datasets and complicate cross-institutional comparisons.

Recent research has explored data-driven approaches to taxonomy development that leverage machine-learning techniques to identify patterns directly within narrative data. Topic modeling, clustering algorithms, and embedding-based similarity analysis can uncover latent thematic structures within large collections of safety reports [3, 7, 12]. Rather than imposing rigid hierarchical categories, these methods allow classification systems to evolve in response to empirical data. Through iterative human-in-the-loop validation processes, domain experts can refine computationally generated clusters into clinically meaningful safety categories.

Dynamic taxonomy construction offers several advantages for safety analytics. First, it enables the identification of emerging event types that may not yet be represented within existing classification schemes. Second, it allows the detection of nuanced subcategories within broad event classes, improving the granularity of safety analysis. Third, adaptive taxonomies facilitate more meaningful benchmarking across healthcare institutions by aligning classification frameworks with real-world patterns of safety events [2, 13]. In this way, the integration of machine learning into taxonomy development transforms safety classification from a static reporting structure into a continuously evolving knowledge framework.

Integration into learning health systems: closing the loop

The analytical potential of narrative mining and dynamic taxonomy construction becomes most impactful when integrated within broader learning health system (LHS) architectures. The LHS paradigm conceptualizes healthcare organizations as continuously adaptive systems in which data generated during routine clinical care are systematically transformed into knowledge that informs future practice [7, 9, 16, 17]. Within this framework, the rapid analysis of safety narratives plays a crucial role in enabling real-time learning from operational experience.

AI-enabled narrative analytics provides the computational engine required to operationalize this learning cycle. Automated pipelines can ingest newly submitted safety reports, classify events, detect emerging patterns, and generate structured insights that inform risk mitigation strategies. When integrated with clinical dashboards and governance workflows, these insights can be rapidly communicated to frontline clinicians, quality improvement teams, and institutional leadership. Such feedback loops allow organizations to respond proactively to emerging safety risks rather than relying solely on retrospective analysis.

Empirical studies examining the integration of NLP-based safety analytics within LHS infrastructures have reported measurable improvements in organizational learning capacity. Automated classification systems accelerate the time required to identify safety trends, enabling earlier detection of recurrent hazards and faster implementation of corrective interventions [8, 14]. Moreover, the transparency of algorithmically generated safety insights can strengthen institutional safety culture by providing clinicians with timely, evidence-based feedback regarding system vulnerabilities. As safety

analytics become increasingly embedded within operational workflows, the boundary between data collection and quality improvement begins to dissolve, reinforcing the self-improving dynamics characteristic of mature learning health systems.

Systems-level perspective and original synthesis framework

Although the literature on AI-enabled safety analytics has expanded rapidly, existing studies often examine isolated components of the analytical pipeline rather than the broader infrastructural ecosystem required for sustainable implementation. Technical research may focus on model performance or algorithmic architecture, while health systems research tends to emphasize governance, workflow integration, or organizational learning processes. This fragmentation obscures the interdependencies among technological, operational, and governance dimensions of safety intelligence infrastructures.

To address this conceptual gap, the present review synthesizes the literature through a systems-level perspective that organizes existing evidence into four interconnected infrastructural layers. The first layer concerns the ingestion and preprocessing of raw narrative safety data, including text normalization, de-identification, and linguistic feature extraction. The second layer encompasses the development of analytical models and the evolution of adaptive taxonomies capable of interpreting narrative content at scale. The third layer examines the operational deployment of these models within clinical and administrative workflows, emphasizing integration with safety monitoring dashboards, reporting systems, and decision-support tools. The fourth layer addresses governance, feedback mechanisms, and ethical oversight structures that ensure responsible deployment of AI-driven safety analytics.

Viewing narrative safety analysis through this multi-layered infrastructural lens highlights the necessity of coordinated development across technical, organizational, and regulatory domains. Advances in machine learning alone cannot transform safety intelligence if models remain disconnected from operational decision pathways. Similarly, governance frameworks cannot fully leverage safety data without robust analytical capabilities capable of extracting actionable insights from narrative reports. Effective safety intelligence infrastructures, therefore, depend on the tight coupling of data ingestion, analytical modeling, operational integration, and governance oversight into a cohesive ecosystem. **Table 1** summarises the functional infrastructure required to operationalize AI-enabled narrative safety analytics across the full Learning Health System pipeline.

Table 1. Functional infrastructure of AI-enabled patient safety narrative analytics across the learning health system pipeline

Infrastructure layer	Core analytical function	Key technologies	Operational outputs	Safety intelligence contribution
Narrative data ingestion	Capture and normalize unstructured safety narratives from heterogeneous reporting systems	Clinical NLP preprocessing, entity recognition, and de-identification pipelines	Cleaned narrative datasets and structured linguistic features	Enables scalable analysis of high-volume safety narratives
Narrative mining and taxonomy development	Extract semantic meaning and construct adaptive safety classification frameworks	Transformer language models, topic modeling, and embedding-based clustering	Event classification, contributing factor identification, and emergent taxonomy categories	Converts free-text narratives into structured safety knowledge
Safety signal detection	Identify recurrent hazards and latent safety patterns across reports	Pattern detection algorithms, anomaly	Safety alerts, risk scores, and trend analyses	Enables early identification of

		detection, and predictive risk models		systemic safety threats
Clinical decision support integration	Embed safety intelligence within operational healthcare workflows	Safety dashboards, EHR integration modules, and automated triage systems	Prioritized incident review lists, workflow alerts, and governance reports	Translates analytics into actionable clinical oversight
Intervention and learning cycle	Implement corrective actions and evaluate outcomes within a learning health system	Quality-improvement workflows and protocol update mechanisms	Process modifications, training interventions, and system redesign	Converts safety insights into organizational learning
Governance and model recalibration	Ensure ethical oversight, model reliability, and taxonomy stability	Bias auditing tools, drift detection algorithms, and human-AI review committees	Model updates, taxonomy revisions, and governance reports	Maintains trust, accountability, and long-term analytical validity

The sections that follow examine each of these layers in detail, synthesizing findings from high-impact peer-reviewed literature to articulate a comprehensive model of AI-enabled safety intelligence within contemporary healthcare systems.

Landscape of AI in Healthcare Systems and Analytics

Layer 1: Data ingestion and narrative preprocessing

Effective PSE analytics begins with robust ingestion pipelines capable of handling heterogeneous reporting formats, multilingual text, and variable narrative quality. Studies consistently highlight the necessity of domain-specific preprocessing pipelines—including negation detection, temporal reasoning, and medical entity recognition—before downstream modeling [1, 4, 6]. When properly implemented, these steps dramatically improve signal-to-noise ratios and enable subsequent taxonomy development to operate on cleaner semantic representations [2, 12].

Layer 2: Model development and automated taxonomy construction

A rich body of work has progressed from classical bag-of-words and support-vector-machine approaches to transformer-based architectures that capture long-range contextual dependencies within safety narratives [1, 3, 13]. Comparative analyses demonstrate that fine-tuned clinical language models outperform generic NLP tools in both classification accuracy and taxonomy coherence [2, 15]. Taxonomy development has evolved from expert-curated hierarchies toward semi-supervised methods that iteratively refine categories as new data accrue, producing living ontologies that adapt to emerging threats such as novel HIT failure modes or pandemic-related safety events [7, 10, 14].

Layer 3: Deployment within operational and clinical workflows

Successful translation requires seamless embedding of AI outputs into existing safety and quality infrastructures. Real-world deployments illustrate that NLP-derived classifications can automatically populate dashboards, trigger peer-review workflows, and prioritize high-harm events for human investigation [4, 11, 16]. Integration with electronic health record systems further enables closed-loop alerting, whereby detected safety patterns prompt immediate workflow adjustments [9, 18].

Layer 4: Governance, feedback, and continuous recalibration

Sustained performance hinges on robust governance frameworks that address model drift, bias amplification, and explainability requirements. Evidence underscores the value of human–AI collaborative review committees that regularly audit and retrain models using newly labeled data, thereby maintaining taxonomic validity and clinical trust [5, 8, 19]. Ethical oversight mechanisms—particularly those ensuring equitable performance across diverse patient populations and reporting cultures—are increasingly recognized as non-negotiable components of mature LHS implementations [7, 17].

Cross-study comparative insights and emergent patterns

Synthesis across the 19 studies reveals three convergent findings: (i) NLP consistently augments rather than replaces human expertise, (ii) dynamic taxonomies outperform static schemes in longitudinal surveillance, and (iii) LHS integration is the critical multiplier that converts analytical capability into measurable safety gains [1, 3, 5, 8, 10]. Heterogeneity in reporting systems and model architectures nonetheless highlights the need for standardized evaluation benchmarks and open datasets to accelerate progress [6, 12, 15].

Intelligent clinical decision and closed-loop healthcare systems

From detection to actionable clinical intelligence

The transition from retrospective PSE analysis to prospective clinical decision support represents the apex of AI-enabled safety infrastructure. When narrative-derived taxonomies feed directly into clinical decision-support engines, systems can anticipate risk trajectories and recommend preventive interventions in real time [9, 11, 18]. This closed-loop paradigm shifts safety from reactive reporting to proactive risk mitigation.

Conceptual clinical intelligence pipeline

A unifying interpretive model formalizes the process as a continuous loop:

Clinical intelligence loop = Data ingestion → Narrative mining and taxonomy application → Risk inference → Decision support → Intervention execution → Outcome feedback → Model recalibration

This conceptual pipeline, synthesized from the reviewed evidence, emphasizes bidirectional information flow and continuous recalibration as essential for sustained LHS performance [7, 10, 14, 17].

Human–AI fusion in safety decision-making

Empirical implementations demonstrate that hybrid human–AI teams achieve superior sensitivity and specificity compared with either component alone [5, 8, 13, 19]. Governance structures that embed explainable AI outputs within multidisciplinary safety committees further enhance adoption and clinical relevance [4, 16].

Evidence of impact and scalability

Across diverse settings, AI-augmented closed-loop systems have demonstrated earlier detection of safety signals, reduced event recurrence, and measurable improvements in organizational safety culture metrics [1, 3, 11, 15]. Scalability to national or multi-institutional LHS platforms remains an active frontier, with early evidence suggesting that federated learning and standardized taxonomy mappings will be key enablers [7, 9, 18]. **Figure 1** illustrates the integrated architecture through which narrative mining, adaptive taxonomy development, and governance feedback collectively transform patient safety narratives into continuous learning cycles within a learning health system.

Figure 1. AI-enabled patient safety intelligence loop integrating narrative mining, adaptive taxonomy construction, and learning health system feedback cycles.

Unstructured patient safety event (PSE) narratives are ingested and preprocessed through domain-specific text pipelines. Natural language processing models then extract semantic features and construct adaptive taxonomies that classify event types and contributing factors. These structured representations feed safety-signal detection engines that generate risk scores and alerts for clinical governance workflows. Decision support dashboards and automated review triggers enable operational responses such as workflow adjustments or policy updates. Outcomes and human review annotations are subsequently fed into governance mechanisms that recalibrate models and refine taxonomies, completing a continuous learning cycle within the learning health system.

Challenges and limitations

Technical and methodological barriers in narrative mining

Despite substantial progress, current NLP and text-mining approaches for patient safety event (PSE) narratives continue to face inherent limitations in handling linguistic variability, ambiguity, and domain-specific jargon [1, 2, 6]. Models trained on one institution's reporting culture often exhibit degraded performance when deployed elsewhere, highlighting persistent generalisability gaps across heterogeneous health systems [3, 14, 15]. Early rule-based and classical machine-

learning pipelines required extensive feature engineering, while even modern transformer architectures struggle with rare event types and implicit causal relationships embedded in clinician narratives [4, 12, 20].

Taxonomy evolution: rigidity versus drift

Dynamically constructed taxonomies, although superior to static frameworks, introduce new risks of concept drift and category proliferation over time [7, 10, 21]. Without rigorous human-in-the-loop governance, data-driven ontologies can become fragmented or biased toward frequently reported event clusters, marginalizing low-frequency but high-severity harms such as diagnostic delays or health-information-technology failures [5, 8, 10]. Comparative analyses across the literature reveal that taxonomy stability remains an under-addressed challenge, with few studies reporting longitudinal recalibration protocols beyond initial validation [13, 16].

Integration into learning health systems: the translation gap

Embedding AI-derived insights into operational LHS workflows has proven more complex than anticipated. Real-world deployments frequently encounter resistance from safety committees accustomed to manual review, resulting in “alert fatigue” and under-utilization of automated classifications [6, 9, 11, 17]. Moreover, the closed-loop feedback required for continuous model recalibration demands seamless bidirectional data flows between reporting systems, electronic health records, and quality-improvement platforms—interoperability that remains fragmented in most organizations [7, 10, 14, 18].

Governance, ethics, and trust

Explainability and accountability constitute critical unresolved barriers. Black-box deep-learning models applied to PSE narratives raise legitimate concerns regarding bias amplification, particularly when training data reflect historical under-reporting in certain demographic or clinical contexts [5, 14, 16, 19, 20]. Regulatory frameworks have yet to mature sufficiently to provide clear guidance on validation standards, post-market surveillance of safety analytics algorithms, or liability allocation when AI-generated taxonomies influence clinical decisions [8, 15, 21]. Ethical oversight mechanisms, while increasingly discussed, are rarely operationalized at scale [6, 12, 16].

Resource and organizational constraints

Implementation costs, including computational infrastructure, specialized personnel, and ongoing model maintenance, remain prohibitive for smaller or resource-limited health systems [3, 6, 13, 14]. The requirement for continuous labeled data to sustain taxonomy accuracy further exacerbates workload on already overburdened patient-safety teams [2, 10]. Cross-study synthesis underscores that technological capability has outpaced organizational readiness, creating a persistent “last-mile” problem in realizing the full safety benefits of narrative analytics [7, 11, 16, 17].

Future research directions

Advancing generalizable and explainable models

One of the most pressing priorities for future research involves improving the generalisability of artificial intelligence models applied to patient safety event (PSE) narratives. Many existing models are trained on datasets derived from a single institution or a limited number of reporting systems, which restricts their ability to transfer effectively across healthcare environments with different documentation practices, clinical terminologies, and reporting cultures. Foundation models pre-trained on large, multi-institutional corpora of PSE narratives offer a promising solution to this limitation, as such models can learn linguistic representations that capture broader patterns of safety-related discourse across heterogeneous healthcare settings [1]. By leveraging large-scale pre-training strategies similar to those used in contemporary language models, researchers may develop safety analytics tools capable of maintaining robust performance when deployed across diverse clinical infrastructures.

In parallel with improving generalisability, the interpretability of AI-driven safety analytics remains a critical requirement for clinical adoption. Healthcare professionals must be able to understand how algorithmic systems arrive at their conclusions, particularly when those conclusions influence safety governance decisions or trigger operational interventions. Explainability techniques such as attention visualization can illuminate which narrative segments contributed most strongly to model predictions, enabling safety analysts to verify that the model is relying on clinically meaningful textual cues rather than spurious correlations [3]. Complementary approaches based on counterfactual reasoning can further enhance transparency by demonstrating how small changes in narrative content would alter model classifications, thereby revealing the decision boundaries underlying the model's reasoning process [6].

Another promising direction involves the integration of hybrid human–AI learning frameworks. Rather than relying solely on static annotated datasets, active-learning approaches allow models to identify instances of high uncertainty and selectively request human annotations for those cases. This strategy enables efficient allocation of expert review resources while continuously improving model performance over time [2]. Such adaptive annotation pipelines are particularly valuable in safety reporting environments where new event types may emerge or where narrative styles evolve in response to changing clinical practices. Through targeted human oversight, active-learning systems can maintain the fidelity of evolving taxonomies while substantially reducing the overall labeling burden associated with training large-scale models [12]. Recent methodological work suggests that combining uncertainty-based sampling with semantic diversity metrics can further optimize the selection of narratives requiring expert review, improving both annotation efficiency and model robustness [21].

Dynamic, self-evolving taxonomies

Future research must also address the limitations of static safety classification frameworks by developing mechanisms for dynamic taxonomy evolution. Traditional taxonomies were primarily designed to standardize reporting structures rather than to capture the full complexity of safety events within modern healthcare systems. As clinical workflows, technologies, and organizational structures evolve, new patterns of risk frequently emerge that fall outside the boundaries of established classification categories. Machine-learning–driven ontology learning offers a promising avenue for addressing this challenge by enabling taxonomies to adapt in response to empirical patterns identified within narrative data streams [4].

Temporal topic modeling techniques provide one potential approach for detecting emerging safety themes over time. By analyzing changes in narrative topic distributions across sequential reporting periods, these models can reveal the gradual emergence of new event clusters or shifts in the prevalence of existing safety concerns [7]. When combined with causal discovery algorithms capable of identifying potential relationships between contributing factors and outcomes, such methods may allow healthcare organizations to detect emerging hazards before they manifest as widespread harm [10]. This capacity for early signal detection is particularly valuable in complex clinical environments where subtle process deviations can accumulate into significant systemic vulnerabilities.

Seamless learning health system integration and closed-loop architectures

The translation of AI-driven safety analytics from research prototypes into operational healthcare environments requires closer integration with learning health system (LHS) infrastructures. Although numerous studies have demonstrated the technical feasibility of narrative analysis using NLP, relatively few investigations have examined how these analytical tools perform when embedded within real-world clinical workflows. Future research must therefore prioritize prospective implementation studies that evaluate the complete lifecycle of safety intelligence generation—from narrative ingestion and algorithmic analysis to the deployment of corrective interventions and measurement of resulting safety outcomes [8].

A key methodological priority involves quantifying the end-to-end cycle time required for safety learning within AI-augmented systems. Traditional safety review processes often require weeks or months to detect recurring hazards due to the manual nature of report analysis. AI-enabled pipelines have the potential to compress this timeline dramatically by automating initial classification and signal detection processes. Prospective LHS studies should measure the time

required for narratives to move from initial submission through analytical processing, governance review, and implementation of corrective actions, thereby providing empirical evidence regarding the operational impact of automated safety analytics [9]. Such evaluations will help determine whether AI integration truly accelerates the learning cycles envisioned within the LHS paradigm.

Another promising direction involves the use of federated learning approaches that enable collaborative model development across institutions without requiring centralized data sharing. In healthcare environments where patient safety reports often contain sensitive information, federated learning architectures allow models to be trained across distributed datasets while preserving institutional data sovereignty. Each participating organization contributes model updates derived from its local data, which are then aggregated to produce a globally improved model without transferring raw narratives between institutions [6]. This strategy could facilitate the development of national or international safety intelligence networks capable of learning from aggregated experience across diverse healthcare systems. Early experimental implementations suggest that federated training can achieve performance levels comparable to centralized approaches while maintaining stronger privacy protections and regulatory compliance [14].

Ethical, equitable, and regulatory frameworks

As AI-driven safety analytics become increasingly embedded within healthcare infrastructures, the ethical and regulatory dimensions of these technologies require careful consideration. Patient safety reporting systems frequently contain sensitive clinical information as well as narrative descriptions that may reference staff actions, communication breakdowns, or organizational factors. The deployment of automated analytical tools within such contexts, therefore, raises important questions regarding data governance, algorithmic accountability, and transparency. Future research should focus on the development of governance frameworks that embed fairness auditing and bias detection mechanisms throughout the entire analytical pipeline [5].

Algorithmic bias represents a particularly important concern in safety analytics. If models are trained on datasets reflecting historical reporting patterns, they may inadvertently reproduce or amplify existing inequities in how safety events are documented or interpreted. For example, differences in reporting behavior across clinical departments or demographic groups may influence the distribution of narrative content available for model training. Researchers must therefore develop methodologies capable of detecting and mitigating such biases within both training data and model outputs [8]. Transparent reporting standards that clearly document model training procedures, evaluation metrics, and known limitations will also be essential for maintaining trust in AI-enabled safety systems [16].

Multimodal and predictive safety analytics

The next frontier of safety intelligence research lies in the integration of narrative analysis with multimodal clinical data sources. Safety narratives capture detailed contextual descriptions of events, but they represent only one component of the complex informational environment surrounding clinical care. Advances in biomedical informatics now enable the integration of textual narratives with physiologic waveforms, medical imaging, genomic data, and structured electronic health record variables. By combining these heterogeneous data streams, researchers may develop predictive safety models capable of identifying emerging hazards before adverse events occur [1].

Multimodal learning architectures that jointly process text, time-series physiologic signals, and structured clinical data have demonstrated substantial potential in other areas of healthcare analytics. Applying similar approaches to safety intelligence could enable the detection of early warning signals associated with deteriorating patient conditions, workflow disruptions, or communication failures. For example, patterns detected within narrative reports may be linked with abnormal physiologic trajectories or unusual clinical workflow patterns, creating richer representations of the causal pathways leading to safety events [7].

Ultimately, the clinical value of predictive safety analytics must be demonstrated through rigorous empirical evaluation. Large-scale pragmatic trials that assess whether AI-driven interventions reduce adverse-event rates in real-world

healthcare environments will be essential for establishing the effectiveness of these technologies. Such trials should measure not only predictive accuracy but also downstream clinical outcomes, operational efficiency, and staff acceptance of AI-assisted safety tools [11]. Evidence demonstrating a causal relationship between algorithmic safety monitoring and measurable improvements in patient outcomes would provide the strongest foundation for widespread adoption of AI-enabled safety intelligence infrastructures across healthcare systems [17].

Conclusion

The synthesis studies demonstrate that AI-enabled narrative mining, dynamic taxonomy development, and LHS integration have matured from experimental prototypes into foundational infrastructure for modern patient safety. The original four-layer systems perspective articulated herein—data ingestion, intelligence and taxonomy evolution, workflow deployment, and governance/feedback—reveals both the remarkable progress achieved and the critical interdependencies required for sustained impact. When properly orchestrated, these technologies transform voluminous unstructured narratives into timely, actionable intelligence that accelerates learning cycles and measurably enhances safety culture.

Nevertheless, technical, organizational, ethical, and resource barriers continue to constrain widespread realization of this vision. Addressing these challenges through generalizable models, self-evolving taxonomies, robust governance, and rigorous prospective evaluation will determine whether AI truly fulfills its promise as the engine of next-generation learning health systems. The conceptual clinical intelligence loop presented in **Figure 1** provides a unifying blueprint for future development and implementation efforts.

Ultimately, the greatest opportunity lies not in technological sophistication alone but in the deliberate alignment of AI capabilities with human expertise, organizational processes, and ethical imperatives. Only through such alignment can narrative mining and taxonomy development deliver on their potential to make preventable harm a relic of the past rather than an enduring feature of healthcare delivery.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 22 Sep 2025 Revised: 27 Dec 2025 Accepted: 10 Feb 2026
Published online: 10 July 2026

References

Young IJB, Luz S, Lone N. A systematic review of natural language processing for classification tasks in the field of incident reporting and adverse event analysis. Int J Med Inform. 2019;132:103971.
https://doi.org/10.1016/j.ijmedinf.2019.103971.

Tabaie A, Sengupta S, Pruitt ZM, Fong A. A natural language processing approach to categorise contributing factors from patient safety event reports. BMJ Health Care Inform. 2023;30(1):e100731.
https://doi.org/10.1136/bmjhci-2022-100731.

Boxley C, Fujimoto M, Ratwani RM, Fong A. A text mining approach to categorize patient safety event reports by medication error type. Sci Rep. 2023;13:18354.
https://doi.org/10.1038/s41598-023-45152-w.

Fong A, Adams KT, Gaunt MJ, Howe JL, Kellogg KM, Ratwani RM. Identifying health information technology related safety event reports from patient safety event report databases. J Biomed Inform. 2018;86:135-42.
https://doi.org/10.1016/j.jbi.2018.09.007.

Choudhury A, Asan O. Role of artificial intelligence in patient safety outcomes: systematic literature review. JMIR Med Inform. 2020;8(7):e18599.
https://doi.org/10.2196/18599.

Ozonoff A, Milliren CE, Fournier K, Welcher J, Landschaft A, Samnaliev M, et al. Electronic surveillance of patient safety events using natural language processing. Health Informatics J. 2022;28(4):14604582221132429.
https://doi.org/10.1177/14604582221132429.

Enticott J, Braaf S, Johnson A, Jones A, Teede HJ. Learning health systems using data to drive healthcare improvement and impact: a systematic review. BMC Health Serv Res. 2021;21(1):200.
https://doi.org/10.1186/s12913-021-06115-7.

Bates DW, Singh H. Two decades since To Err Is Human: an assessment of progress and emerging priorities in patient safety. Health Aff (Millwood). 2018;37(11):1736-43.
https://doi.org/10.1377/hlthaff.2018.0738.

Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med. 2018;1:18.
https://doi.org/10.1038/s41746-018-0029-1.

De Micco F, Di Palma G, Ferorelli D, De Benedictis A, Tomassini L, Tambone V, et al. Artificial intelligence in healthcare: transforming patient safety with intelligent systems-A systematic review. Front Med (Lausanne). 2025;11:1522554.
https://doi.org/10.3389/fmed.2024.1522554.

Fairie P, Santana MJ, Ahmed S. Categorising patient concerns using natural language processing. BMJ Health Care Inform. 2021;28(1):e100274.
https://doi.org/10.1136/bmjhci-2020-100274.

Luschi A, Nesi P, Iadanza E. Evidence-based clinical engineering: Health information technology adverse events identification and classification with natural language processing. Heliyon. 2023;9(11):e21723.
https://doi.org/10.1016/j.heliyon.2023.e21723.

Mertes PM, Morgand C, Barach P, Jurkolow G, Assmann KE, Dufetelle E, et al. Validation of a natural language processing algorithm using national reporting data to improve identification of anesthesia-related ADVerse evENTs: The "ADVENTURE" study. Anaesth Crit Care Pain Med. 2024;43(4):101390.
https://doi.org/10.1016/j.bja.2024.03.015.

Fong A. Realizing the Power of Text Mining and Natural Language Processing for Analyzing Patient Safety Event Narratives: The Challenges and Path Forward. J Patient Saf. 2021;17(8):e834-e836.
https://doi.org/10.1097/PTS.0000000000000805.

Platt JE, Raj M, Wienroth M. An Analysis of the Learning Health System in Its First Decade in Practice: Scoping Review. J Med Internet Res. 2020;22(3):e17026.
https://doi.org/10.2196/17026.

Chen H, Cohen E, Wilson D, Alfred M. A Machine Learning Approach with Human-AI Collaboration for Automated Classification of Patient Safety Event Reports: Algorithm Development and Validation Study. JMIR Hum Factors. 2024;11:e53378.
https://doi.org/10.2196/53378.

Fong A, Komolafe T, Adams KT, Cohen A, Howe JL, Ratwani RM. Exploration and Initial Development of Text Classification Models to Identify Health Information Technology Usability-Related Patient Safety Event Reports. Appl Clin Inform. 2019;10(3):521-7.
https://doi.org/10.1055/s-0039-1693427.

Ball R, Talal AH, Dang O, Muñoz M, Markatou M. Trust but Verify: Lessons Learned for the Application of AI to Case-Based Clinical Decision-Making From Postmarketing Drug Safety Assessment at the US Food and Drug Administration. J Med Internet Res. 2024;26:e50274.
https://doi.org/10.2196/50274.

Ramar K, Oxentenko AS, Dowdy SC. Transforming Health Care Through Quality and Safety. Mayo Clin Proc. 2025;100(8):1385-401.
https://doi.org/10.1016/j.mayocp.2025.02.012.

Kang H, et al. A novel schema to enhance data quality of patient safety event reports. Appl Clin Inform. 2017;8(4):1028-43.
https://doi.org/10.4338/ACI-2017-03-RA-0042.

Ramar K, Oxentenko AS, Dowdy SC. Transforming Health Care Through Quality and Safety. Mayo Clin Proc. 2025;100(8):1385-401.
https://doi.org/10.1016/j.mayocp.2025.04.012.