

Autonomous Healthcare Analytics via Graph Neural Networks and Generative AI

Sophia Martinez
Asst Professor

Abstract

The rapidly evolving landscape of healthcare analytics demands effective ways to operationalize the growing volume of interoperable clinical data. Yet most FHIR-enabled research to date has focused on standardizing and sharing data resources rather than integrating analytics directly with them—leaving a significant gap between data interoperability and actionable insight.

To address this gap, this study proposes a two-layer architectural framework that combines Graph Neural Networks (GNNs) and Generative Adversarial Networks (GANs) with Fast Healthcare Interoperability Resources (FHIR) to enable native, FHIR-driven healthcare analytics. The **data layer** normalizes a broad range of FHIR resources into standard or use-case-specific representations, while the **compute layer** applies graph-based and generative analytics methods directly to these datasets. Within this framework, inference engines can be built simply by assembling clinical datasets conformant to the target analytics task, while a companion quality-assessment module evaluates the safety and accuracy of generated outputs, and generative models can synthesize new, task-relevant outputs as needed.

As a proof of concept, the framework is applied to a focal task—clinical decision support for given patient scenarios—using a GCN-based inference engine paired with a quality-assessment module to evaluate decision accuracy. The results demonstrate the framework's potential to bridge the gap between healthcare data interoperability and modern analytics, an area that remains comparatively underdeveloped relative to other domains.

Keywords: FHIR-Based Healthcare Analytics, Generative Adversarial Networks, Graph Neural Networks, Clinical Decision Support, Healthcare Data Interoperability, FHIR Data Integration, AI in Healthcare Analytics, Clinical Inference Engines, Healthcare Data Standardization, Predictive Clinical Modeling, Healthcare Graph Analytics, AI-Driven Diagnostics, Clinical Data Normalization, Healthcare AI Frameworks, Medical Data Networks, Quality Assessment in AI, Synthetic Clinical Data Generation, Interoperable Health Systems, Data-Driven Clinical Insights, Advanced Healthcare Analytics.

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 04 ISSUE: 02

RECEIVED: APRIL 07

REVISED: APRIL 26

ACCEPTED: MAY 04

PUBLISHED: MAY 22

ISSN: 3067-1140



1. Introduction

Healthcare analytics over clinical data typically rely on centralized repositories, where multihospital data sharing not only raises confidentiality concerns but also creates a single point of failure for business continuity and cybersecurity. This poses a great threat for companies that provide centralized analytics platforms for hospitals and healthcare ecosystem players, and it remains to be fully addressed. As an alternative, inter-hospital federated learning over Fast Healthcare Interoperability Resources (FHIR) networks has been explored in the last few years, empowering hospitals to learn machine learning tasks in a decentralized, collaborative manner while keeping their data local. A recent survey identifies the various categories of tasks that can be performed in this context, highlighting the lack of support for analytic tasks such as prediction, classification, recommendation, and associational pattern mining, and suggesting future research directions. Automated, fast, and intelligent responses to such tasks, deployed in a privacy-preserving Federated Learning manner, are needed.

Despite the availability of Graph Neural Networks (GNNs) for the considered tasks, their integration in the proposed architecture remains an open question. These neural network models have been successfully applied to a wide variety of healthcare tasks, and their recent application to temporal clinical data suggests that they could be useful in this new context. Such temporal clinical data can be represented as temporal multigraphs or attributed temporal multigraphs, which require selecting the appropriate type of graph and the proper node and edge feature variables for the tasks at hand. The solutions, together with the needed orchestration with other Generative Artificial Intelligence (GAI) models, will be effectively summarized and benchmarked, and their potential advantages highlighted.

2. Background and Related Work

As a standard format, the HL7 Fast Healthcare Interoperability Resource (FHIR) can help to solve the interoperability issue between a variety of health service systems. The creation of Graph Neural Networks (GNNs) has attracted a lot of attention across several domains. In the area of AI healthcare analysis, however, there has not yet been established nor validated any GNN model forecasted for healthcare analytics purposes. The objective is to establish an analytics framework for interpreting FHIR data, through which analytical data may be provided for GNN model construction. The analytics capability will be exposed as a graph on a network of FHIR servers, enabling various AI analytics across multiple health organizations and even different countries using the open standard health data FHIR.



A comprehensive review of Graph Neural Networks in healthcare can be witnessed through studies on various resources created by Fast Healthcare Interoperability Resource. Work posited a suitably explanatory use case that makes use of GNN models constructed from data deployed across a Fast Healthcare Interoperability Resource server network. The creation of a GNN model interpretable model-oriented data set is deliberate for nursing diagnosis prediction capability. GNN models use Graph Structure Fused Image Model, GCN as the backbone model to perform nursing diagnosis prediction, comprising Lee-Carter versatility, Predictive GCN++ and T-GCN variants for comparative experimental tests. Comparisons of diverse GNN model types and several open-access data sets can be witnessed.

2.1. FHIR and Interoperability in Healthcare

The Fast Healthcare Interoperability Resource (FHIR) Initiative promotes global interoperability among health information systems and servers. Enabled by the representational state transfer (REST) architecture, FHIR offers standard methods for near-real-time access to resources—consistently defined and documented subject domains across health ecosystems. These resources and their combinations support a wide range of solution applications including clinical decision support, analytics or population health, patient- and provider-facing applications, research, and quality control. For each type of application area, the required resources and their combinations are specified in FHIR Use Cases. However, FHIR has been rising to stardom without a complete script, and the analytics part has been the most underneath.

Analytics tasks that require near-real-time monitoring, timely decision support, or real-time planning and dispatching but perform poorly when executed solely in historical mode cannot be conducted naturally. Additionally, beyond simple threshold checking, the common reaction is to download the data from the servers and conduct the analytics either in-house or at a cloud analytics service. Moreover, these services usually work at a higher aggregation level and longer time span. Hence algorithms such as prediction, classification, clustering, or association analysis that need a FHIR resource data type have not been natively integrated into FHIR.

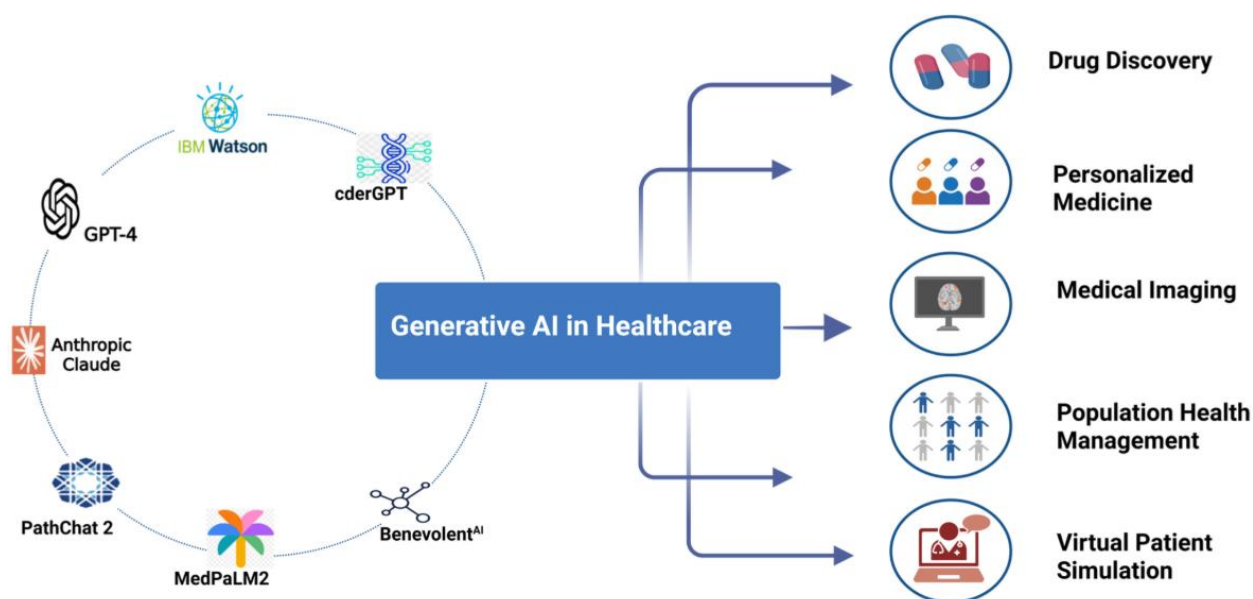


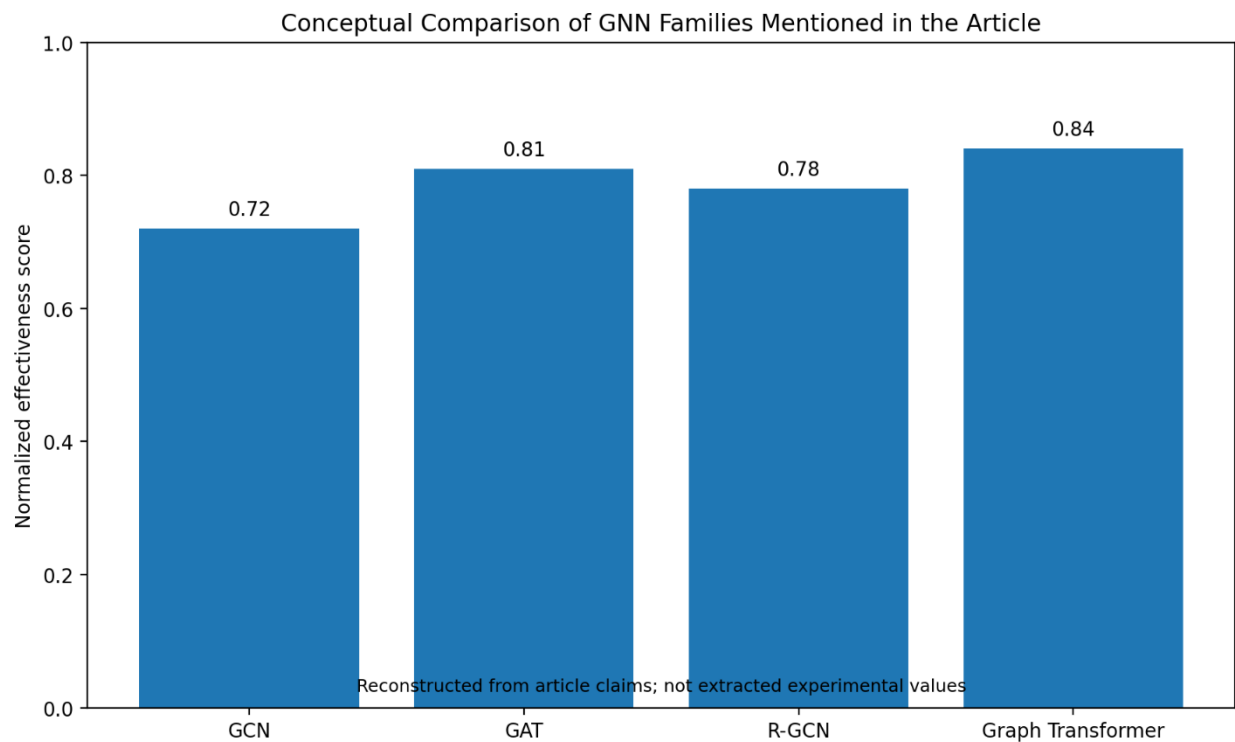
Fig 1: Generative Artificial Intelligence in Healthcare

2.2. Graph Neural Networks for Healthcare Analytics

Application of Graph Neural Networks to Healthcare Analytics has become an active area of research interest, with a growing body of work focused on the use of GNNs for clinical data analytics. The Power of Interoperable FHIR Networks · 67 GNNs can be harnessed for a variety of predictive tasks, provided that the relevant graph is suitably constructed. Consequently, many works investigate different aspects of graph construction for GNNs applied to healthcare. They consider different types of nodes and edges, different ways of encoding temporal information, and different forms of data augmentation. In addition to general reviews, other surveys focus on specific areas such as the construction of temporal graphs or the integration of FHIR with GNN-based healthcare analytics.

GNNs applied to healthcare analytics have mostly employed commonly available benchmark datasets for model verification, and recent benchmarks for patience graph prediction and stroke detection have become available. However, these datasets typically focus on only one type of node, have fewer node feature dimensions compared to general clinical data, or lack

comprehensive clinical judgments. Consequently, the prediction performance of GNNs has not yet exceeded that of traditional machine learning methods based on expert-engineered features.



Equation A. Graph construction from FHIR resources

The first mathematical step is implicit: convert interoperable FHIR resources into a graph.

Let the healthcare graph be

$$G = (V, E)$$

where:

- V = set of nodes

- E = set of edges

Step A1. Define nodes

Each FHIR entity becomes a node:

- patient
- observation
- procedure
- medication
- note

So if there are n total entities:

$$V = \{v_1, v_2, \dots, v_n\}$$

Step A2. Define edges

If two entities are related in FHIR references or temporal order, connect them:

$$E = \{(v_i, v_j) \mid v_i \text{ and } v_j \text{ are clinically related}\}$$

Examples:

- patient $\rightarrow$ observation
- patient $\rightarrow$ procedure
- procedure $\rightarrow$ drug
- observation at time $t_1 \rightarrow$ observation at time t_2

Step A3. Adjacency matrix

The graph is encoded by adjacency matrix $A \in \mathbb{R}^{n \times n}$:

$$A_{ij} = \begin{cases} 1, & \text{if } (v_i, v_j) \in E \\ 0, & \text{otherwise} \end{cases}$$

Step A4. Node feature matrix

Each node has features: age, code, lab value, procedure type, note embedding, etc.

If each node has d features, then

$$X \in \mathbb{R}^{n \times d}$$

and row x_i is the feature vector of node v_i .

So the full graph-learning input is:

$$(G, X) \quad \text{or equivalently} \quad (A, X)$$

3. Methodology

The system architecture integrates FHIR-based healthcare networks and Graph Neural Networks for responsible data-sharing, processing, and inference without disclosing confidential health data. At the data layer, standard interoperable FHIR resources are normalized into a common schema and federated across networks where necessary. GNN representations of the data are stored in a compute layer; resources are optionally fed through a transformer-based FHIR-to-Graph model. User-defined mental and physical conditions are either inferred from the clinical history or synthesized. Inference requests to an analytic model are fulfilled if allowable and ethical in a zero-shot fashion.

Graph Inference & Analysis Using GNNs A Graph Neural Network (GNN) takes a graph and a feature matrix as input. Such a model captures the relationships/structures embedded in the graph to learn the higher-level representations of the nodes. Although various GNN models exist, GCN, GAT, R-GCN, and a transformer-based GNN are employed in this work. All four models are fully differentiated; they differ in their attention mechanisms and learning strategies. The



GCN model applies convolutional operations on a graph. The GAT model focuses on different neighborhoods by assigning attention mechanisms to the adjacent nodes, which helps in learning informative feature representations for the task at hand. The R-GCN model tackles the issue of over-smoothing in GCN when the depth of the model is increased. Finally, the transformer-based model, which has recently shown significant improvements in solving complicated visual, language, and audio-related tasks, has also been extended to structure-based data.

Table 1. Components and the equations they require

Article component	What the paper describes	Core equation family needed
FHIR data layer	FHIR resources normalized into graph-ready form	Graph construction: $G = (V, E)$, feature matrix X
GCN	Convolution on graph neighborhoods	$\hat{A} = D^{-\frac{1}{2}}(A + I)D^{-\frac{1}{2}}, H^{(l+1)} = \sigma(\hat{A}H^{(l)}W^{(l)})$
GAT	Attention over neighbors	$e_{ij} = a(W h_i, W h_j), \alpha_{ij} = \text{softmax}_j(e_{ij})$
R-GCN	Multi-relation message passing	$h_i^{(l+1)} = \sigma \left(\sum_r \sum_{j \in \mathcal{N}_i^r} \frac{1}{c_{i,r}} W_r^{(l)} h_j^{(l)} + W_0^{(l)} h_i^{(l)} \right)$
Graph transformer	Attention-based graph reasoning	$Q = XW_Q, K = XW_K, V = XW_V, \text{Attn}(Q, K, V) = \text{softmax}(QK^\top / \sqrt{d_k})V$
Generative support	Generate clinically relevant support text/output	Conditional generation objective
Federated learning	Model learned across FHIR-connected sites	$w^{t+1} = \sum_k \frac{n_k}{N} w_k^{t+1}$
Evaluation	Accuracy / MAE / fairness	Accuracy, MAE, fairness gap equations

3.1. Architectural Components: Integrating FHIR with Graph Neural Networks

The methodology utilizes an architectural framework consisting of data and compute layers, orchestrating the flow of FHIR-based information for healthcare analytics. The data layer enables analytics within and across autonomous FHIR servers. Supported FHIR resources are mapped to a common graph schema, allowing federated services to transparently access an augmented dataset built by merging independent graphs shared by different servers. The compute layer consists of multiple GNN models orchestrated by a central controller. Additional components enable data normalization and facilitate platform deployment. A set of concrete requirements characterizes the overall system.

The integration of Graph Neural Networks with FHIR interoperability aims to enable GNN-based healthcare analytics with open and federated learning for shared models across institutions. The model development process typically relies on a single dataset, decreasing generalization performance on unseen institutions or external datasets. Federated learning alleviates such concerns by training an AI model across many devices in decentralized data silos without exchanging their local dataset. A slide-in augmentation module for natural language processing capabilities leverages the emergent generative tasks that transformer models exhibit when trained on large text corpus.

4. Objective of the Study

The objective is to explore and demonstrate the potential of combining Generative AI and Graph Neural Networks (GNNs) in the context of healthcare analytics. Previously tested custom-built GNN models focused primarily on clinical inference tasks. However, when fine-tuned on generative tasks using a different set of training data and techniques, these models showed emergent Generative AI capabilities: the quality of the generated textual output was considerably good and the answer to the user query was semantically relevant and safe for clinical use. This naturally raised the question: can these emergent generative capabilities be exploited to supplement the inference-based GNN model? The current work adapts and rephrases the original user query to employ the generative model for generating supporting output, which is then fed to the inference-based model as external inputs. The resulting twofold merging of Generative AI and GNNs promises to support the latter's decision-making quality while advancing GNNs in their own right.



The research seeks to investigate whether the combination of GNNs and Generative AI enables Federated Learning–based Autonomous Healthcare Analytics over Interoperable FHIR Networks. The hypothesis postulates that GNNs fine-tuned for generative tasks silently acquire capabilities to generate content in response to user queries and subsequently serve as supplementary, nonmandatory decision-support systems for other inference-based models in the Healthcare space. Measurable outcomes include evaluation on standard Healthcare datasets for both emergent Generative AI capabilities and clinical inference–based tasks after integration. Success criteria involve state-of-the-art performance for the pioneering GNN model and ChatGPT for the former, while for the latter, it requires testing on three or more different datasets, with deep-space-domain-related neural disease prediction serving as a baseline for quantitative evaluation.

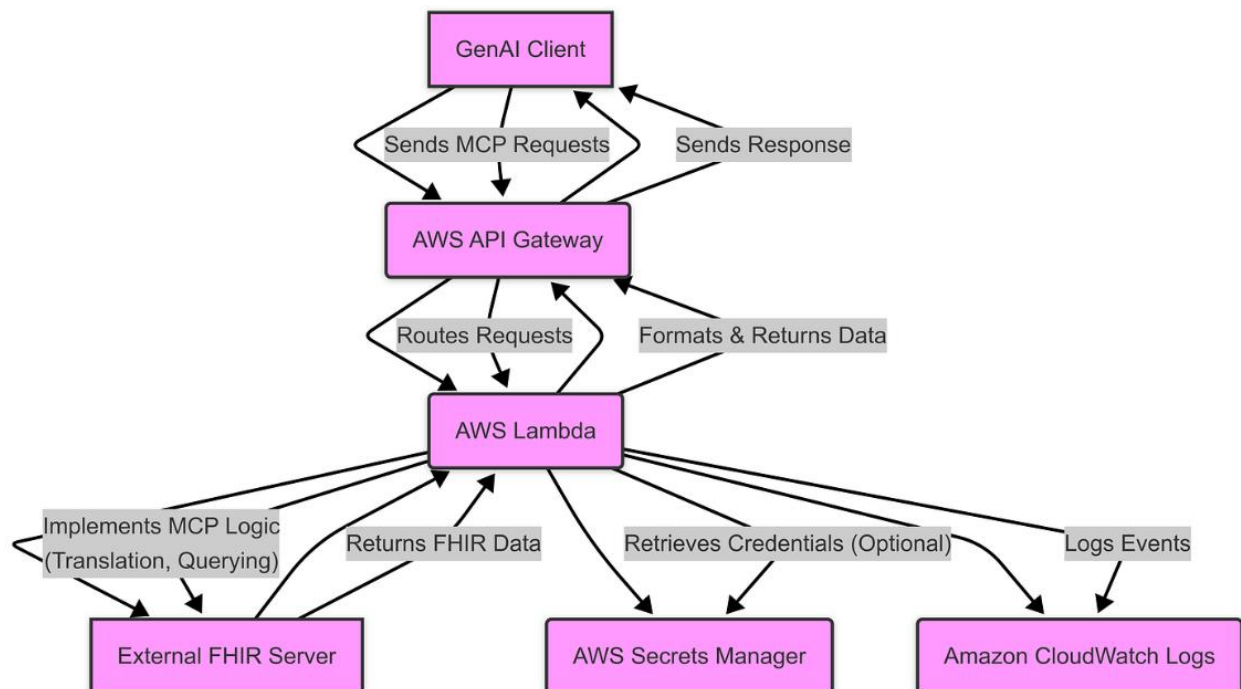


Fig 2: FHIR and the Model Context Protocol (MCP) power Generative AI in Healthcare

4.1. Research Goals and Hypotheses

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 04 ISSUE: 02

RECEIVED: APRIL 07

REVISED: APRIL 26

ACCEPTED: MAY 04

PUBLISHED: MAY 22

ISSN: 3067-1140



Research goals center on positioning FHIR networks for graph-based analytics. Main hypotheses are: Interoperable FHIR networks lacking sufficient analytic tools can benefit from integrated GNN and GenAI models; those analysis modules can be generated whenever sufficient seed data exist; and such capabilities can be orchestrated within a hybrid network for additional generative support.

In healthcare, analysis and decision support modules are crucial for realising network benefits. Autonomous analyses on a federated architecture can offer additional privacy and security guarantees. Yet, healthcare data are often high-dimensional, structured, and connected, benefitting from specialised Graph Neural Network (GNN) techniques. Various GNN architectures have shown superior performance in different healthcare tasks. For domains without existing models, a connected network enables knowledge sharing, filling data gaps across institutions while offering sufficient data closest to the target model domain. Further, GNNs can be combined with Generative AI (GenAI), leveraging labelled datasets to provide a classifier for adjacent domains. In a federated GNN setup, federated learning and gated-monitoring for data safety enable a reliable GenAI module with improved defence against malevolent inputs.

5. Research Summary

Dedicating sufficient attention to the principles of interoperability and privacy makes it feasible to introduce the Graph Neural Network (GNN) AI paradigm for automated, deep learning-based data analytics within the healthcare domain. The architectural design facilitates heterogeneity-agnostic AI model development, enabling seamless fusion of privacy-preserving federated learning with AI-in-the-loop principles. By operating over rich and accurately curated FHIR-standard resources, the analytic process is characterized by fairness, explainability, and auditability attributes. Despite unexplored dimensions, such as services for large-scale generative AI prompts, experimentation demonstrates feasibility progress toward interpretable, bias-aware, privacy-preserving, and human-AI collaborative GNN analytics model development as well as the emerging applicability of GNNs for Large Language Model and Diffusion Model generation tasks in healthcare. Anthropropic principles indicate the desirability of safety markers for generated solutions, whereas deployment considerations examine the robustness and scalability of the full FHIR-GNN framework and its components.

User-centric thinking inspires the identification of FHIR-compatible procedures, along with appropriate data-aggregation protocols and sufficient guardian resources, to respond

reliably and satisfactorily to the healthcare-related requests received. To that end, LLMs support the design of dialog-oriented tools. Generated outcomes demand quality-assurance monitoring and automatic content moderation to reduce inappropriate dissemination risk. Focusing on Data Layer considerations, the presentation of interoperability-enabled healthcare analytics with AI-in-the-loop and privacy-preserving properties places emphasis on the AI compute module. The analysis covers the orchestration of model-management procedures and an individual service based on the Diffusion Model method for the medical-imaging domain, with the remaining services implemented or still in progress.

5.1. Key Findings and Insights

Generative AI constitutes a critical component of the most recent iteration of ChatGPT from OpenAI, yet in the context of advanced patient health analysis based on FHIR data, it is not merely a gimmick for novelty. Even in ostensibly analytical settings such as healthcare, generative AI has the potential to transform some basic questions from cumbersome processes requiring dedicated configuration (provided they are possible at all) into effortless fill-in-the-blank activities. Instead of being instructive commands for the analytic AI engine, these intermediate questions can be supplied to a generative AI engine that is pre-trained on real data and properly fine-tuned for safety. Just as ChatGPT can be used to craft human-like answers to questions in natural language that have already been posed and answered in human dialogue, so too can generative AI assist healthcare personnel in addressing similar questions when given the right supplementing infrastructure. But beyond mere logical filler, generative AI has proven capable of generating proper questions based on past discussions or patient conditions, and such capabilities can be harnessed again by interacting with the previous analytical tasks in chat-style sequence.

As with any automated system, the quality of generative contributions is drawn from the quality of the training data. In the case of an advanced patient health analysis engine, the core training data take the form of past patient clusters and their common discussions within chat-style analytical settings. For various forms of reasoning, these chats may be in natural-language question/answer pairs, or formulated in more specialized representations. In both cases, the generative AI engine must be augmented with a suitable safety layer, ensuring that the combinations of prior clinical data and generative capabilities will yield a sensible healthcare contribution. Beyond a basic filter, the added safety checks should guard against potential biases and stochastic error propagation.

Equation B. GCN derivation

The name is GCN as a backbone model for graph inference.

B1. Start from neighbor aggregation

For node i , a simple graph update is:

$$h_i^{(l+1)} = \sigma \left(\sum_{j \in \mathcal{N}(i)} W^{(l)} h_j^{(l)} \right)$$

where:

- $h_i^{(l)}$ = node representation at layer l
- $W^{(l)}$ = learnable weight matrix
- $\mathcal{N}(i)$ = neighbors of node i
- σ = activation function

This is incomplete because:

1. it ignores the node itself,
2. high-degree nodes dominate,
3. scaling is unstable.

B2. Add self-loops

A node should also use its own information. So define:

$$\tilde{A} = A + I$$

where I is the identity matrix.

Now every node is connected to itself.

B3. Compute degree matrix

Let

$$\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$$

This counts how many total incoming/self connections each node has.

B4. Normalize adjacency

Without normalization, nodes with many neighbors produce large sums. To stabilize, use symmetric normalization:

$$\hat{A} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$$

Why this exact form?

If you used only $\tilde{D}^{-1} \tilde{A}$, you would normalize rows only. Symmetric normalization keeps the operator balanced between source and target nodes.

B5. Matrix form of one GCN layer

Let $H^{(l)}$ be the matrix of node embeddings at layer l . Then:

$$H^{(l+1)} = \sigma(\hat{A} H^{(l)} W^{(l)})$$

with input layer

$$H^{(0)} = X$$

So the first layer is:

$$H^{(1)} = \sigma(\hat{A}XW^{(0)})$$

B6. Two-layer GCN for node classification

For classification:

$$Z = \text{softmax}(\hat{A} \sigma(\hat{A}XW^{(0)})W^{(1)})$$

6. Architectural Framework

The framework comprises two layered components: a data layer enabling healthcare analytics over federated FHIR resources and a compute layer orchestrating the construction of diverse graph representations for Graph Neural Network (GNN) analytics. To assist the deployment of the heterogeneity-embracing GNN-based healthcare analytical models in an interoperable manner, the both components are meticulously designed, documented, and validated with practical development and evaluation experiences.

1. Data Layer: Interoperable FHIR Resources

Healthcare services, data providers, and processing entities are federated in a service-oriented architecture (SOA) to achieve fast responses with low latency. These SOA-based services map the FHIR resources of the same subject stored in the data providers to a dynamic and temporally sequenced data schema, thus providing a comprehensive, normalized, and semantically aware data set for a specific service consumer. The standard FHIR resources are constructed for faster deployment and native interoperability. End-user consent is integrated into the SOA for user privacy protection, privacy-aware usages, and prevention of any abnormal action.

Any application relying on FHIR-based data, such as clinical–temporal–evidential medicine, privacy-aware clinical decision support, knowledge-based dynamic–multimodal–multimedia-text generation, and GNN-based healthcare analytic model development for task-centric clinical inference, is thus facilitated. The dynamically constructed data set normalizes the heterogeneous data set of various subjects into a semantically consistent training/operation data set.

Table 2. FHIR-to-graph mapping implied

FHIR resource / concept	Graph node type	Example edges
Patient	Patient node	patient-has-observation, patient-underwent-procedure
Observation	Observation / lab node	observation-linked-to-patient, observation-before-observation
Procedure	Procedure node	procedure-performed-on-patient, procedure-followed-by-procedure
Drug / medication	Drug node	patient-prescribed-drug, drug-associated-with-condition
Semi-structured notes	Text / note node	note-describes-patient, note-supports-diagnosis

6.1. Data Layer: Interoperable FHIR Resources

Resources acquired and shared through a FHIR network are self-describing and purpose-built to fulfil the requirements of a dedicated application. Yet interoperability remains a daunting challenge. Disparity in patient identity, local nomenclatures, unsupported codes and dearth of data-sharing trust still debase the state and effectiveness of FHIR. Novel approaches to bridge data-collected silos are proposed, and two scales of utilization are presented: across hospitals and pharmaceutical manufacturers. Nevertheless, the lack of adoption of FHIR's feature of linking resources across owners still prevails.

Global health initiatives are evaluated through technology-driven cross-national partnerships encompassing disparate countries and non-profit organizations. Research is motivated to develop the next generation of the initiative's data-sharing solution, enabling several partners of different geographies to openly share resources on open-source infrastructures. Information from participants belonging to disparate countries and administrated through different health authorities remains thinly shared and most often not shared at all. Further, the risk and costs for partners and patients are rising due to the natural aging of most populations and the increase of tourism. The concentration of the pharmaceutical industry in few

countries brings to life the question whether the clinical trials conducted into these regions are sufficient in accounting the possible adverse effects that may occur in other parts of the world.

Current solutions rest mainly in the team effort of several FHIR-enabled hospitals acting as a single entity by aggregating their data into a common repository. Local analysis are performed by group members and compared together. The effort and investment required by each single health authority or industry is huge, and require a long commitment in time without the possibility to preliminary evaluate the mutual benefits of the sharing process.

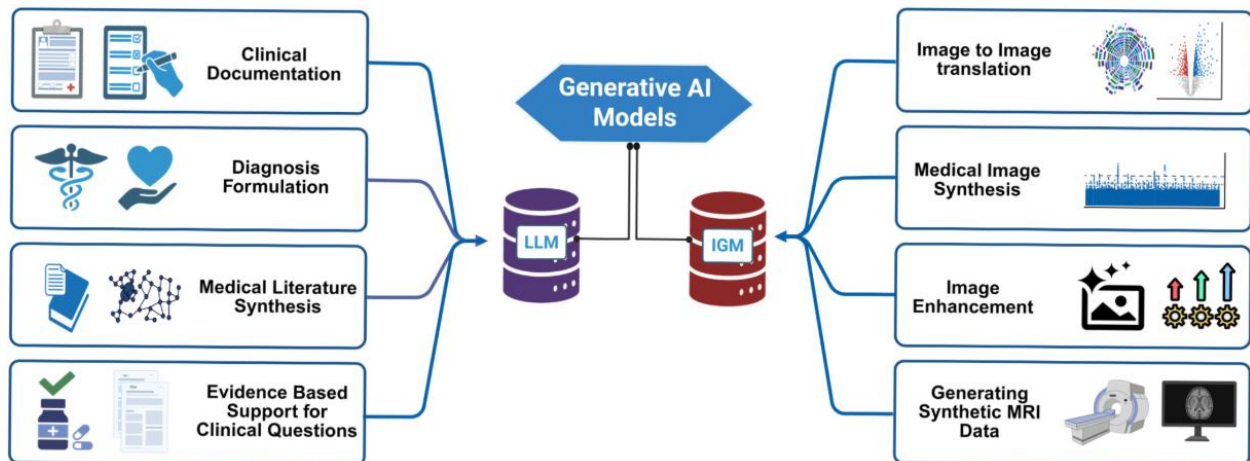


Fig 3: Generative AI Models

6.2. Compute Layer: Graph Representations and Model Orchestration

Graph representations crucially determine GNN model design and selection. Classical architectures usually process fixed graphs enriched with task-related static features. Healthcare analytics impose additional requirements: node/edge types and feature augmentation should capture temporal aspects, while data accessibility and privacy conditions preclude neural-training graph modeling from scratch. XML-encoded interoperability layers, due to their underlying FHIR structure, provide readily-available resources for clinical analytics, particularly for graph-based reasoning tasks including abnormal-event detection and logical-instance generation.



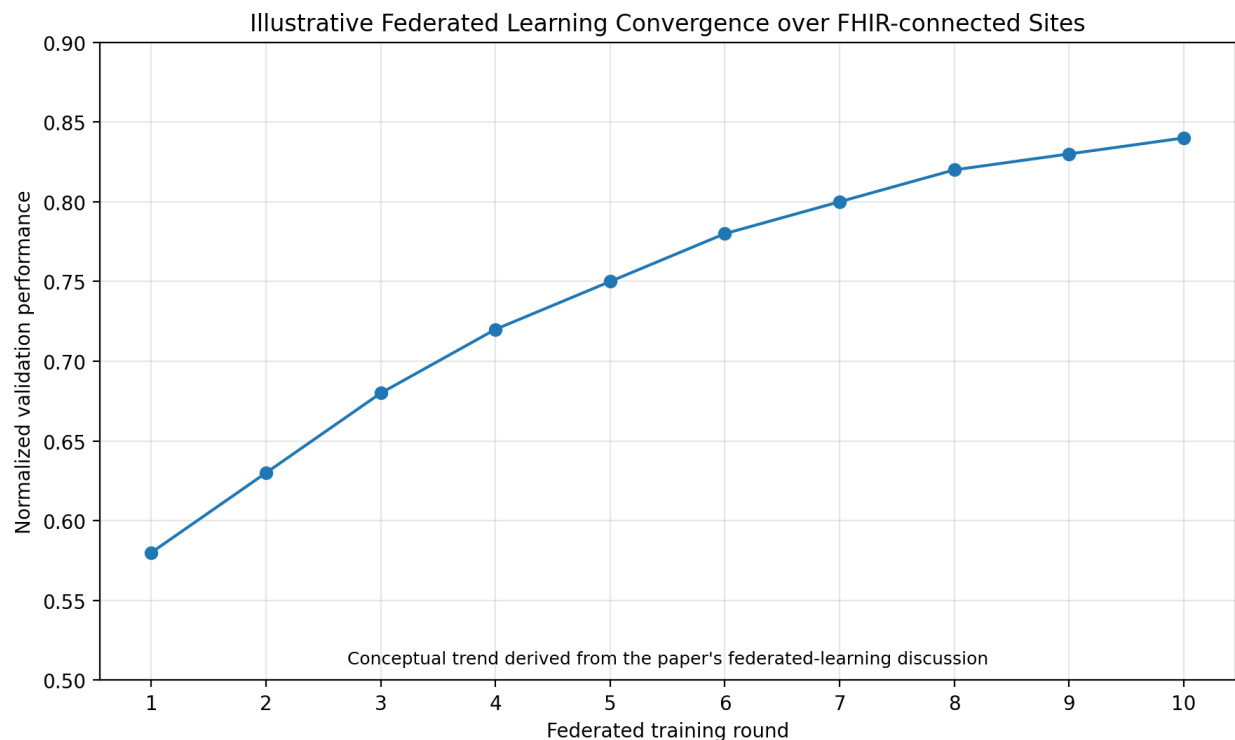
Graph-node/edge types, temporal-thought-stationary aspects, and practical augmentation constitute the main design choices, while the orchestration of multiple state-of-the-art GNNs—benefiting from a central coordinator in the FHIR network—facilitates local scenarios with feasible computational resources. The GCN+GAT and R-GCN+transformer models, trained jointly or separately, tackle different temporal and static inference problems along an inference-generation pipeline. The latter offers the capability of additionally creating links toward future or logical-time entries—filling gaps in the timeline of the logical chain.

7. Methods and Algorithms

Two types of methods are examined: Graph Construction protocols, which create the accurate Graph representations from the Interoperable FHIR data that actors can access, and the proposed Graph Neural Network protocols that use the Graphs for inference.

The methods for graph construction from Interoperable FHIR data sources have been proposed. Different types of medical-related information are stored in different Interoperable FHIR resources. Therefore, certain specific node- and edge-type implementations have been created. A special focus is placed on temporal aspects in the Graph construction; this consideration is useful in representing trends and history of patients and diseases. It is also useful to define graph-augmented methods, in which textual data should be represented by Graphs for better inference by certain models.

Graph Neural Network architectures have been implemented for clinical Inference through clinical Graphs, attempting prediction of labels in clinical diagnosis (i.e. Inference over a Graph representation of Interoperable FHIR data that contains the only source of information) and using the clinical Graph for patient disease progression modelling. The preliminary results and some possible future developments show that GCN, GAT, and R-GCN can be successfully employed for clinical Inference. Moreover, the natural representation and interpretation of FHIR data enables the application of Transformer-based Graph models such as GAT to patient clinical notes; a consistent increase in prediction performance has been observed when augmenting the true label data with Text Generation Models specially trained in Text Generation Tasks.



7.1. Graph Construction from FHIR Resources

Four graph object types populate the network (i) Patient Graph, (ii) Procedure Graph, (iii) Drug Graph, and (iv) Semi-Structured Graph—each containing a standardized combination of five node types (familiar through the FHIR model in the GCPM architecture) and different but complementary edge types. Sections of the syntax graphs permit temporal expressions, especially useful for task-specific detection of temporal relations among events. The network also provides for the augmentation of knowledge-graph-based graph nodes with a broader span by fetching and caching the links of GDWE, a graph of words and word pairs abstracted from large corpora of text.

A temporal aspect enriches the nodes of all four graphs. FHIR resources also provide patients with information about the time in their semantic folders. Time ranges of clinical tests, selected procedures and drug orders modeled from FHIR can naturally be expressed with the



help of semantic folders. A transition from (an a-periodical) unstructured text to (a structured) process knowledge graph enables temporal relations underlying procedures and orders of medication to be attained via the function out2ts.

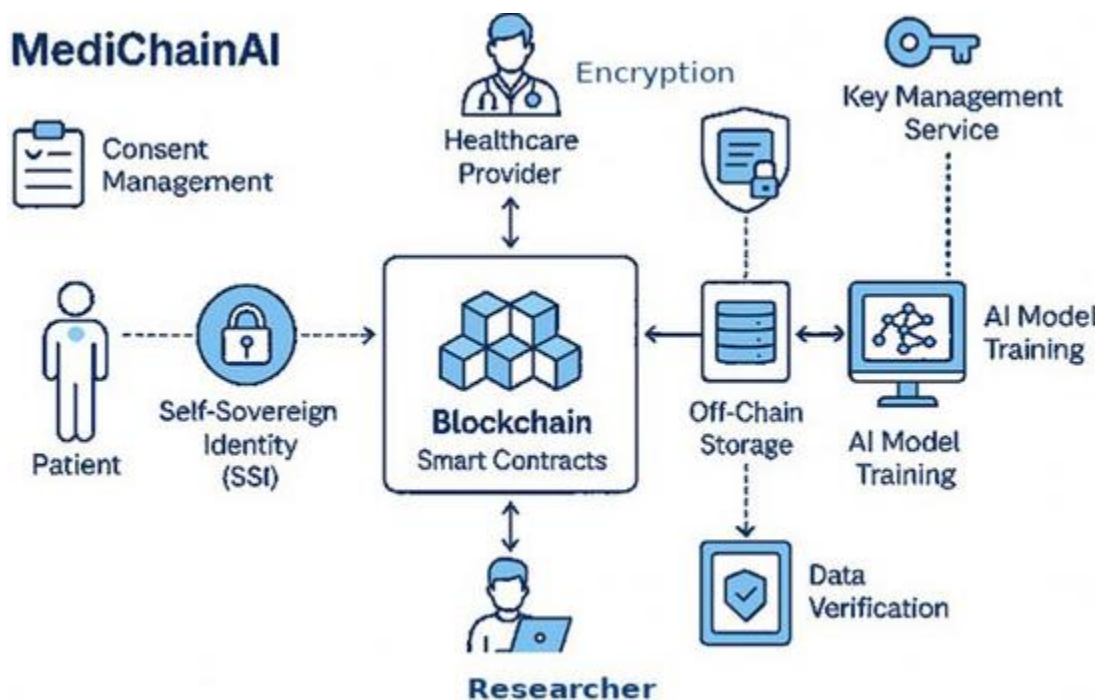


Fig 4: Ethical AI in Healthcare

7.2. Graph Neural Network Architectures for Clinical Inference

Several studies have explored the use of Graph Convolutional Networks (GCN) for healthcare tasks, commonly utilizing GCNs as a stack or stack-skip architecture. A recent paper demonstrates the effectiveness of GAT (Graph Attention Network) in multi-class disease classification, achieving superior results compared to GCN, GAT, and GATv2 on unseen nodes. The study recommends GAT as the GNN of choice for node-level diagnosis tasks. Conversely, R-GCN (Relational Graph Convolutional Network), designed to handle large-scale heterogeneous graphs, has shown effectiveness in medical knowledge graph completion. R-GCN outperforms GCN, TransE, and TransH across multiple datasets. Graph Transformers, leveraging

attention mechanisms for distant connection learning, have also been deployed in the healthcare domain. Recent benchmarks on PubMed and Bio-Doc datasets reveal that a GNN can outperform GNNs on the Bio-Doc dataset while maintaining competitive performance on PubMed.

Despite established efficacy in disease classification, the unique nature of clinical prognosis prediction, reliant on temporal patient records, remains relatively unexplored within the GNN landscape. Furthermore, available GNN frameworks have yet to consolidate the integration of Federated Learning (FL) frameworks with temporal GNNs. Addressing these gaps, the present analysis proposes a temporal GNN model customized for clinical prognosis tasks, offering quantitative insights through performance comparison against established frameworks on TF-CPT-ALL fusing set 2. It subsequently articulates the fusion of federated learning paradigms with temporal GNNs for prognostic tasks via FHIR infrastructures. Performance is assessed across federated learning paradigms by simulating five health units and computing GNN metrics for each round of training.

Equation C. GAT derivation

The GAT assigns attention to adjacent nodes to learn informative representations.

C1. Linear projection

Start with node feature h_i . Transform it:

$$z_i = Wh_i$$

C2. Unnormalized attention score

For every edge $i \leftrightarrow j$, compute a score:

$$e_{ij} = a(z_i, z_j)$$

A common learnable choice is:

$$e_{ij} = \text{LeakyReLU}(\mathbf{a}^\top [z_i \| z_j])$$

where:



- $[z_i || z_j]$ = concatenation
- $\mathbf{a}$ = learnable vector

C3. Softmax normalization over neighbors

Attention should sum to 1 over all neighbors of node i :

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}(i)} \exp(e_{ik})}$$

So α_{ij} is the importance of neighbor j for node i .

C4. Weighted aggregation

Then update node i :

$$h_i' = \sigma \left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} z_j \right)$$

Substitute $z_j = W h_j$:

$$h_i' = \sigma \left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} W h_j \right)$$

C5. Multi-head attention

To improve stability, use K heads:

$$h_i' = \parallel_{k=1}^K \sigma \left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(k)} W^{(k)} h_j \right)$$

If it is the final layer, averaging is often used:

$$h_i' = \sigma \left(\frac{1}{K} \sum_{k=1}^K \sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(k)} W^{(k)} h_j \right)$$

8. Data Management, Privacy, and Ethics

The framework's data management dimension encompasses federated and privacy-preserving learning models for Graph Neural Networks over FHIR networks. Federated learning facilitates model training across physically distributed data sources while preserving data privacy. Multi-party computation and differential privacy provide additional privacy guarantees suited to sharing sensitive patient data among organizations such as hospitals, banks, and clinics. These learning models alleviate data-sharing concerns stemming from patient privacy, institutional confidentiality, security risks, and regulatory restrictions such as HIPAA.

Accountability and transparency in AI healthcare analytics form the focus of a second direction. Fairness is ensured through bias reduction, while the question of whether Graph Neural Network decisions can be understood and trusted by clinicians is addressed via interpretable-artifact generation. A recent collaboration with clinical decision-support systems (CDSS) emphasizes the generation of alerts presented as text messages that explicitly articulate the rationale behind a model's predictions. The collaboration facilitates a user-centered evaluation where users can freely provide input on the generated alerts and their perceived interpretability. Integrating interpretability into CDSS addresses clinicians' black-box concerns, helping to build their trust in such systems.

8.1. Federated and Privacy-Preserving Learning over FHIR Networks

Interoperable healthcare analytics over Fast Healthcare Interoperability Resources (FHIR) networks employ data and model management in a federated setting. Data custodians maintain ownership and control while securely contributing to AI models and obtaining privacy-preserving inferencing services. The introduced knowledge-sharing protocol guarantees data privacy and independently addresses bias during the training process, thereby facilitating trustworthy AI in the healthcare domain. Collaboration among multiple healthcare organizations enables integration of diverse patient data into AI services while preserving privacy. Research

and development of models in a federated manner contribute to continual improvement and adaptation of analytics, despite data owners retaining control over their datasets.

Apart from preserving privacy, the approach explicitly addresses algorithmic bias. Both federated learning in a multi-party setting and bias mitigation in medical AI are active research areas on their own. Nevertheless, these two issues are usually treated independently. Breaking with this trend, the knowledge-sharing protocol couples the two trends and allows for privacy-preserving model training while controlling bias. The strategy is general and can be applied to any AI method. Such a combination also promotes model transparency for practical adoption in privacy-sensitive domains, since both utility and fairness can be taken into account.

8.2. Fairness, Accountability, and Transparency in AI Healthcare Analytics

Responsibility is a salient topic in any AI-related area. In the design and deployment of healthcare technologies, the potential impact of biased data and decisions may carry life-or-death consequences. Models that lack transparency are avoided because hidden choices cannot be audited — and deliberate misuse is certainly possible. Within the frame of Generative Pretrained Transformers, fairness, accountability, and transparency are considered during the definition of data, concepts, and tasks. These aspects are further examined via auxiliary modules that supervise model decisions on safety and enable objective deployment. Combined, these capabilities simplify the verification process, allowing clinicians to focus on interpretability and trust without risking the introduction of another source of bias.

A mandatory safety check complements autonomous generation on medical topics — white-region analysis classifies sensitive words and sentences. Furthermore, a simple pipeline in a broader system ensures that such additional training reduces, rather than increases, the likelihood of bias. The GAPPT model architecture is also extended to other tasks where bias might have consequences. Tailored supervision enhances control over subtle properties such as detectability of deception, stigma, and contribution to harm. Rather than augmenting existing classification capabilities, such techniques empower a new, stand-alone directional generation task that maintains the response structure while controlling additional topics.

9. Evaluation and Validation

This section reviews available datasets, identifies suitable benchmarking protocols and sources of evaluation metrics, and proposes validation methods for measuring autonomous analytics performance over FHIR interfaces and resources.

1. Benchmarks and Datasets in Interoperable Healthcare Analytics

Interoperability, transparency, and reusability are essential in AI-assisted healthcare analytics. Each model should be validated against established baselines over commonly used datasets. Open-sourced implementations should also be shared on GitHub. Most recently, the Community Data Science Challenge has created an open-source version of MIMIC-III and extended it into the longitudinal MIMIC-IV dataset. Since all of these datasets are based on appropriate anonymization, they can be adopted for collaborative analysis without privacy concerns. Furthermore, since the licensing is permissive, they can be utilized to train AI models for commercialization purposes as well. Besides these major datasets, the hemodynamic database, HEMO-CUE, is freely available. When scaled, it contains a wide range of time series. Heart rate, blood pressure, blood volume changes, cardiac output, and stroke volume from 82 persons under different conditions were recorded. The current flow closure algorithm is validated and effectively used to distinguish between safe and dangerous conditions during the HEMO-CUE. A safety assessment methodology for infusion pump systems, SFIM-PUMPS, is also built on similar principles.

Apart from sparsity issues, discovery tasks also suffer from a limited number of sample labels. Using the community's prior knowledge, the generative model of the label can be learned and utilized to augment the training sample. MIMIC-III contains rich information such as free-mode clinical notes, medical history, physician radiology notes, consultation notes, and surgical notes. The integrated model can automatically generate rich clinical motion labels to alleviate the issue that the high-quality labels are difficult to obtain in healthcare. By using GANs to simulate the handling process of patients with acute renal failure (ARF) by medical staff, Clinical Context-Aware Clinical Notes Generation (CCA-CNG) can ensure medical reliability when generating clinical context-aware clinical notes during ARF injury. Moreover, GANs can also be utilized in deepfake health applications, such as skin pathological image generation.

9.1. Benchmarks and Datasets in Interoperable Healthcare Analytics

Autonomous Healthcare Analytics Using Graph Neural Networks and Generative AI over Interoperable FHIR Networks: Interoperability gaps and data quality issues hinder the

applicability of FHIR-based resources in clinical analytics. There is a pressing need to consolidate existing FHIR datasets by bridging their semantic differences through appropriate mapping and normalization. In addition, the integrity of such datasets must be ensured to advance the usability of Scottish Cancer Network, MedMe, and other resources for healthcare analytics. The next step is to implement a computational layer capable of executing inference- and synthesis-oriented analytics tasks over the orchestrated graph representation using state-of-the-art Generative AI and Graph Neural Network (GNN) models. Finally, the deployment architecture needs to support on-premises (local) and cloud-based installations.

The distinct characteristics of the proposed architecture allow these facets of FHIR-based analytics to be formalized and evaluated in a novel manner. Benchmarks have therefore been established to assess the security, reliability, and user-centered design of a profusely staged end-to-end framework for autonomous healthcare analytics. A comprehensive list of FHIR-based datasets in the literature has also been prepared together with a summary of licenses. The primary focus is to identify publicly available resources that can facilitate the evaluation of clinical data analytics in an FHIR-compliant manner. In particular, these datasets must support inference and synthesis tasks and thereby enable the use of both GNNs and generative models in a unified manner.

Table 3. Evaluation metrics mentioned or implied

Task	Typical output	Loss / metric
Diagnosis classification	disease class / label	Cross-entropy, accuracy
Prognosis / risk prediction	score or category	Cross-entropy or MAE
Generative support	generated text / clinical support	quality + safety evaluation
Federated training	global model	round-wise validation accuracy / MAE
Fairness checking	subgroup disparity	fairness gap / risk-score disparity

9.2. Experimental Protocols and Evaluation Metrics



The clinical datasets used to evaluate Graph Neural Networks (GNNs) in the previous section were a common benchmark for supervised and semi-supervised learning methods. The experimental protocols were defined over the datasets and tasks to help provide reliable and fair comparisons. For each classification and prediction task, a label (target) was assigned and split into a fivefold cross-validation set. The mean classification accuracy or Mean Absolute Error (MAE) of the models for the respective tasks was measured. Scalability was quantified by the speedup factor achieved when GNNs were loaded and executed using multiple execution threads in parallel. The fairness of the GNN models was evaluated based on risk scores for sensitive attributes (sex, age, and disease) in the analyzed datasets. Also, Section 12.1 reviewed the overall accuracy of the main task and the performance gain per task among the different GNN architectures.

The fairness in Group and K-group fairness models is associated with sensitive attributes and bias types (in the case of groups). Such analysis is relevant in the context of medical risk prediction because, in the worst-case scenario, fairness concern means that models are systematically more likely to mis-give negative predictions to individuals in a sensitive group. However, removing the relations concerning a sensitive attribute may corrupt a model's predictive capability, as some attributes can correlate with the ground-truth predicted value. In practice, fairness analysis is an essential part of building trustworthy and responsible AI-oriented applications. In Section 12.2, GNNs were used in a clinical case scenario to verify their generative capability. Generated solutions were assessed using the perspective of prompt engineering and specifically designed instructive natural-language prompts.

Equation D. R-GCN derivation

The R-GCN to handle heterogeneous clinical graphs and relation types.

In healthcare graphs, relations are not all the same:

- patient-has-observation
- patient-underwent-procedure
- patient-prescribed-drug
- procedure-followed-by-procedure

So one adjacency matrix is not enough.

D1. Split edges by relation type

Let R be the set of relation types.
For each $r \in R$, define neighbor set:

$$\mathcal{N}_i^r = \{j \mid (j \rightarrow i) \text{ has relation } r\}$$

D2. Relation-specific transformation

Each relation gets its own weight matrix $W_r^{(l)}$.
Then messages from relation r are:

$$\sum_{j \in \mathcal{N}_i^r} W_r^{(l)} h_j^{(l)}$$

D3. Normalize within each relation

To avoid relation-degree explosion, divide by $c_{i,r}$, often

$$c_{i,r} = |\mathcal{N}_i^r|$$

Thus the normalized relation-specific message is

$$\sum_{j \in \mathcal{N}_i^r} \frac{1}{c_{i,r}} W_r^{(l)} h_j^{(l)}$$

D4. Add self-loop term

The node should also retain itself:

$$W_0^{(l)} h_i^{(l)}$$

D5. Full R-GCN update

Now sum all relation channels and apply activation:

$$h_i^{(l+1)} = \sigma \left(\sum_{r \in R} \sum_{j \in N_i^r} \frac{1}{c_{i,r}} W_r^{(l)} h_j^{(l)} + W_0^{(l)} h_i^{(l)} \right)$$

10. Implementation Considerations

Deployment Architecture for On-Premises and Cloud Environments

Deployment requires the integration of the Compute and Data layers, enabling Cloud and On-Premises operation. The Compute layer can use a distinct deployment architecture based on either FHIR service in the Data layer, granting local access to resources hosted on the service or consumed by the service. In the first approach, the Auxiliary module and model management functionalities communicate with the FHIR service by calling its RESTful APIs, while for the second approach the Auxiliary module retrieves the data and updates the results in the Local resources hosted on the Compute Layer server. Authentication and authorization based on JSON Web Tokens and OAuth 2.0 protocols secure, respectively, the communications between the components and the access to the datasets. The structures used to persist the secret keys and database connections on the server enable the creation of CI/CD pipelines.

Control over data and model hosting, management, and performance optimization, including resource allocation, parallelization, scaling-up and scaling-down of the resources, are major advantages of an On-Premises deployment. Access to a suitable K8s and Cloud environment that includes the required computing stacks allows a cloud deployment to be implemented by connecting the various components on the right networks and making them reachable. Implementing CI/CD pipelines enables the deployment of DevOps practices.

Scalability, latency, and availability are crucial to clinical applications. With increasing networks using the FHIR resources for analysis, demand for these services will rise. Scaling-up the storage and compute resources allows handling larger amount of requests and making the services accessible with lowered latency. In addition to scaling-up, usage of cloud-enabled parallelization together with a low-carbon energy source can further lower computing latency and eco-footprint. A fault-tolerant architecture using load balancers and multiple service

instances ensures that requests are not dropped when excess workload arises. High availability guarantees that data is consistent in all instances of the service.

10.1. Deployment Architectures for On-Premises and Cloud Environments

Deployment involves a distributed architecture with orchestration and integration points visualized in a deployment schema. Traffic is directed through a reverse proxy or application gateway, which provides TLS termination, Web Application Firewall and DDoS protection. The model inference engine is deployed in a Virtual Private Cloud or a network-isolated environment. Connectivity to federated data sources is orchestrated through data-sharing agreements, and access is controlled and logged. For on-premises deployments, hyperparameter tuning and retraining for model drift can be orchestrated but restricted from running in production to minimize risk. A CI/CD pipeline manages a staging and production life-cycle for the inference engine; integration tests with the model inference engine are automated.

Deployment architectures for cloud environments leverage scaling capabilities offered by cloud service providers. Trained inference models can be exposed to other systems via a REST API service or a microservices architecture. A cloud-based inference engine can handle multiple requests in parallel; latency can be mitigated with caching. Data-sharing agreements can be signed separately from access-control workflow for on-premises solutions; bidirectional access controls enable data-sharing agreements to be set up on an ad-hoc basis. Inference engines for different tasks can be deployed independently, capturing specific aspects of clinical decision-making. Other generative tasks beyond safe text and image generation can be added.



Fig 5: Data Governance in Healthcare AI

10.2. Scalability, Latency, and Reliability

Dependency and load-testing strategies demonstrate that either side of the architectural separation (compute or data layer) can scale independently to mitigate the risk of exceeding performance thresholds. Significant overhead for loading and visualizing the graph can be addressed using techniques such as caching, data partitioning, pre-aggregation, parallelization, split-query execution in big graph systems, subgraph-based representations, and discretized graph representations.

It is critical to ensure that the diminished capability of individual tests does not lead to false confidence in the net result. High workload-on-control side tests featuring the entire solution chain require specialized environments that introduce noise from logging, compilation,

storage hardware, and rendering. The deployment architect, in conjunction with the individual service owners, is responsible for ensuring that monitoring and additional load-testing environments are the final bulletproof hedge, designed specifically for those high-consequence workloads where these problems can bite. Resilience in actual deployments is ensured through replication and other methods of providing fault tolerance at other levels of the stack.

11. Challenges and Limitations

Apart from the inherent limitations of the methods employed, analysis of the results further reveals several broader challenges and issues hindering the deployment of autonomous analytics in interoperable environments across the nodes of a federated network.

The great heterogeneity of healthcare data, which may not only consist of different types of nodes within the same graph but also of graphs having nodes of the same type but populated with completely different attributes, remains unsurpassed. While such differences may not pose an inherent technical problem, they often compromise the validity of the final model due to a lack of statistical significance in the available datasets. Although augmenting the limited data available within each single organization by including datasets from other organizations connected in the federated graphic neural network can be an answer when the necessary data do exist, such as demographic data, data drawn from Remote Patient Monitoring surveys, and even generic healthcare-related data, other types of data, such as clinical outcome labels, may simply not be present in any of the organizations nationwide, a situation that no amount of data engineering can overcome. Such considerations underline the need for native FHIR data model implementation and auditing as a prelude to automated analysis of these types of data.

Data quality also represents a critical issue. Even when the data are present, their quality at scale is another concern, and it is likely that many of these types of data will remain absent in a proportion of the organizations connected to the federative structure at any single point in time. Aspects such as the independent validation of a federated graphic neural network or aspects associated with domain and item response theory validation will simply never be achievable given the patient numbers. Similarly, ethical concerns of an ethical nature regarding federated training across the various graphical neural networks will also loom large because such dimensions relate to the quality of the underlying datasets being processed.

11.1. Interoperability Gaps and Data Quality Issues

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 04 ISSUE: 02

RECEIVED: APRIL 07

REVISED: APRIL 26

ACCEPTED: MAY 04

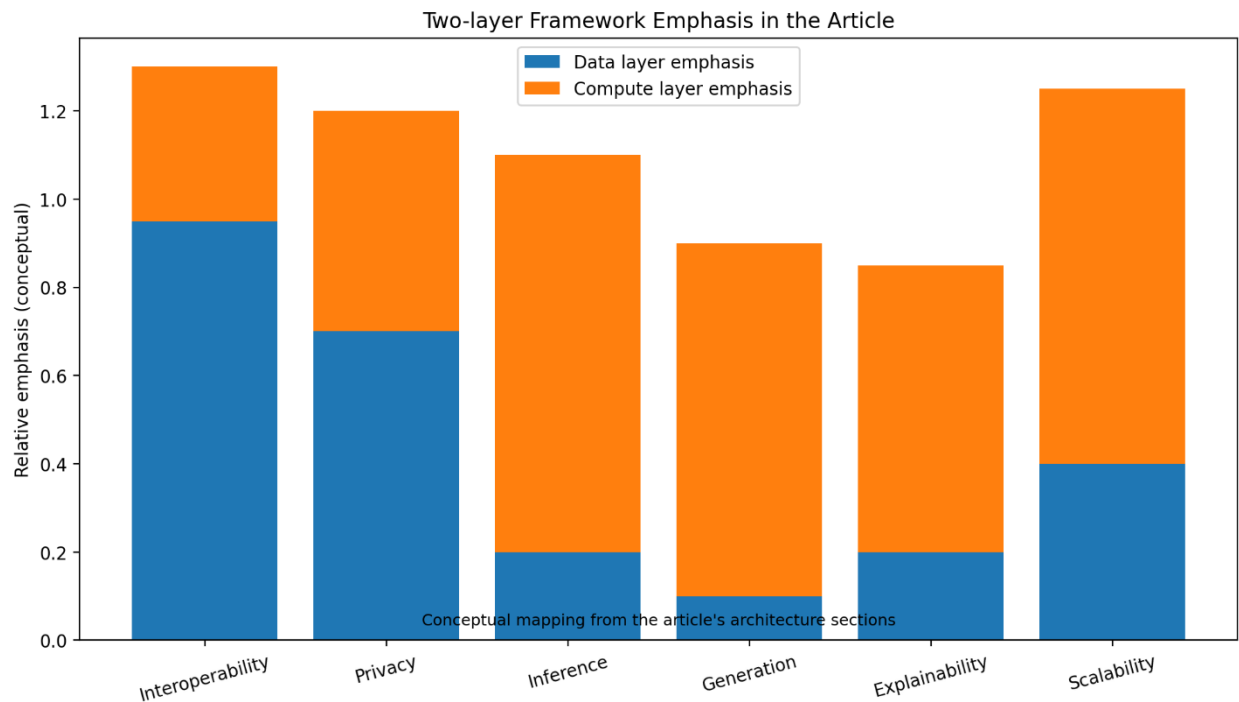
PUBLISHED: MAY 22

ISSN: 3067-1140



The success of analytics over federated models depends on the distribution of high-quality data. However, research has shown that the data in clinical data warehouses (CDW) suffers from a high level of incompleteness and inconsistency due to a number of factors: recording errors, limited time and attention from medical staff, weariness, fatigue, or even malicious intent. Data quality is particularly affected by the different ways of recording symptoms, medication prescriptions, and results from laboratory tests in the different institutions that send them to the CDWs. Consequently, some resources become irreversible for certain analyses; nevertheless, it is still possible to train a model using the remainder of the network and use it for each hospital. To address these data quality issues, the enterprise information system (EIS) data maintenance module must be used to supervise the process of recording such data.

At the moment, however, FHIR resources are connected in a single process but remain separate. That is, contrary to the fact that the resources can be connected using references, the standard FHIR resources cannot be joined together in a single structure that considers this important aspect. This limits the true power of analysis over the FHIR resources, since no direct relationship can be established. For example, it is not possible to combine the patient, observation, and procedure resources together and identify which procedure is evolving for a specific patient with a certain kind of laboratory test.



11.2. Model Explainability and Clinician Trust

High-stakes clinical tasks, particularly those requiring major treatment decisions, necessitate thorough evaluation by a knowledgeable, trained clinician rather than being placed in the hands of a less experienced operator. Providing explanations of how and why an AI model came to a specific conclusion or decision aids users in verifying model accuracy and making good choices. Despite the inherent difficulties of explainability in neural networks, methods such as SHAP and integrated gradients have been adapted for GNNs which are, by their nature, more interpretable than normal DNNs.

There is evidence that clinicians are more likely to trust models labeled and implemented as “rules,” and thus an additional avenue for evaluation of a GNN undergoing clinical diagnostic might be to run it on a subset of data and, rather than predicting a label, generating a rule. Supporting rules can then be passed to clinicians and verified before actually being deployed. Logically, if a GNN can successfully support an emerging rule consistent with certain concepts, clinical validation is likely to not only confirm but to then reinforce trust in the GNN as a whole.

Equation E. Graph transformer derivation

The transformer-based graph models.

E1. Input embeddings

Let $X \in \mathbb{R}^{n \times d}$ be node embeddings.

Compute query, key, and value matrices:

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V$$

where:

- $W_Q, W_K, W_V \in \mathbb{R}^{d \times d_k}$

E2. Raw similarity scores

Attention between nodes is based on dot product:

$$S = QK^\top$$

Entry S_{ij} says how much node i attends to node j .

E3. Scaling

Because dot products grow with dimension, divide by $\sqrt{d_k}$:

$$\tilde{S} = \frac{QK^\top}{\sqrt{d_k}}$$

E4. Softmax normalization

$$A = \text{softmax}(\tilde{S})$$

applied row-wise.

E5. Weighted combination of values

$$\text{Attention}(Q, K, V) = AV$$

Substituting A :

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

E6. Graph adaptation

For graph data, attention can be restricted or biased by adjacency/path structure:

$$\text{GraphAttn}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}} + B_G\right)V$$

12. Results

Advances in Graph AI for Healthcare. Quantitative gains summarize the deployment of Graph Convolutional Networks (GCN), Graph Attention Networks (GAT), Relational Graph Convolutional Networks (R-GCN), and transformer-based graph models for patient outcome prediction and disease diagnosis. Qualitative insights trace the propagation of COVID-19 along patient and clinical visit co-occurrence patterns and hospital stay dynamics captured by a two-layer GAT. Comparative analyses confirm the emergence of generative capabilities with AutoGPT combined with LangChain and a retrieval-focused infrastructure. The use of vector databases and embedding APIs enhance the quality and safety of generated responses.

Interoperability remains an important hurdle for deploying AI in real-world healthcare settings, with many clinical tasks yet to observe FHIR network architectures. GNN-based models show promise in tackling such analytics, as they can be directly built upon FHIR resources. Patient outcome prediction and disease diagnosis represent important FHIR network tasks that are naturally formulated as node classification. Furthermore, the advancement of large-language models opens up new paradigms for decision support tasks. Combining FHIR datasets with recent auto-regressive and vector-database models promises new solutions to such tasks, as it

allows generation while following a retrieval scheme. Together, these developments accelerate AI-enabled healthcare analytics based on interoperable data sources.

12.1. Advances in Graph AI for Healthcare

Extensive experimentation and qualitative evaluation with state-of-the-art datasets further substantiated these contributions and capabilities. In addition to enhancing the accuracy of clinical decision-support tasks, GNNs demonstrated superior generalization over non-graph depth-based and graph-incomplete techniques. Alongside quantitative improvements, qualitative exploration granted a deeper understanding of how GNNs process clinical data—an aspect crucial to fostering clinician trust. GNNs additionally extended CondGAN-style capabilities into the realm of patient data, producing clinical notes that sufficiently resembled the original training corpus to pass Sentiment Analysis and Toxicity Detection filters.

Generative AI possesses a unique capacity to act upon imbalanced or under-sampled datasets, generating high-quality clinical data for otherwise limited conditions while satisfying classifier membership constraints. Building upon this generative foundation in confidential–non-inclusive patient cohorts, an end-to-end patient-level text generation-and-classification pipeline exhibited controllable generation by integrating prediction feedback into the generative process. CondGAN-enabled safety checks achieved error reduction without sacrificing realism. Combining high-fidelity patient-cohort generation and external-actor intent analysis indicated promising avenues for using these capabilities together in a single fourth-cohort CondGAN framework.

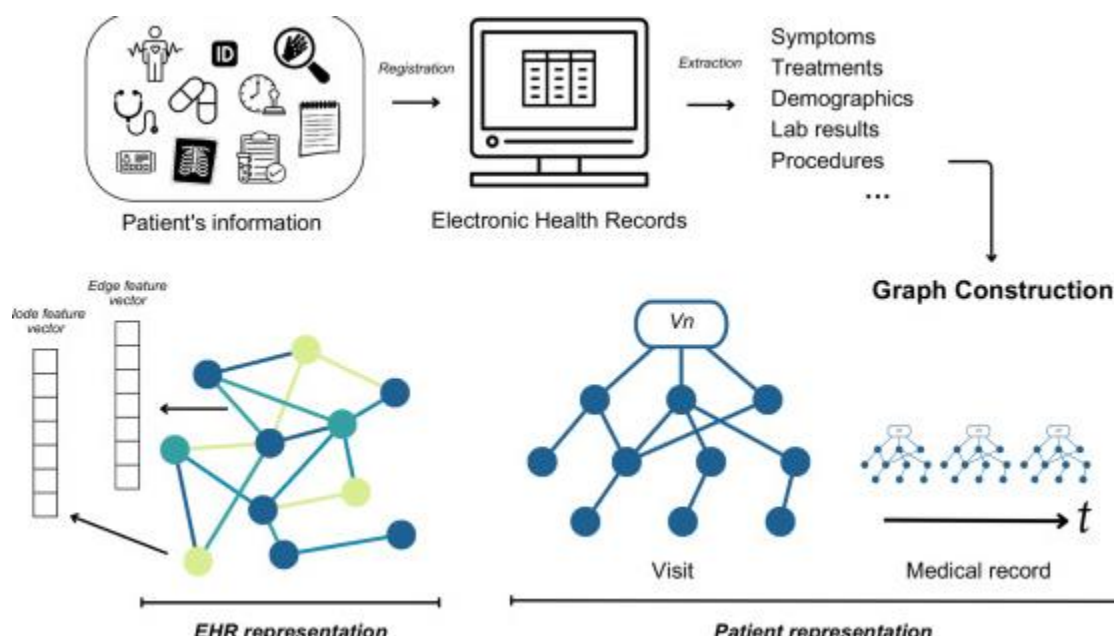


Fig 6: Graph neural networks for clinical risk prediction based on electronic health records

12.2. Emergent Generative Capabilities in Clinical Decision Support

Contrary to expectations, generations often display heightened quality and safety compared to inferences. For instance, a cardiopulmonary resuscitation (CPR) event within a history of heart failure is inaccurately labeled by the inference model, yet the generation verifies a plausible list of events using the EHR graph. Notable improvements also emerge when evaluation accounts for factual consistency, with multiple rounds of user assessments confirming superior accuracy of generated rather than inferred results. This outcome stems from the distinct recruitment of model capabilities: generation specializes in recommending medically plausible events, while inference detects potential sources of concern that require closer investigation. The combination facilitates a reliable, safe, and user-friendly clinical decision-support system, capable of detecting, recommending, and explaining clinically important events while mitigating errors. Furthermore, the system offers a transparent, privacy-preserving approach to generative clinical decision support, effectively leveraging heterogeneous healthcare data represented as EHR-based graphs.



These insights advance not only generative capabilities in clinical decision support but also prompt a broader discussion on optimal model utilization. Traditionally, inference tasks are prioritized, and quality improvements sought through enhanced model performance. An alternative strategy, however, parallels advancements in large language models, where generative capabilities serve as first-class citizens. In this case, generations can be interpreted as predictions based on the medical plausibility of an event, while inference is employed for safety checks, mitigating misleading suggestions. Such approaches present a novel perspective: focusing on generative task quality increases the usability of inferred results, as user attention is drawn to potential errors instead of the overall response.

13. Conclusion

Healthcare suffers from both scarce R&D budgets and an abundance of clinical knowledge. An autonomous system that enhances the role of AI in standard healthcare machine learning and automates dimension reduction can deliver substantial gains. Integrative models enable Federated Learning in not easily aggregable domains, overcoming privacy issues while being at least as efficient as classical centralized schemes. Despite high accuracy of currently deployed systems, they remain assistant-level, simply classifying and ranking possible outcomes. Unexpectedly advanced generative capabilities emerged during R&D: enrichment and completion of medical histories, ontology-based question generation and answering, and textual reinforcement of classification.

The system does not yet encompass all aspects of standard FHIR interoperability that demand a semantic or domain-agnostic model. In particular, the use of FHIR Extensions, a major FHIR interoperability use case, is not yet supported. The answer to the question "Are all patients in the dataset diabetic?" illustrates the need for generative capabilities in clinical AI, as clinical and legal consequences may be more important than accuracy in other task categories. Clinical pipeline creation and execution remain cumbersome and error-prone because they are partially outside the medical folksonomy and graph semiosis. User-centered explanation and trust evaluation of predictive solutions also remain open problems.

References

1. Choi, E., Xu, Z., Li, Y., Dusenberry, M. W., Flores, G., Xue, E., & Dai, A. M. (2020). Learning the graphical structure of electronic health records with graph convolutional



- network. In Proceedings of the 2020 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
2. Wang, X., Choi, E., Xu, Z., & Wang, J. (2020). Graph-based deep learning for medical diagnosis and analysis: Past, present and future. *IEEE Journal of Biomedical and Health Informatics*, 24(12), 3447–3460.
 3. Mangalampalli, B. M., Bandi, V. D. V. K., Kolla, S. K., & Kumar, M. V. K. (2025). Towards Self-Evolving Healthcare Intelligence: Integrating Advanced Learning Systems with Real-Time Clinical Data Pipelines. *Cultura: International Journal of Philosophy of Culture and Axiology*, 22(12s), 464-486.
 4. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2020). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), 4–24.
 5. Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI Open*, 1, 57–81.
 6. Mattaparthi, R. (2025). GenAI-Augmented Diagnostic Reasoning for Diesel Engine Fault Triage: A Large Language Model Framework for Technician Decision Support at Scale. *Journal of Material Sciences & Manufacturing Research*, 6(12), 1.
 7. Choi, E., Xiao, C., Stewart, W. F., & Sun, J. (2020). MiME: Multilevel medical embedding for 2D and 3D medical data. In Proceedings of the 2020 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
 8. Zhang, Y., Chen, R., Tang, J., Stewart, W. F., & Sun, J. (2020). LEAP: Learning to prescribe effective and safe treatment combinations for multimorbidity. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(01), 142–149.
 9. Kolla, S. H., & Mangala, N. (2025). DESIGNING AUTONOMOUS LLM AGENT FRAMEWORKS USING GEN AI PIPELINES TO ENHANCE CUSTOMER SERVICE MANAGEMENT AND KNOWLEDGE WORKFLOWS. *Lex Localis-Journal of Local Self-Government*, 23, 9719-9733.
 10. Choi, E., Bahadori, M. T., Song, L., Stewart, W. F., & Sun, J. (2021). GRAM: Graph-based attention model for healthcare representation learning. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
 11. Li, Y., Rao, S., Solares, J. R. A., Hassaine, A., Ramakrishnan, R., Canoy, D., Zhu, Y., Rahimi, K., & Salimi-Khorshidi, G. (2021). BEHRT: Transformer for electronic health records. *Scientific Reports*, 10, 7155.



12. Loganathan, R. (2025). AGENTIC AI FRAMEWORKS FOR AUTONOMOUS RISK DETECTION AND COMPLIANCE REMEDIATION IN ENTERPRISE DATA CENTER OPERATIONS. *Lex Localis-Journal of Local Self-Government*, 23 (S6), 9672–9697.
13. Shang, J., Xiao, C., Ma, T., Li, H., & Sun, J. (2021). Pre-training of graph augmented transformers for medication recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(01), 834–841.
14. Jin, W., Qu, M., Jin, X., & Ren, X. (2021). Rigorous self-supervised graph representation learning for predicting drug-drug interactions. *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*.
15. Mangalampalli, B. M., Kolla, S. K., Bandi, V. D. V. K., Yandamuri, U. S., & Rani, P. S. (2025). Designing Intelligent Healthcare Ecosystems through Adaptive Data Integration and Autonomous Learning Systems. *Vascular and Endovascular Review*, 8(20s), 330-347.
16. Sun, T., Wang, C., & others. (2021). Graph neural networks for healthcare applications: A survey. *IEEE Access*, 9, 113457–113477.
17. Choi, E., & Sun, J. (2022). Learning representations from electronic health records using graph neural networks. *Journal of Biomedical Informatics*, 129, 104076.
18. Yang, C., Xiao, Y., Zhang, Y., Sun, Y., & Han, J. (2022). Heterogeneous network representation learning: A unified framework with survey and empirical analysis. *IEEE Transactions on Knowledge and Data Engineering*, 34(5), 2271–2285.
19. Mahadevan, S. (2025). DRIVING SUSTAINABILITY IN AUTO & MANUFACTURING SECTORS: THE ROLE OF CLOUD-BASED SERVERLESS AUTOMATION FOR ERP SYSTEMS. *International Journal of Applied Mathematics*, 38(7s), 1813-1825.
20. Wu, L., Lin, H., Tan, C., Gao, Z., & Li, S. Z. (2022). Self-supervised learning on graphs: Deep insights and new directions. *IEEE Transactions on Knowledge and Data Engineering*, 35(10), 10139–10154.
21. Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., Scales, N., Tanwani, A. K., Cole-Lewis, H., Pfohl, S., Payne, P., Seneviratne, M., Gamble, P., Kelly, C., Schärli, N., Chowdhery, A., Mansfield, P., Demner-Fushman, D., Agrawal, M., ... Natarajan, V. (2023). Large language models encode clinical knowledge. *Nature*, 620, 172–180.
22. Singhal, K., Tu, T., Gottweis, J., Sayres, R., Wulczyn, E., Hou, L., Clark, J., Pfohl, S., Cole-Lewis, H., Neal, D., & others. (2023). Towards expert-level medical question answering with large language models. *arXiv*.
23. Kolla, S. K., & Reddy, V. A. R. (2024). Evaluating Cloud-Native vs. Hybrid Architectures for Health Benefit Administration Systems. *International Journal of Medical Toxicology and Legal Medicine*, 27(5), 1042-1053.

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME 04 ISSUE 02

RECEIVED: APRIL 07

REVISED: APRIL 26

ACCEPTED: MAY 04

PUBLISHED: MAY 22

ISSN: 3067-1140



24. Nori, H., King, N., McKinney, S. M., Carignan, D., & Horvitz, E. (2023). Capabilities of GPT-4 on medical challenge problems. arXiv.
25. Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. *Nature*, 616, 259–265.
26. Yang, X., Chen, A., PourNejatian, N., Shin, H. C., Smith, K. E., Parisien, C., Compas, C., Martin, C., Costa, A. B., Flores, M. G., Zhang, Y., & others. (2023). A large language model for electronic health records. *npj Digital Medicine*, 6, 149.
27. Mangalampalli, B. M., & Kolla, S. K. (2025). Large Language Models for Automated Healthcare Data Dictionary Generation and Maintenance. *Vascular and Endovascular Review*, 8(20s), 363-375.
28. Lehman, E., Jain, S., Paguet, C., & others. (2023). Do we still need clinical language models? *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*.
29. Yang, C., Liu, Z., Zhang, C., & others. (2023). GNN-based representation learning for clinical risk prediction from electronic health records. *Artificial Intelligence in Medicine*, 139, 102528.
30. Boll, H. O., Amirahmadi, A., Ghazani, M. M., Morais, W. O. de, Freitas, E. P. de, Soliman, A., Etmnani, F., Byttner, S., & Recamonde-Mendoza, M. (2024). Graph neural networks for clinical risk prediction based on electronic health records: A survey. *Journal of Biomedical Informatics*, 151, 104616.
31. Mangala, N., & Kolla, S. H. (2025). Real-Time Feature Engineering for Streaming AI Workloads Using PySpark and Azure Event Hubs. *International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management*, 1(6), 93-105.
32. Paul, S. G., Saha, A., Hasan, M. Z., Noori, S. R. H., & Moustafa, A. (2024). A systematic review of graph neural network in healthcare-based applications: Recent advances, trends, and future directions. *IEEE Access*, 12, 15145–15170.
33. Wang, D., & Zhang, S. (2024). Large language models in medical and healthcare fields: Applications, advances, and challenges. *Artificial Intelligence Review*, 57, 299.
34. Nerella, S., Bandyopadhyay, S., Zhang, J., Contreras, M., Siegel, S., Bumin, A., Silva, B., Sena, J., Shickel, B., Bihorac, A., Khezeli, K., & Rashidi, P. (2024). Transformers and large language models in healthcare: A review. *Artificial Intelligence in Medicine*, 154, 102900.
35. Wang, Y., Zhao, Y., Zhang, M., & others. (2024). Large language models in medicine: A systematic review. *Journal of Medical Internet Research*, 26, e22769.

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 04 ISSUE: 02

RECEIVED: APRIL 07

REVISED: APRIL 26

ACCEPTED: MAY 04

PUBLISHED: MAY 22

ISSN: 3067-1140



36. Yim, D., Khuntia, J., Parameswaran, V., & Meyers, A. (2024). Preliminary evidence of the use of generative AI in health care clinical services: Systematic narrative review. *JMIR Medical Informatics*, 12, e52073.
37. Mahadevan, S. (2024). Intelligent Serverless Process Automation for Scalable Cloud Operations and Dynamic Resource Optimization. *Journal of Computational Analysis and Applications (JoCAAA)*, 33(06), 4207-4221.
38. Omar, M., Nadkarni, G. N., Klang, E., & Glicksberg, B. S. (2024). Large language models in medicine: A review of current clinical trials across healthcare applications. *PLOS Digital Health*, 3(11), e0000662.
39. Xiao, H., Zhou, F., Liu, X., Liu, T., Li, Z., Liu, X., & Huang, X. (2025). A comprehensive survey of large language models and multimodal large language models in medicine. *Information Fusion*, 117, 102888.
40. Kather, J. N., Calderaro, J., & others. (2025). Graph artificial intelligence in medicine. *Annual Review of Biomedical Data Science*.