IFERP Explore Your Research Journey...

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 05, May 2026

An Agentic Hybrid Architecture for Grounded Longitudinal Clinical Reasoning and Explainable EHR Summarization

[1] M. Eliazer, [2] Asmita Mandal, [3] Sneha Puri

[1] [2] [3] Department of Computer Science and Engineering, School of Computing, SRM Institute of Science and Technology, Kattankulathur, India

Corresponding Author Email: [1] eliazerm@srmist.edu.in, [2] am4090@srmist.edu.in, [3] sp8526@srmist.edu.in

***Abstract*— Longitudinal Electronic Health Records (EHRs) have clinical signals that are spread out and required to be thought about in several steps for correct understanding. Traditional systems either depend on separate rule-based alerts or directly create narrative summaries, which lack a structured time frame. This paper talks about an agentic hybrid architecture that combines deterministic longitudinal analytics with controlled large language model (LLM) generation. Before making a structured insight representation, the agent first co- ordinates normalization, trend extraction, recurrence modeling, and medication transition analysis, and then this representation is turned into summaries for clinicians using limited prompts. The framework helps to make things more clear, modular, and easy to understand in a clinical setting by combining autonomous analytical orchestration with grounded language synthesis.**

***Index Terms*— Electronic Health Records, Agentic Systems, Hybrid Intelligence, Longitudinal Reasoning, Explainable Clinical AI.**

I. INTRODUCTION

Electronic Health Records have become complete stores of patient information that is spread out over time. This structured storage makes data more accessible, but clinical interpretation still needs to combine data from multiple visits to find trends in progression, persistent deviations, and changes in treatment for accurate results.

Most decision-support tools work in one of two ways. Rule- based systems send out threshold alerts, but they don't group information by context. End-to-end generative methods help to make summaries that flow well, but they don't always have clear analytical reasons for doing so. The lack of structured temporal reasoning in summarization pipelines makes it hard to rely on and audit them.

To solve this gap, we introduce an agent-driven hybrid framework that separates deterministic computation from se- mantic generation while also maintaining coordinated reasoning flow. The agent also acts as a conductor; it calls analytical modules in order, combines structured findings, and makes sure that grounded summarization gives accurate results.

In high-volume clinical environments, the difficulty not only lies in detecting abnormal values but also in integrating temporally dispersed signals into cohesive clinical narratives. For example, a gradual rise in biomarkers over multiple visits

may be a sign that the disease is getting worse, even if each reading is only slightly above the threshold. Incremental medication adjustments can also show patterns of therapeutic escalation that are hard to see when reports are looked at on their own. These long-term dependencies need structured reasoning systems that can combine evidence from multiple visits instead of looking at each record on its own.

II. LITERATURE SURVEY

Clinical intelligence systems have evolved from rule-based anomaly detection to more advanced data-driven and language-enhanced frameworks.

A. Automated Clinical Summarization

Earlier clinical summarization methods focused on ex tracting key information from medical records or converting structured data into brief text summaries. With the rise of transformer-based models, the quality improved in terms of fluency and contextual understanding. However, applying these generative models directly to raw EHR data still presents challenges, particularly in accurately representing temporal patterns and validating numerical trends.

B. Temporal Modeling in Healthcare

Longitudinal analysis methodologies have utilized regression models, recurrence statistics, and persistence indicators to assess disease progression. These methods focus on changes in direction and how often something goes wrong, not just when a threshold is crossed. Most temporal models are good at analyzing data, but they are made to predict risk rather than summarize it in a way that makes sense.

C. Hybrid Analytical–Generative Systems

Recent research has investigated the integration of structured computation with neural language models to improve interpretability. Hybrid pipelines frequently retain intermediate representations before text generation, facilitating partial traceability. However, there has been little research specifically on agent-driven orchestration of deterministic longitudinal reasoning followed by constrained summarization in EHR contexts. The suggested framework

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 05, May 2026

fills this gap.

III. AGENTIC HYBRID ARCHITECTURE

The suggested system uses a coordinated hybrid pipeline, with an agent in charge of both deterministic analytical components and semantic generation modules.

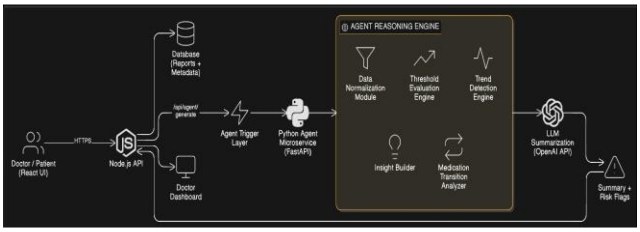

Figure 1. Agent-coordinated hybrid pipeline for longitudinal clinical reasoning and summarization.

A. Normalization and Temporal Structuring

Chronological alignment, unit harmonization, and clear handling of incomplete entries help standardize incoming reports. This makes sure that the analysis is consistent before the agent is called.

B. Trend Extraction Module

Regression analysis measures how each parameter changes over time:

$$m = \frac{\sum (t_i - \bar{t})(x_i - \bar{x})}{\sum (t_i - \bar{t})^2}$$

The slope coefficient measures how things change over time during visits.

C. Recurrence Modeling

Persistent abnormality is computed as:

$$F = \frac{N_{abnormal}}{N_{total}}$$

Parameters that go above the persistence threshold θ are marked as chronic deviations.

D. Medication Transition Analysis

Symmetric set operations are used to find therapeutic changes:

$$\Delta M = M_{current} \triangle M_{previous}$$

This keeps track of changes, additions, and replacements between visits.

E. Insight Object Construction

The agent puts together computed metrics into a structured Insight Object that has:

- Directional trend indicators
- Abnormal recurrence ratios
- Therapy evolution logs

This intermediate representation preserves analytical trace- ability.

F. Controlled LLM Synthesis

The Insight Object is given to a limited prompt template. The language model creates summaries for clinicians that are based only on structured inputs, so it can't make any guesses that aren't based on calculated metrics.

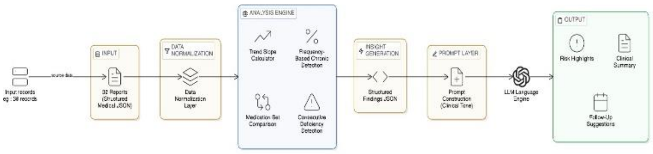

Figure 2. Agent-coordinated hybrid pipeline for longitudinal clinical reasoning and summarization.

IV. ANALYTICAL EVALUATION

A. Validation of Temporal Computation

Controlled synthetic longitudinal sequences were created to test slope estimation, recurrence modeling, and medication transition detection under specific conditions in order to check for analytical correctness.

We made sequences that steadily went up, down, or stayed the same to check the validity of the directional trend. When the regression formula was used in (1), we got positive slope coefficients for progressive patterns and negative slope coefficients for regressive patterns. Stable sequences produced slopes that are almost close to zero, which shows that they were numerically stable and sensitive to changes in direction.

B. Recurrence Classification Consistency

We tested the recurrence formulation in (2) using different types of abnormal frequency distributions. The persistence threshold θ was exceeded in situations that had significant abnormal occurrence ratios, which is the reason that they were called sustained deviations. On the other hand, outlier values that were not associated with anything else did not have a big effect on the recurrence ratio, which kept false chronic classification from occurring.

C. Finding Therapeutic Transitions

We used structured drug lists from multiple reports to conduct medication evolution analysis. The symmetric set formulation in (3) consistently detected additions, deletions, and substitutions of therapy.

D. Grounded Summary Generation

After deterministic computation, structured Insight Objects were sent to the controlled LLM module to make summaries for clinicians. Figure 3 shows how structured analytical metrics change into grounded narrative output.

The generated summaries combined progression indicators, recurrence statistics, and therapy transitions into one clear interpretation, which is better than looking at multiple reports on their own. Importantly, narrative outputs

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13, Issue 05, May 2026

stayed in line with the computed metrics, which made sure that there was a clear link between analytical reasoning and semantic expression.

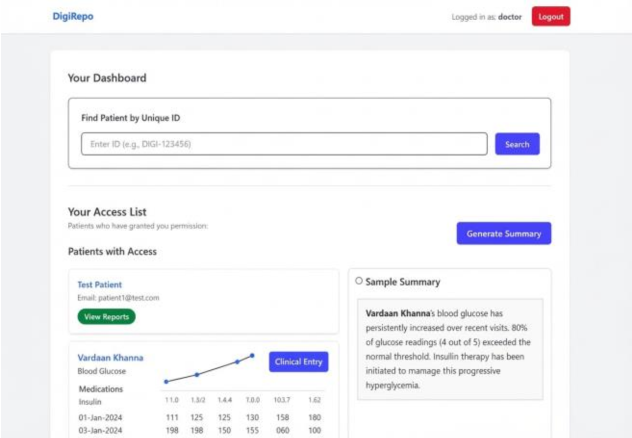

Figure 3. Agent-generated clinical summary derived from structured longitudinal analytical findings.

V. WORKFLOW INTEGRATION

The recommended agentic framework is designed for integration as a modular microservice incorporated into existing Electronic Health Record systems. The architecture doesn't replace essential clinical systems. Instead, it acts as an analytical layer which is used when analyzing patient records on a regular basis.

A. Operational Trigger Mechanism

When a clinician accesses a patient profile, the system gets the most recent longitudinal reports and operates on the deterministic analytical pipeline. The agent calls the normalization, temporal computation, recurrence modeling, and medication comparison modules one after the other before making the Insight Object. Once structured reasoning is done, the supervised LLM only makes a brief summary of the context.

This staged execution assures that narrative outputs remain consistently grounded in accurate calculated metrics.

B. Audit and Transparency Layer

The Insight Object keeps all of the intermediate analytical values, like slope coefficients, recurrence ratios, and detected therapy transitions. Before accepting narrative interpretations, clinicians may look at these numbers. This ability to audit reinforces accountability and prevents individuals from relying too much on generative outputs.

Table I: Workflow Integration Components

Component	**Function**
EHR Database	Stores structured clinical reports
Agent Controller	Orchestrates analytical modules
Temporal Engine	Computes trends and recurrence
Insight Object	Stores intermediate metrics
LLM Module	Generates grounded summary

C. Execution Flow Characteristics

The overall workflow keeps the order of execution and the speed of computation the same. Since analytical modules work in linear time based on the number of reports, it's still feasible to call them in real time when evaluating a patient for standard longitudinal histories.

Table II: Analytical Module Validation Accuracy

Module	**Test Scenario**	**Accuracy (%)**
Trend Detection	Increasing/Decreasing Patterns	80
Recurrence Modeling	Persistent vs Transient Cases	78
Medication Transition	Add/Remove/Substitute	100
Insight Construction	Structured Output Integrity	100

The framework improves interpretive efficiency by incorporating structured longitudinal reasoning directly into standard clinical review processes while maintaining predefined decision pathways. Since computation and summarization are separate, automation helps doctors instead of replacing them.

VI. FUTURE SCOPE

While the hybrid agentic framework described here provides a baseline for longitudinal clinical summarization, there remain ways to make advancements.

First, adaptive thresholding methods could be added to make abnormality detection more personal. By establishing individualized baselines derived from a patient's own historical data, the system could identify minute shifts in disease trajectory that static benchmarks might not capture.

Second, integrating predictive risk-forecasting modules with the deterministic pipeline used currently holds significant potential. The current focus remains on rule-based interpretation, however, adding a probabilistic layer for early warning could improve proactive care after rigorous clinical validation.

Third, integration of multiple sources of data including unstructured physician notes, imaging metadata, wearable sensor data, or lifestyle indicators is a chance for growth. Integrating structured analytics with contextual clinical descriptions may better the depth of interpretation.

Fourth, a large-scale clinical evaluation is needed to measure usability and impact in the real-world. User studies could estimate how much time clinicians save in reviewing cases, how much less time is spent on paperwork, and how much lower mental effort is required to use.

Lastly, integrating federated learning systems could allow analytical specifications to be aligned across institutions with- out bringing sensitive patient data into a single location. This kind of privacy-preserving approach would support the responsible growth of healthcare networks.

VII. DISCUSSION

A. Architectural Significance

The proposed hybrid agent-based framework demonstrates how structured, step-by-step longitudinal reasoning can be combined with carefully managed language generation. By ensuring that analytical computations are completed before any narrative is produced, the design prevents semantic interpretation from occurring without a solid numerical basis. Keeping reasoning separate from expression is especially important in clinical environments where traceability and auditability are essential.

Unlike purely generative pipelines that operate directly on raw medical data, this framework introduces an intermediate Insight Object. This structured layer ensures that every generated statement can be traced back to clearly defined calculations, such as slope values or abnormality frequency ratios.

B. Comparison with Existing Paradigms

Traditional rule-based systems can clearly flag abnormalities but often fail to present how conditions evolve over time alongside treatment in a unified way. In contrast, LLM-based summarization produces smooth, readable narratives, yet these may lack solid numerical backing.

The proposed architecture bridges this gap by combining structured computation with controlled generative synthesis. Rather than replacing analytical reasoning with probabilistic predictions, it focuses on improving interpretability while preserving modular transparency. Compared to unrestricted language-only approaches, this hybrid design reduces the risk of hallucinations and supports more reproducible outcomes.

C. Interpretability and Trust

For AI systems to be applied in real-world settings, strong performance is not the only criterion; they also need to be trusted. By retaining intermediate numerical outputs, this framework lets doctors review the reasoning process before accepting the final narrative summaries. Being able to examine details like slope values, recurrence thresholds, and medication changes encourages more responsible use and reduces dependence on fully automated results.

D. Operational Considerations

From a systems perspective, the deterministic components operate in linear time relative to the number of reports, making them fitting for routine clinical use. Since the design functions as a modular overlay rather than a standalone predictive engine, it can be integrated into existing EHR systems without requiring extensive retraining.

This modular separation also lets adjustments be done to regression settings, recurrence thresholds, or prompt templates independently, without reworking the entire system. Such flexibility is valuable in healthcare environments where strict regulations and continuous validation are required.

VIII. LIMITATIONS

While the proposed framework is structured and designed with emphasis on ease of understanding, it still has a few limitations that should be acknowledged.

First, the system assumes that the EHR data is well-structured and consistently formatted. In real clinical environments, lab reports can vary in units, may have missing entries or timestamps, and often differ in format across different institutions. Although a normalization layer is included, weighty pre-processing may still be required to ensure reliable analysis at larger scale.

Second, the framework focuses on deterministic longitudinal aggregation rather than predicting inference. It does not generate probabilistic risk scores, perform diagnoses, or forecast progression of a disease. While this makes the system easier to verify, it limits its usefulness in scenarios where predictive insights could add value.

Third, finding abnormalities relies on threshold values that are set in advance. Unvarying thresholds may not capture personalized physiological baselines or individual patient variability. For example, small but consistent straying from a patient's normal range may be clinically important, even if they fall within the general population thresholds. There are no adaptive calibration mechanisms at present.

Fourth, the evaluation in this study is based on controlled synthetic longitudinal scenarios intended to validate analytical correctness. While this supports deterministic reliability, broader clinical validation across diverse populations of patients is needed to find usability, impact on review time, and overall effectiveness in practice

Fifth, the LLM-based summarization unit depends on how the prompt is written and how external APIs work, even though it operates on structured inputs. Updates to language models or changes in deployment settings could alter how outputs are generated. Although grounding reduces hallucination risk, it cannot eliminate it entirely.

Finally, the current implementation is limited to structured data from medications and labs and it does not cover un-structured medical notes, imaging-derived information, data collected from wearables, or other types of multi-modal health data. Integrating these sources would call for added analytical modules and validation processes.

These limitations indicate the need for further refinement and comprehensive clinical evaluation before the framework can be widely adopted.

IX. CONCLUSION

This paper presents a hybrid agent-based framework de-signed for longitudinal analysis of clinical reports and the generation of explainable summaries within Electronic Health Record systems. It addresses a key challenge in modern healthcare workflows: bringing together medical data that is spread across time, in order to identify progression patterns, persistent abnormalities, and treatment changes.

Unlike conventional threshold-driven alert systems or purely generative summarization approaches, the framework enforces deterministic longitudinal reasoning before producing any narrative output. By explicitly calculating regression-based trends, abnormal recurrence measures, and medication transition events, the generated summaries remain grounded in structured analytical evidence. The inclusion of an intermediate Insight Object further supports traceability, allowing clinicians to review underlying metrics before accepting the final interpretation.

Validation results demonstrate stable performance in controlled longitudinal scenarios along with computational efficiency. By separating structured reasoning from language generation, the system achieves a practical balance between interpretability, modularity, and usability, all of which are essential for responsible AI use in clinical contexts.

Rather than replacing clinical judgment, the framework is intended to support it by reducing cognitive load and transforming scattered data into clear, interpretable summaries. Overall, the study highlights the promise of agent-driven hybrid intelligence systems that combine deterministic computation with constrained language models to provide transparent and clinically relevant decision support.

REFERENCES

[1] X. Yang *et al.*, "A large language model for electronic health records," *npj Digital Medicine*, vol. 5, no. 1, pp. 1–9, 2022.

[2] Y. Zhang, M. Chen, and Q. Li, "Clinical text summarization using transformer-based architectures," *Journal of Biomedical Informatics*, vol. 120, p. 103859, 2021.

[3] S. Tjoa and C. Guan, "A survey on explainable artificial intelligence (XAI): Toward medical XAI," *IEEE Access*, vol. 8, pp. 128620–128651, 2020.

[4] R. Sutton *et al.*, "An overview of clinical decision support systems: benefits, risks, and strategies for success," *npj Digital Medicine*, vol. 3, no. 1, pp. 1–10, 2020.

[5] J. Futoma, J. Morris, and J. Lucas, "A comparison of models for pre- dicting early hospital readmissions," *Journal of Biomedical Informatics*, vol. 56, pp. 229–238, 2015.

[6] H. Nori *et al.*, "Capabilities of GPT-4 on medical challenge problems," arXiv:2303.13375, 2023.

[7] T. B. Brown *et al.*, "Language models are few-shot learners," in *Proc. NeurIPS*, 2020, pp. 1877–1901.

[8] A. Holzinger, "Explainable AI and multi-modal causability in medicine," *IEEE Intelligent Systems*, vol. 36, no. 2, pp. 67–77, 2021.