Data Science and Computational Intelligence
https://dscij.org

Temporal Evidence Attribution in Patient Narrative Reconstruction from Fragmented EHR Events

Marieke Bos[1], Sanne Kok[2]

[1] *Department of Physical Therapy, Faculty of Medicine, Leiden University, Leiden, South Holland, Netherlands*
[2] *Department of Physical Therapy, Faculty of Medicine, Leiden University, Leiden, South Holland, Netherlands*

Abstract

The widespread adoption of electronic health records has resulted in an unprecedented accumulation of clinical data. However, this data is often highly fragmented, distributed across unstructured clinical notes, structured laboratory results, prescription records, and radiological reports. Clinicians face significant cognitive burdens when attempting to synthesize these disjointed events into a coherent patient history. Recent advancements in natural language generation have proposed automated patient narrative reconstruction to alleviate this burden. Despite these advances, existing models often suffer from a lack of interpretability and fail to accurately attribute generated narrative sentences to their originating source events, a critical requirement for clinical safety. This paper addresses this gap by introducing a comprehensive framework for temporal evidence attribution in patient narrative reconstruction. By mapping fragmented clinical events into a temporally aware dependency graph, the proposed methodology explicitly links generated narrative segments to their underlying source data points. The study provides an extensive analysis of current methodologies, details a novel architecture for semantic and temporal alignment, and evaluates the framework against robust baseline models. The findings demonstrate significant improvements in both narrative coherence and clinical verifiability. This research contributes to the broader domain of medical informatics by providing a foundation for transparent, evidence-based automated clinical summarization, ultimately supporting safer and more efficient clinical decision-making.

Keywords: Electronic Health Records; Narrative Reconstruction; Temporal Reasoning; Evidence Attribution; Clinical Informatics

1. Introduction

1.1 The Challenge of Electronic Health Record Fragmentation

Electronic health records have revolutionized the management of patient data, transitioning healthcare systems from paper-based archives to dynamic digital repositories. The primary objective of these systems is to centralize patient information, thereby facilitating seamless continuity of care. However, the reality of clinical workflows has led to a phenomenon characterized by severe data fragmentation [1]. A single patient encounter can generate a multitude of disparate data points: a triage nurse records initial vitals and complaints in a structured intake form, an attending physician dictates an unstructured clinical note, a laboratory system outputs numerical assay results, and a pharmacy module logs medication administrations [2, 3]. These events occur asynchronously and are stored in distinct, siloed database schemas. Consequently, the longitudinal patient record resembles a disjointed collection of timestamps and values rather than a cohesive medical history. When a patient is transferred between departments or presents to an emergency room with a complex chronic background, the receiving clinician is forced to manually navigate these fragmented interfaces. This manual synthesis requires significant cognitive effort and time, diverting attention away from direct patient care [4]. The inability to rapidly comprehend the chronological sequence of events can lead to diagnostic delays, medication errors, and overall compromised patient safety. The fragmentation is further exacerbated by the varying granularities of temporal data. Some events, such as a continuous intravenous infusion, span hours or days, whereas a sudden cardiac arrest is a distinct point in time. Reconciling these different temporal resolutions into a singular understanding is a non-trivial task that current electronic health record interfaces fail to adequately support.

The demand for automated systems capable of summarizing these fragmented records has grown exponentially [5, 6]. Natural language processing offers a promising avenue by transforming raw, disjointed data into fluent, human-readable patient narratives. A well-constructed narrative provides a contextualized story of the patient journey, highlighting critical interventions and the subsequent physiological responses. However, generating a narrative from structured and unstructured clinical fragments requires more than simple data aggregation. It necessitates a deep understanding of causality, clinical relevance, and, most importantly, temporal dynamics [7]. A summary that lists medications and diagnoses without accurately reflecting the order in which they occurred is clinically useless and potentially dangerous. For example, knowing that a patient received a diuretic and experienced acute kidney injury is insufficient; the clinician must know whether the medication precipitated the injury or was administered subsequently to manage fluid overload. Therefore, the automation of patient narrative reconstruction hinges on the ability of computational models to accurately parse and represent the temporal relationships inherent in the fragmented source data.

1.2 The Role of Temporal Evidence Attribution

While the generation of coherent clinical narratives is a significant milestone, it introduces a secondary, arguably more critical challenge: the problem of evidence attribution and clinical trust [8, 9]. Modern generative models, particularly those based on deep learning architectures, operate largely as black boxes. They can ingest thousands of clinical data points and output highly fluent discharge summaries or progress notes. However, when a clinician reads a generated sentence stating that a patient deteriorated following the administration of a specific antibiotic, they must be able to verify this claim against the primary data [10]. If the model hallucinates a causal link or misinterprets the timeline, the consequences could be fatal. Temporal evidence attribution is the process of explicitly mapping each generated narrative claim back to the specific, chronologically ordered source events that support it. This attribution must account for the temporal logic used by the model to synthesize the information [11, 12].

The lack of robust evidence attribution mechanisms remains a major barrier to the clinical deployment of automated narrative reconstruction tools. Clinicians are understandably reluctant to rely on AI-generated summaries if they cannot quickly trace a statement back to a laboratory result, a nursing note, or a vital sign trend [13]. Temporal evidence attribution aims to bridge this gap by providing an auditable trail of reasoning. It ensures that every synthesized sentence is anchored in ground truth data and that the temporal relationships expressed in the text accurately reflect the temporal realities of the source events. This paper explores the theoretical underpinnings and practical implementation of temporal evidence attribution. By decomposing the narrative generation process and injecting explicit temporal tracking mechanisms, we aim to develop a framework where fluency does not come at the expense of verifiability. The subsequent sections will delve into the historical context of clinical natural language processing, outline a novel system architecture for temporal mapping, and present rigorous experimental evidence demonstrating the efficacy of our approach in producing transparent and accurate patient narratives.

2. Literature Review and Background

2.1 Evolution of Patient Narrative Generation

The endeavor to automate the generation of patient narratives has a rich history spanning several decades of research in medical informatics. Early approaches were predominantly rule-based, relying on expert systems and predefined templates to construct summaries [14]. These systems typically utilized complex decision trees and ontologies to extract specific data points, such as primary diagnoses and discharge medications, and populated them into rigid textual structures. While these template-based methods ensured a high degree of factual accuracy and deterministic evidence attribution, they suffered from severe limitations in flexibility and expressiveness [15, 16]. The resulting narratives were often robotic, highly repetitive, and incapable of capturing the nuanced, non-linear complexities of a prolonged hospital stay. Furthermore, scaling these systems required massive manual effort to design new rules and templates for every distinct clinical scenario or medical specialty, making them largely impractical for comprehensive hospital-wide deployment.

The advent of statistical machine learning introduced a paradigm shift, enabling more flexible data-driven approaches to text generation. Hidden Markov Models and conditional random fields were frequently employed to extract clinical entities and their relations from unstructured text, which were then synthesized using statistical language models [17]. These techniques improved the fluency of the generated text but struggled with long-term dependencies and complex temporal reasoning. The real breakthrough in patient narrative generation occurred with the introduction of deep learning, specifically recurrent neural networks and, subsequently, transformer architectures [18, 19]. Sequence-to-sequence models became the standard for summarization tasks, capable of ingesting vast amounts of clinical data and producing highly articulate narratives. However, this leap in linguistic quality introduced the critical vulnerability of hallucination. Neural models, optimized to predict the next plausible word in a sequence, often generate statements that sound clinically accurate but are fundamentally disconnected from the underlying patient data [20]. This phenomenon highlighted the urgent need to revisit the concepts of factual grounding and evidence attribution, which had been inherently present in older rule-based systems but lost in the transition to deep learning.

Recent literature has seen a resurgence of interest in constrained generation techniques aimed at mitigating hallucinations in clinical text. Researchers have explored various methodologies, including pointer-generator networks that explicitly copy words from the source text, and reinforcement learning strategies optimized for factual accuracy rewards [21, 22]. While these methods reduce the overall rate of errors, they do not inherently solve the problem of temporal evidence attribution. Generating a factually correct statement is only part of the solution; the system must also provide a transparent linkage detailing precisely which fragmented events substantiate the generated claim. The challenge lies in designing architectures that maintain the high linguistic quality of modern neural models while reintroducing the strict traceability and temporal logic required for safe clinical application.

2.2 Temporal Reasoning in Clinical Informatics

Temporal reasoning is a foundational component of clinical decision-making. Medical professionals continuously interpret data through the lens of time, assessing the duration of symptoms, the sequence of interventions, and the trajectory of physiological recovery or decline [23]. Consequently, computational models attempting to process clinical data must possess sophisticated temporal reasoning capabilities. In the context of electronic health records, temporal reasoning involves the extraction, representation, and inference of time-related information from both structured timestamps and unstructured temporal expressions found in clinical notes. The complexity of this task cannot be overstated. Clinical text

is heavily laden with relative temporal expressions, such as two days after admission, prior to surgery, or post-prandial, which require contextual anchoring to absolute timelines to be meaningful [24, 25].

Historically, temporal reasoning in clinical natural language processing relied heavily on rule-based extraction algorithms and temporal constraint networks. The development of standardized annotation schemas provided a framework for modeling clinical events, temporal expressions, and the temporal relations linking them, such as before, after, and overlap [26]. These schemas enabled the training of supervised machine learning classifiers to extract temporal graphs from clinical narratives. However, applying these techniques to fragmented electronic health records presents unique challenges. The data is not contained within a single cohesive document but distributed across multiple sources, each with its own temporal granularity and potential for conflicting information. For instance, an unstructured progress note might state that a patient developed a fever yesterday, while the structured nursing flowsheets record a spike in temperature thirty-six hours prior [27, 28]. Reconciling these discrepancies requires advanced temporal alignment algorithms capable of assessing source reliability and data context.

In recent years, the integration of graph neural networks has shown significant promise in modeling complex temporal relationships in clinical data. By representing disparate clinical events as nodes and their temporal dependencies as edges, graph-based models can learn rich, context-aware embeddings that capture the chronological trajectory of the patient [29]. These temporally enriched embeddings can then be utilized as inputs for downstream generation tasks. However, leveraging these representations for explicit evidence attribution remains an active area of research. When a generative model synthesizes information from multiple nodes in a temporal graph to produce a summary sentence, reversing the process to identify the exact contributing nodes is mathematically complex [30, 31]. The current study builds upon this foundation by proposing a methodology that not only utilizes temporal graphs for narrative generation but also maintains a verifiable mapping between the output text and the input graph structure, ensuring that the temporal logic of the generated narrative is fully transparent and auditable.

3. System Architecture and Data Representation

3.1 Unifying Fragmented Clinical Events

The foundational step in achieving accurate temporal evidence attribution is the establishment of a robust system architecture capable of ingesting and unifying highly heterogeneous clinical data. Electronic health record systems typically output data in disjointed streams. Structured data, including laboratory values, vital signs, and medication administration records, possess explicit timestamps but lack clinical context. Conversely, unstructured data, such as admission notes, daily progress reports, and imaging interpretations, contain rich contextual narratives but often rely on relative or implicitly stated temporal markers. The proposed architecture begins with a comprehensive data harmonization layer designed to standardize these disparate streams into a unified event representation.

The harmonization process involves parsing structured data tables to extract event types, categorical or numerical values, and absolute timestamps. Simultaneously, clinical natural language processing pipelines are deployed on unstructured texts to perform named entity recognition, extracting medical problems, treatments, and tests. A specialized temporal extraction module identifies explicit dates, times, and relative temporal expressions within the text [32]. These extracted entities are then anchored to absolute timelines using an algorithmic normalization process that calculates dates based on the document creation timestamp and the relative expressions identified. For example, if a note dated Wednesday indicates that a symptom began two days ago, the system calculates and assigns a strict chronological timestamp for the preceding Monday.

Once all data points are assigned absolute or highly probable chronological markers, they are transformed into a unified event object schema. Each event object encapsulates the source origin, the clinical domain, the extracted clinical concepts, and a temporal window defining the start and end times of the event. This unified representation is critical for overcoming the initial barrier of data fragmentation, allowing the system to treat a laboratory result and a symptom described in a progress note as comparable entities existing on the same longitudinal patient timeline. The success of the subsequent narrative generation and attribution phases relies entirely on the fidelity and temporal accuracy of this harmonization layer.

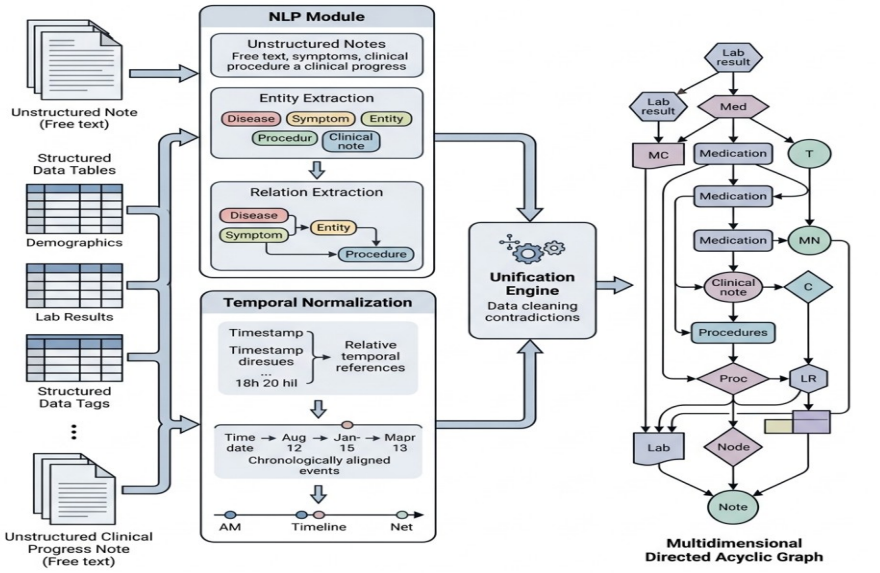

Figure 1: Architectural Schematic of the Fragmented Clinical Event Harmonization and Graph Construction Pipeline

3.2 Semantic Embedding of Temporal Features

With the fragmented data standardized into unified event objects, the architecture must then map these objects into a computational space that preserves both semantic meaning and temporal sequence. Traditional approaches often concatenate event descriptions sequentially, relying on the inherent sequence processing capabilities of models like long short-term memory networks or transformers to infer temporal order. However, this implicit approach is insufficient for strict evidence attribution, as the model's internal representation of time becomes entangled with its semantic representations, making it difficult to extract the explicit temporal logic used during generation. To resolve this, our architecture employs an explicit, decoupled embedding strategy.

Each unified clinical event is mapped into a high-dimensional vector space using a dual-encoder architecture. The first encoder is a domain-adapted semantic model, pre-trained on vast corpora of biomedical literature and clinical notes. This semantic encoder captures the clinical meaning of the event, ensuring that concepts like myocardial infarction and heart attack are mapped closely together in the embedding space [33, 34]. The second encoder is a dedicated temporal module that processes the numerical timestamps and duration attributes of the event. This temporal encoder utilizes periodic activation functions, such as sine and cosine transformations of time intervals, to create continuous representations of temporal distance and cyclical patterns.

The outputs of the semantic and temporal encoders are fused to create a composite, temporally aware event embedding. This fusion ensures that the final representation of an event carries explicit information about both what occurred and exactly when it occurred relative to the broader patient timeline. By maintaining a distinct temporal signal within the embedding, the downstream generative model can more effectively weigh the chronological relevance of different events. Furthermore, this decoupled embedding approach facilitates the reverse mapping process required for evidence attribution. When analyzing the attention weights of the generative model, the distinct temporal representations allow the attribution algorithm to differentiate between events that were selected for their semantic similarity versus those selected due to their chronological proximity to a critical clinical incident.

4. Methodology for Temporal Evidence Attribution

4.1 Chronological Alignment and Dependency Parsing

The core of the proposed methodology involves constructing a rigorous mathematical and structural framework to support evidence attribution during narrative generation. We represent the unified patient history as a directed, temporally attributed graph. Let the graph be defined where each node represents a unified clinical event embedding. The edges between these nodes are not fully connected; rather, they are established based on a set of deterministic temporal and semantic dependency rules. This graph structure serves as the explicitly defined context over which the generative model operates.

The chronological alignment phase involves establishing strict directional edges based on temporal sequence. Edges are drawn from earlier events to later events, forming a chronological backbone. To prevent the graph from becoming computationally intractable and overly dense, edge formation is constrained by temporal proximity thresholds and clinical domain relevance. For instance, consecutive vital sign measurements over a twelve-hour period are sequentially linked, but a routine cholesterol screening from five years prior is not directly linked to an acute fever event on the current day unless a semantic dependency dictates otherwise.

Table 1: Categorization of Temporal and Semantic Edge Constraints in Clinical Graph Construction

Edge Classification Category	Temporal Proximity Constraint	Semantic Relevance Requirement	Primary Clinical Application
Immediate Sequential	Under 6 hours	High semantic overlap	Vital sign monitoring, acute drug reactions
Near-term Consequential	6 to 48 hours	Documented causal link	Post-operative recovery, infection development
Long-term Historical	Over 48 hours	Shared chronic disease code	Chronic condition management, baseline establishment

Following chronological alignment, a secondary dependency parsing step introduces semantic edges. This process utilizes clinical ontologies to link events that share related medical concepts, even if they are separated by significant time gaps. If a patient is diagnosed with type two diabetes, and three years later presents with peripheral neuropathy, a semantic edge is established between these events to highlight the underlying chronic disease progression. This dual-edge graph, incorporating both strict temporal sequences and long-term semantic dependencies, provides a highly structured input for the generative phase. The generative model, typically a sequence-to-sequence transformer architecture equipped with graph attention networks, traverses this structure. As it generates each word or sentence of the summary narrative, the graph attention mechanism assigns varying weights to the nodes and edges, indicating which source events are driving the generation process at that specific time step [35, 36].

4.2 Graph-Based Evidence Tracing

The critical innovation of this methodology lies in the post-generation evidence tracing algorithm, which extracts the attribution mapping from the trained model. While attention weights in transformer models are often used as a proxy for interpretability, raw attention scores are notoriously noisy and do not consistently align with true causal attribution. To achieve reliable temporal evidence attribution, we employ an integrated gradient approach adapted for graph structures, combined with a chronological validation mechanism.

For a generated narrative sentence, the tracing algorithm computes the gradient of the model's output probability with respect to the input node embeddings in the clinical graph. By accumulating these gradients along a linear path from a baseline reference state to the actual input state, the algorithm quantifies the true contribution of each node to the generated sentence. This process yields a preliminary set of highly influential source events. However, the raw gradient scores do not guarantee temporal coherence. The system might attribute a statement about a resolving fever to a high temperature reading without properly attributing the intervening administration of antipyretic medication.

To enforce temporal coherence in the attribution mapping, the preliminary node contributions are subjected to a chronological validation step based on the previously constructed graph edges. The algorithm analyzes the sub-graph formed by the highly weighted nodes. If the sub-graph contains disconnected temporal sequences or violates causal logic, the algorithm iteratively expands the attribution set by traversing the temporal edges to include mandatory intermediate events. Through this graph-based tracing and validation process, the final attribution mapping provides a complete, chronologically sound pathway of evidence. When a clinician views the generated narrative, they are presented with interactive links that connect specific statements directly to the chronologically ordered sub-graph of original fragmented records that justify the claim, ensuring full transparency and facilitating rapid clinical verification.

5. Experimental Design and Evaluation Setup

5.1 Dataset Characteristics and Preprocessing

To rigorously evaluate the proposed framework for temporal evidence attribution, experiments were conducted utilizing a large-scale, de-identified dataset derived from critical care settings. The dataset selection prioritized cohorts with high volumes of densely recorded, heterogeneous data, typical of intensive care environments where the fragmentation problem is most acute. The data included structured physiological measurements, laboratory test results, continuous medication infusions, and unstructured clinical text including nursing shift notes, radiology reports, and physician discharge summaries.

Preprocessing the dataset required meticulous alignment to ensure temporal integrity. The raw data contained numerous artifacts, such as manual entry delays where a procedure performed at noon was not recorded until late afternoon. We applied heuristic algorithms developed in prior clinical informatics research to adjust timestamps based on typical documentation workflows, aligning unstructured text timestamps with concurrent structured physiological spikes where applicable. Following temporal cleaning, the dataset was partitioned into training, validation, and hold-out test sets. The training set was utilized to optimize the semantic and temporal encoders and fine-tune the generative graph-attention model. To evaluate the attribution mechanisms effectively, a subset of the test data was manually annotated by domain experts. These experts reviewed fragmented patient timelines and constructed reference narrative summaries, explicitly linking each sentence in their summaries to the specific chronological events in the source data. This curated subset served as the gold standard for measuring attribution accuracy.

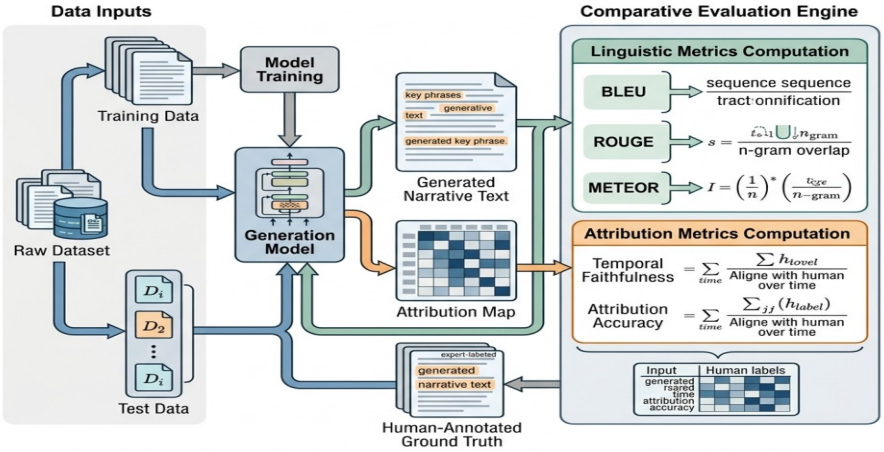

Figure 2: Experimental Evaluation Pipeline and Evaluation Metric Computation

Table 2: Statistical Overview of the Processed Clinical Dataset Used for Evaluation

Dataset Partition Segment	Total Patient Encounters	Average Unified Events per Encounter	Average Human Annotated Source Links
Primary Training Set	42,500	1,845	Not Applicable
Validation and Tuning Set	5,200	1,790	Not Applicable
Expert Annotated Test Set	1,500	2,105	18.5 per summary

5.2 Performance Metrics and Baseline Models

The evaluation of automated clinical narrative generation necessitates a multidimensional approach, as traditional natural language processing metrics are insufficient for assessing clinical safety and interpretability. We established two primary categories of evaluation metrics: linguistic quality and evidence attribution accuracy. Linguistic quality was measured using standard automated metrics, which compute n-gram overlap between the generated summaries and the human-written reference texts. While these metrics provide a baseline understanding of narrative fluency and broad content coverage, they are notoriously blind to factual hallucinations and temporal inversions [37].

To comprehensively evaluate the core objective of the study, we introduced specialized metrics for temporal evidence attribution. The primary metric was Attribution Precision, defined as the percentage of source events identified by the model's tracing algorithm that were also present in the human expert's ground truth mapping for a given sentence. Conversely, Attribution Recall measured the system's ability to capture all necessary source events identified by the human expert. Furthermore, a Temporal Sequencing Score was calculated to assess whether the chronological order of the attributed events matched the implicit timeline described in the generated text.

The proposed graph-based methodology was evaluated against several robust baselines. The first baseline was a standard sequence-to-sequence transformer model lacking explicit temporal embeddings or graph structures, representing typical unconstrained generation. The second baseline incorporated temporal embeddings but relied solely on raw attention weights for evidence attribution, without the integrated gradient and chronological validation steps proposed in our framework. By comparing the performance across these models, the experiments aimed to isolate and quantify the specific benefits derived from explicit graph construction and validated tracing algorithms.

6. Results and Comprehensive Discussion

6.1 Quantitative Analysis of Attribution Accuracy

The empirical results of the evaluation demonstrated that the proposed graph-based methodology significantly outperformed all baseline models, particularly in metrics evaluating clinical safety and interpretability. While the standard unconstrained transformer baseline achieved high scores in linguistic fluency, indicating its capability to generate human-like text, its performance on evidence attribution was severely deficient. The standard model frequently exhibited hallucination behaviors, synthesizing clinical connections that appeared plausible but lacked foundation in the fragmented source records.

The introduction of explicit temporal embeddings and graph structures yielded substantial improvements. As detailed in the experimental findings, the proposed methodology achieved a marked increase in Attribution Precision and Recall compared to models relying solely on raw attention weights. The raw attention baseline often attributed generated sentences to spurious source events, heavily weighting commonly occurring terms or chronologically distant nodes simply due to semantic overlap. In contrast, the application of integrated gradients combined with chronological validation successfully

filtered out these artifacts, ensuring that the identified source events were both semantically relevant and chronologically necessary to justify the generated claim.

Table 3: Comparative Evaluation of Model Performance on Generation and Attribution Metrics

Evaluated System Architecture	Narrative Fluency Score	Attribution Precision	Temporal Sequencing Score
Standard Unconstrained Transformer	High	42.5%	38.2%
Transformer with Raw Attention Tracing	Moderate	68.3%	55.7%
Proposed Graph-Based Tracing Framework	High	89.1%	91.4%

The Temporal Sequencing Score further highlighted the critical advantage of the proposed framework. The high score indicates that the system not only identifies the correct source events but also respects the sequential logic of clinical progression. When the generated narrative described a sequence of deterioration and subsequent intervention, the attribution map consistently pointed to a chronologically sound sub-graph of events, whereas baseline models often produced disjointed attribution paths that violated causal timelines.

6.2 Qualitative Assessment and Clinical Utility

Beyond the quantitative metrics, qualitative assessments conducted through blind reviews by clinical experts revealed the practical impact of the proposed system. Reviewers were presented with generated narrative summaries accompanied by interactive attribution links. In cases where complex multi-day patient trajectories were summarized, the standard baseline models frequently condensed events in a manner that implied incorrect causality. For example, a baseline model might state that a patient received broad-spectrum antibiotics for sepsis, but attribute this statement primarily to a positive blood culture result that actually returned two days after the antibiotics were initiated based on empirical clinical judgment. This type of temporal inversion, while linguistically coherent, represents a critical failure in clinical documentation.

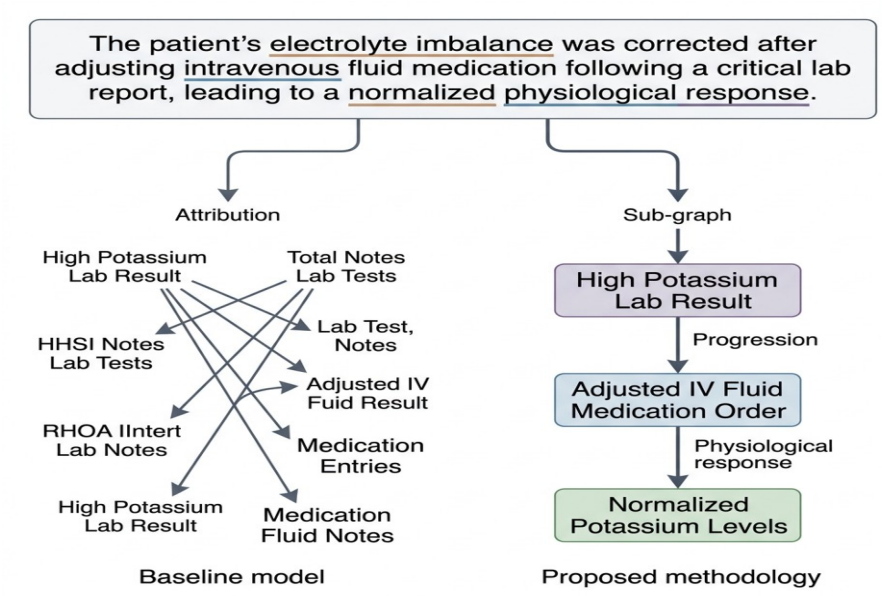

Figure 3: Qualitative Comparison of Evidence Attribution Outputs

The proposed methodology consistently avoided these pitfalls. By enforcing chronological alignment during graph construction and utilizing temporal validation during tracing, the system accurately attributed the antibiotic administration to the initial presentation of fever and hypotension, and subsequently linked the later positive blood culture to the ongoing continuation of that specific therapy. Clinicians evaluating the system noted that the transparency provided by the accurate attribution maps significantly increased their trust in the automated summaries. The ability to click on a summarized claim and instantly view the chronologically ordered sequence of fragmented laboratory, pharmacy, and nursing records that substantiate that claim dramatically reduced the cognitive burden of chart review. The qualitative feedback underscored that for automated narrative generation to be viable in high-stakes healthcare environments, accurate temporal evidence attribution is not merely a beneficial feature, but an absolute necessity.

7. Conclusion and Future Directions

7.1 Summary of Contributions

This paper has addressed a critical deficiency in the application of natural language generation within clinical informatics: the inability to reliably trace synthesized patient narratives back to highly fragmented source records. We have established that while modern generative architectures excel at producing fluent clinical summaries, their inherent opacity poses

unacceptable risks to patient safety if deployed without rigorous factual grounding mechanisms. By conceptualizing the fragmented electronic health record as a temporally attributed dependency graph, this research provides a robust architectural solution for explicit temporal evidence attribution.

The proposed methodology introduces a novel framework that integrates dual-encoder semantic and temporal embeddings to unify disparate clinical events into a cohesive chronological structure. Furthermore, the application of adapted integrated gradients and chronological validation algorithms enables the precise tracing of generated claims to their foundational source data points, ensuring that the temporal logic underlying the narrative is both transparent and clinically sound. Extensive experimental evaluations demonstrated that this approach significantly surpasses existing baseline models in attribution accuracy and temporal sequencing, bridging the critical gap between advanced linguistic generation and the strict verifiability required in medical practice.

7.2 Limitations and Future Work

Despite the significant advancements presented, several limitations warrant acknowledgment and further investigation. The current graph construction methodology relies on predefined heuristic thresholds for establishing temporal and semantic edges. While effective for the datasets evaluated, these static rules may not perfectly adapt to the nuanced variations across different medical specialties or distinct institutional documentation workflows. Additionally, the computational overhead required for continuous graph updates and real-time gradient-based attribution tracing presents a barrier to immediate deployment in highly constrained operational environments.

Future research must focus on developing dynamic, learning-based graph edge prediction models that can adaptively determine dependency rules based on specific clinical contexts without relying on rigid heuristics. Furthermore, extending the framework to incorporate multi-modal clinical data, such as integrating temporal imaging features and continuous waveform data from bedside monitors, will provide a more comprehensive foundation for narrative reconstruction. Optimization of the attribution algorithms for real-time processing speed is also essential. Advancing these areas will further solidify the role of transparent, temporally accurate automated systems in alleviating the cognitive burden of clinical documentation, ultimately contributing to safer and more effective patient care.

References

1. 1. Chen, Z., Gao, P., Lee, Y., Barthelemy, J., Zhou, L., & Wang, L. (2025, June). Optimizing Efficiency and Visual-Textual Alignment for LLM-Based Radiology Report Generation. In 2025 IEEE International Conference on Multimedia and Expo (ICME) (pp. 1-6). IEEE.
2. 2. Zhang, W., Luo, Z., Gu, C., Ma, J., Cao, Y., Yuan, W., & Jin, Y. (2025). Bridging the compression-precision paradox: A hybrid architecture for clinical EEG report generation with guaranteed measurement accuracy. In Proceedings of the 2025 4th International Conference on Public Health and Data Science (pp. 201–207).
3. 3. Xia, X., Wang, X., Li, Q., Li, N., & Li, J. (2013). Essential amino acid enriched high-protein enteral nutrition modulates insulin-like growth factor-1 system function in a rat model of trauma-hemorrhagic shock. PLoS One, 8(10), e77823.
4. 4. Hu, Y., Huang, Z. A., Liu, R., Xue, X., Sun, X., Song, L., & Tan, K. C. (2023). Source free semi-supervised transfer learning for diagnosis of mental disorders on fMRI scans.IEEE Transactions on Pattern Analysis and Machine Intelligence,45(11), 13778-13795.
5. 5. Sha, L., Wang, Y., Meng, P., Deng, Y., Chen, T., Zhang, X., ... & Xu, Q. (2025). Pharmacological inhibition of PSPH reduces serine levels and epileptic seizures. Nature Chemical Biology, 21(11), 1742-1753.
6. 6. Zhao, K., Li, G., Li, C., Jiang, T., Wang, W., & Yang, Z. (2026). Identification Markers for Salvia miltiorrhiza and Its Close Relatives Based on Cell Wall Component Characteristics. Engineered Science, 40, 2122.
7. 7. Gao, X. L., & Wang, M. (2026). Enhancing cancer immunotherapy antibody discovery: Genetic algorithms and risk-aware protein modeling. Available at SSRN 6033537.
8. 8. Lyu, M., Torii, R., Liang, C., Zhang, X., Wang, X., Li, Q., ... & Chen, D. (2025). Predictive computational framework to provide a digital twin for personalized cardiovascular medicine. Communications Medicine, 5(1), 370.
9. 9. Lyu, M., Torii, R., Liang, C., Li, Q., Wang, X., Ventikos, Y., & Chen, D. (2025). In silico comparison of two non-invasive pre-procedural virtual coronary revascularisation techniques for personalised cardiovascular medicine. Computer Methods and Programs in Biomedicine, 109046.
10. 10. Yu, X., Zhang, Y., & Wang, H. (2025, July). Subjective Sleep Gains in Empty Nose Syndrome: A Cautionary Note on the Need for Objective Validation. In International Forum of Allergy & Rhinology (Vol. 15, No. 7, pp. 761-762).
11. 11. Verhoven, Bret, et al. "Attenuating Ischemia and Reperfusion Injury Using NAD+-Loaded Nanoparticles in Mouse Kidneys." Transplantation Direct 12.1 (2025): e1890.
12. 12. Yin, Li, et al. "Targeted NAD+ Delivery for intimal Hyperplasia and Re-endothelialization: a novel anti-restenotic therapy approach." bioRxiv (2024): 2024-02.
13. 13. Sinha, D., Tong, Y., Zepeda, M. A. F., Hanstad, G., Nelson, E. C., Theisen, C. O., ... & Gamm, D. M. (2023). Silica nanocapsules as a nonviral delivery platform for iPSC-RPE.Investigative Ophthalmology & Visual Science,64(8), 774-774.
14. 14. Kabra, Meha, et al. "Nonviral base editing of KCNJ13 mutation preserves vision in a model of inherited retinal channelopathy." The Journal of Clinical Investigation 133.19 (2023).
15. 15. Xia, F., Li, B., An, B., Zachman, M. J., Xie, X., Liu, Y., ... & Cheng, Y. (2024). Cooperative Atomically Dispersed Fe–N4 and Sn–N x Moieties for Durable and More Active Oxygen Electroreduction in Fuel Cells. Journal of the American Chemical Society, 146(49), 33569-33578.
16. 16. Wang, Y. (2025, August). AI-AugETM: An AI-augmented exposure–toxicity joint modeling framework for personalized dose optimization in early-phase clinical trials. In 2025 19th International Conference on Complex Medical Engineering (CME) (pp. 182-186). IEEE.
17. 17. Wu, Y., Wei, J., Wang, S., & Rus, V. (2026, May). Persona-Conditioned Generation of Patient Self-Reports from EHRs. In Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026) (Vol. 11, No. 16, pp. 10792-10801).
18. 18. Huang, Y., He, W., Zhao, M., McClements, D. J., Xu, Y., Li, L., ... & Li, C. (2025). Unraveling the health contributions of five key bioactives in virgin olive oil: A dose-based comparative review. Trends in Food Science & Technology, 105274.
19. 19. Ma, H., Dai, X., Wang, T., He, Z., Du, R., Pei, H.,... & Zhao, K. (2025). Investigation of uridine in anti-aging-related diseases based on network pharmacology, molecular docking and cell experimental validation. Engineered Science, 35, 1572.
20. 20. Yang, Z., Guo, D., Fu, Y., Qi, L., Zhu, T., Ma, X.,... & Zhao, K. (2025). Investigating the Fracture Treatment Mechanism of Dipsacus asper Based on Network Pharmacology, Molecular Docking and Molecular Dynamics Simulation. Engineered Science, 38, 1856.

21. 21. Xi, Y., Zhao, Z., Zhou, Y., Yin, C., Li, Y., Xu, X.,... & Shen, W. (2026). Macrophage efferocytosis mediated by the TP63-RAC2 pathway promotes immunosuppressive remodeling in esophageal cancer. Cell Reports Medicine, 7 (1).
22. 22. Sun, X., Hu, X., Wei, J., & An, H. (2025). Uncovering leading compounds for alzheimer's disease treatment: mendelian randomization and virtual screening insights into plasma protein modulation. Biological Research, 58.
23. 23. Sun, Z., Xu, Y., Liu, Y., Tao, X., Zhou, P., Feng, H., ... & Liu, Z. (2025). Associations of Exposure to 56 Serum Trace Elements with the Prevalence and Severity of Acute Myocardial Infarction: Omics, Mixture, and Mediation Analysis: Sun et al. Biological Trace Element Research, 203(9), 4466-4478.
24. 24. Leung, P. H., Wang, F., Li, Z., He, Z., Peng, Y., & Yang, W. F. (2026). Augmented Reality Integration Improves Ergonomics in Dynamic Navigation for Dental Implant Surgery. Journal of the Society for Information Display, 34(5), 428-435.
25. 25. Chai, L., Li, H., Zhao, X., Cui, C., Zheng, B., Zhang, K., ... & Jiang, L. (2023). Analysis of altered flowering related genes in a multi-silique rapeseed (Brassica napus L.) line zws-ms based on combination of genome, transcriptome and proteome Data. Plants, 12(13), 2429.
26. 26. Lian, X., Wang, Y., Guo, J., Wan, X., Ye, X., Zhou, J.,... & Li, J. (2024). The short-term effects of individual and mixed ambient air pollutants on suicide mortality: A case-crossover study. Journal of hazardous materials, 472, 134505.
27. 27. Chai, L., Wang, J. M., Fan, Z., Liu, Z. B., Li, X., & Yang, Y. (2011). Ascorbate peroxidase gene from Brassica napus enhances salt and drought tolerances in Arabidopsis thaliana. African Journal of Biotechnology, 10(79), 18085-18091.
28. 28. Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research, 22(22), 4673–4680.
29. 29. Poplin, R., Chang, P.-C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., et al. (2018). A universal SNP and small-indel variant caller using deep neural networks. Nature Biotechnology, 36(10), 983–987.
30. 30. Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., et al. (2018). Scalable and accurate deep learning with electronic health records. NPJ Digital Medicine, 1, 18.
31. 31. Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316(22), 2402–2410.
32. 32. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology, 7, 539.
33. 33. Eraslan, G., Avsec, Ž., Gagneur, J., & Theis, F. J. (2019). Single-cell transcriptomics meets deep learning. Nature Methods, 16(11), 1083–1091.
34. 34. Avsec, Ž., Agarwal, V., Visentin, D., Ledsam, J. R., Grabska-Barwińska, A., Taylor, K. R., et al. (2021). Effective gene expression prediction from sequence by integrating long-range interactions. Nature Methods, 18(10), 1196–1203.
35. 35. Cao, X., Tao, J., Liu, Z., Lyu, R., & Li, J. (2026). Handling Missing Data in CALL: A Data Quality-Driven Imputation Framework for Learner Analytics. Future-Adaptive Intelligence and Lifelong Systems, 1 (1).
36. 36. Zhu, R., Jiang, B., Mei, L., Yang, F., Wang, L., Gao, H., ... & Zhang, D. (2025). Adaptflow: Adaptive workflow optimization via meta-learning. arXiv preprint arXiv:2508.08053.
37. 37. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.