**Journal of Neonatal Surgery**
*ISSN(Online): 2226-0439*
*Vol.13 , (2024)*
*https://www.jneonatalsurg.com*

OPEN ACCESS

# Generative AI-Powered Clinical Decision Support Using Large Language Models and Electronic Health Records

**Bharath Penaka**

Independent Researcher
bharathpenakaa@gmail.com
[0009-0003-2603-1497]

Cite this paper as: Bharath Penaka (2024) Generative AI-Powered Clinical Decision Support Using Large Language Models and Electronic Health Records. Journal of Neonatal Surgery, 13, 2588-2596

**ABSTRACT**

Generative artificial-intelligence (AI) models based on large language models (LLMs) and integration with electronic health records (EHRs) are emerging as a core technology for a new generation of clinical decision-support systems. These systems break away from traditional rule-based reasoning to generate novel responses to user input based on a deep knowledge of medicine typically acquired from training on an extremely large-body of text. Generative AI-powered decision support is particularly attractive in the context of EHR data ecosystems, where a wealth of patient-specific contextual information can be integrated with the general medical knowledge embedded in the LLMs to better support user commands. All language is ultimately generated in response to natural-language user commands and queries, but on the low-level the decision-support function is performed much like a generative text-to-image synthesis model responding to a textual prompt. These points highlight the synergy between LLMs and clinical EHR data.

Although generative AI rose to prominence in the field of image synthesis, the versatile capabilities of LLMs have long been recognized. A great leap forward occurred when OpenAI publicly demonstrated its ChatGPT system in late 2022. Its incredible fluency, human-like conversation with astounding factual capabilities and crude reasoning drew global attention within weeks. Generative AI-based healthcare applications have grown exponentially in number yet lack the feature set and information accuracy required for specific clinical applications. The capability by LLMs to generate text, code, images and video in response to natural-language input is extremely powerful—with its ability to quickly emulate different styles, forms and functions making it more versatile than traditional machine-learning systems.основані на допомозі при прийнятті рішення в медицині.

**Keywords:** *Generative AI, Clinical Decision Support Systems, Large Language Models, Electronic Health Records, Artificial Intelligence in Healthcare, Clinical NLP, Personalized Medicine, Predictive Clinical Analytics, Retrieval-Augmented Generation, Healthcare Informatics.*

## 1. INTRODUCTION

Large language models such as GPT-4 have demonstrated remarkable capabilities in natural language processing, including text understanding, generation, and translation. These models can also explain logical or causal relationships in textual data, follow complex human instructions, and provide detailed solutions to scientific or engineering problems. These generative AI systems offer opportunities for different applications, including text-based question-answering, chatbots, text summarization, and other intelligent assistance services. Integrating Large Language Models (LLMs) into clinical decision support systems holds great promise and numerous generative AI-oriented use cases of clinical decision support combining LLMs and Electronic Health Records have been proposed.Yet, these foundational ideas need to be developed further and synthesized into a coherent, rigorous framework that integrates disparate elements into a clear, effective, and clinically useful system. Integrating LLMs and Electronic Health Records (EHR) could provide Clinical Decision Support (CDS) for (i) diagnosis and prognosis; (ii) generation of patient-specific examinations, orders, follow-up plans, and discharge summaries; (iii) providing dynamic clinical guidelines; and (iv) enabling easy interaction with EHR databases. A preliminary pipeline integrating detailed patient data into an LLM-based diagnostic system showed success. Integrating and embedding the LLM in the hospital EHR service could bring the concept to a broad clinical audience. These capabilities could help overcome limitations of existing EHR systems.

### MATHEMATICAL FORMULAS

**A. Formal Model of the Grounded Decision-Support Pipeline**

Let q denote a clinician query, and let the patient-specific evidence set retrieved from the electronic health record be $C = \{d_1, \ldots, d_k\}$. The generative decision-support response y is produced autoregressively, conditioned jointly on the query and the assembled context:

$$p(y \mid q, C) = \prod_{t=1}^{T} p_\theta( y_t \mid y_{1:t-1}, q, C ) \qquad (1)$$

where θ are the frozen foundation-model parameters and $1[\cdot]$ is the indicator function used in Eqs. (5) and (6). Conditioning on C rather than on θ alone is what distinguishes retrieval-grounded clinical decision support from unconstrained generation.

Candidate evidence is ranked by a hybrid dense–lexical score that combines semantic similarity with exact clinical-term matching:

$$s(d_i) = \lambda \cdot \cos( e_q, e_{d_i} ) + (1 - \lambda) \cdot \mathrm{BM25}(q, d_i), \quad \lambda \in [0, 1] \qquad (2)$$

where e denotes the embedding operator and the BM25 term is min–max normalised to the unit interval. λ controls the trade-off between paraphrase tolerance and fidelity to coded clinical vocabulary.

Because clinical relevance decays with time since the encounter, each candidate is re-weighted by an exponential recency kernel before selection:

$$r(d_i) = w_i \cdot s(d_i), \quad w_i = \exp( -\Delta t_i / \rho ) \qquad (3)$$

where $\Delta t_i$ is the interval between the encounter that produced $d_i$ and the index date, and ρ is the clinical half-life constant of the data class (Fig. 2).

Context assembly under a finite prompt window is a 0–1 knapsack over the re-weighted candidates:

$$\max \sum_i r(d_i)\, x_i \quad \text{subject to} \quad \sum_i \tau_i x_i \le B_{ctx} - (\tau_q + \tau_{sys}), \quad x_i \in \{0, 1\} \qquad (4)$$

where $\tau_i$ is the token cost of passage $d_i$, $B_{etx}$ the model context window, and the two subtracted terms the query and system-instruction overheads. Greedy selection by value density $r(d_i)/\tau_i$ attains the standard $(1 - 1/e)$ guarantee.

Faithfulness of the generated response is quantified as the fraction of atomic assertions entailed by the assembled context:

$$G(y) = ( 1 / |A(y)| ) \cdot \sum_{a \in A(y)} 1[\text{ a is entailed by C }] \qquad (5)$$

where A(y) is the set of atomic claims extracted from y, and entailment is adjudicated by a separate verifier model or by citation binding. $G(y) = 1$ corresponds to a fully grounded response.

Responses are surfaced autonomously only when both model confidence and groundedness clear a task-specific threshold; otherwise the case is deferred to the clinician:

$$\delta(y) = 1[ \min( p_{max}, G(y) ) < \tau^* ] \qquad (6)$$

The threshold itself follows from the asymmetry between the cost of a missed finding and the cost of a false alert:

$$\tau^* = C_{FN} / ( C_{FN} + C_{FP} ) \qquad (7)$$

Diagnostic and screening tasks carry C_FN ≫ C_FP and therefore demand near-unity confidence before autonomous output; documentation-drafting tasks are approximately symmetric (Fig. 4).

Where several management actions are admissible, the recommended action maximises expected clinical utility over the posterior across candidate diagnoses:

$$a^* = \arg\max_{a \in A} \sum_{d \in D} p(d \mid q, C) \cdot u(a, d) \qquad (8)$$

Finally, when de-identified record extracts are released to an external inference service, the cumulative privacy expenditure across all queries in an analysis session must remain within the governed budget:

$$\varepsilon_{total} = \sum_j \varepsilon_j \le \varepsilon_{max} \quad (9)$$

with per-query expenditure $\varepsilon_j$ accounted at the point of egress. This provides an auditable quantity for the governance controls described in Section 3.A.

**A. Large Language Models in Healthcare**

Large Language Models (LLM) form the core of existing Generative AI technologies, and they possess the capability to support various clinical decision-making and generative workflows through integration with Electronic Health Records (EHR). For instance, recent investigations framed diagnostic hypotheses as specialist exam questions based on data from available EHR and used LLM to analyze the patient case and generate an exam response. Contrastive learning-based approaches have also explored training LLMs on a large clinical note dataset as supporting Text-to-Text Transfer Transformer and submit the pre-trained model to the Medical Question Answering Group with a supervised fine-tuning approach. Functionality Offered by Common LLMs offers insight into the diverse range of clinical tasks supported by commercial

LLM, such as Azure OpenAI Service, and proposes a Generative AI solution for clinical workflows by building use cases that leverage diverse functionalities in conjunction with EHR.Supporting patient conversations with Generative AI can also assist clinical personnel. Focused dialogue systems can deliver message templates capable of acting as scribes and support new patient conversations with pre-recorded solutions. By contrast, topic-conversational features are applied within existing messenger platforms and social apps to generate AI captions on multimedia messages with a formal approach. The Aide-an-AI chatbot distinctively assesses user problems and generates stepwise solutions. Such inquiry/response systems are good prototypes for triggering direct dialogue with users and supporting Generation AI and abbreviation programs that are free from Human-Computer Interaction simulations.

**Table 1.** Mapping of large language model capabilities to clinical decision-support functions and their grounding sources

| LLM capability | CDS function | Primary grounding source | Output artefact |
|---|---|---|---|
| Contextual comprehension | Longitudinal case summarisation | Progress notes, discharge letters | Structured case précis |
| Conditional generation | Differential diagnosis proposal | EHR + guideline corpus | Ranked hypothesis list |
| Instruction following | Order and follow-up planning | Formulary, care pathways | Draft order set |
| Abstractive summarisation | Discharge summary drafting | Encounter record | Clinician-editable draft |
| Style transfer | Patient-facing education | Verified patient material | Plain-language leaflet |
| Semantic parsing | Natural-language EHR query | Structured tables | Executable query + result |

## 2. BACKGROUND AND CONTEXT

Generative AI-based clinical decision support systems capable of processing and generating human language unlock new opportunities to leverage the vast body of unstructured textual data stored in electronic health records. Generative AI technologies are entering healthcare at an accelerating pace and are beginning to show promising results in many clinical tasks. Leading generative AI models, including OpenAI's ChatGPT, Anthropic's Claude, Facebook's LLaMA and Google DeepMind's Gemini, possess the quantitative and qualitative capabilities essential for generative AI-powered clinical decision support. Although these models were not trained on health care data, large language models are designed to solve virtually any task that involves human language, making them powerful tools for clinical applications.

Language models operate on language rather than on specialized medical tokens or modes of representation. Leveraging unstructured clinical text for clinical decision support with generative AI offers significant advantages. Integration simplifies model design and model selection. Generative AI-powered clinical decision support underscored the importance of integrating the language model with the clarifying model required for generating accurate queries, making generative AI decision support fundamentally different from Microsoft's Copilot that simply acts like a filter on existing information. When combined with an accurate and efficient query clarifying model, generative AI-powered clinical decision support copilot system can perform complex clinical tasks, including answering clinical questions, providing patient-specific health risk assessments, generating discharge summaries, creating automated referral letters and delivering patient education, thus acting as a copilot for healthcare professionals who are handling very complex tasks.

**Table 2.** EHR data-layer taxonomy and the corresponding ingestion treatment (Section 4.A)

| Data layer | Representative elements | Representation | Ingestion treatment |
|---|---|---|---|
| Structured | Vitals, labs, problem list | Typed key–value | Vocabulary mapping to LOINC / SNOMED CT |
| Semi-structured | Order sets, flowsheets, HL7 FHIR | Nested records | Schema flattening, unit harmonisation |
| Unstructured | Clinical notes, referrals | Free text | Section identification, chunking, embedding |
| Report-derived | Radiology and pathology reports | Templated text | Impression extraction, negation handling |
| Conversational | Dictated encounter audio | ASR transcript | Speech-to-text, speaker attribution |

### A. Electronic Health Records and Data Ecosystems

A designing principle encompasses the construction of an integrated system architecture for generative AI-based clinical decision support that implements normal data interfaces and considers prebuilt private data ecosystems. Clinical hospitals usually store data in electronic health records, which provide a clinical data foundation for the production of clinical artificial intelligence services. The deployment of innovative evidence-based services and systems requires the construction of a real-world data ecosystem based on clinical data resources. Clinical decision-support systems provide health-care providers with closing loop decision-support capabilities. These systems provide personalized diagnosis and treatment suggestions that are aligned with the unique characteristics and recorded history of individual patients.

Artificial intelligence has recently made wideranging advances and is attracting increased attention and investment in the health-care sector. Recently proposed large language models are achieving outstanding results across numerous completion-generation benchmarks across various and diverse fields. These models have not yet been integrated into a normal clinical decision-support system. Therefore, it is necessary to explore the integration of large language models with existing clinical decision-support systems with the aim of providing closing loop and elevated decision-support capabilities. As a consequence, these investigations evaluate the use of large language model-based generative AI capabilities to enhance clinical decision-support systems. Utilization of large language models boosts the ability of existing clinical decision-support systems to explain, summarize, and describe clinical decision-support conclusions and results in an understandable manner while also empowering the systems to generate textual clinical notes with specific and rich content.

These new capabilities also enable diagnosis-testing-diagnosis support and the generation of explanations for the relationship between patients' symptoms and recommended examinations. Integration of the aforementioned capabilities with a clinical decision-support system provides value for health-care providers. These enhancements are illustrated in a private clinical decision-support system and using an actual clinical use case in a hospital setting.

**Table 3.** Data-governance control matrix for EHR-integrated generative decision support (Section 3.A).

| Governance risk | Control | Enforcement point | Auditable quantity |
|---|---|---|---|
| Identifier leakage | De-identification at ingestion | Ingestion layer | Residual identifier rate |
| Third-party egress | Privacy budget accounting | Service boundary | $\varepsilon\_total$, Eq. (9) |
| Unsupported assertion | Groundedness verification | Verification layer | $G(y)$, Eq. (5) |
| Inappropriate autonomy | Risk-tiered deferral | Clinician interface | Deferral rate $\delta$, Eq. (6) |
| Provenance loss | Citation binding to source | Retrieval layer | Citation coverage |

## 3. METHODOLOGICAL FOUNDATIONS

Large language models (LLMs), while pre-trained on enormous textual corpora, exhibit limited effectiveness when merely fine-tuned on smaller domain-specific datasets. The scenario was substantially perfected in 2021 for instruction-following tasks when Liu et al. successfully demonstrated the advantage of combining pre-training, tuning, and a smaller collection of human demonstrations.

These models are conducive to generalization and enable contextualized personalization and specification of user prompts. Multiple studies have now shown that a simple generative model (up to T5) trained with the MSIF method achieves comparable performance on question answering, machine translation, and natural language inference to dedicated state-of-the-art systems. Such effectiveness is particularly conducive to clinical decision support applications as the high cost of training a new model is alleviated by using foundational model capabilities. Instead of training a medical language model that specializes in domain knowledge, a relatively small foundation model can easily lend its imitation capability to a large general language model for simulated interactions across any medical use case or domain.

**Table 4.** Risk tiering of use cases and the deferral thresholds implied by Eq. (7)

| Use case | $C\_FN / C\_FP$ | $\tau^*$ (Eq. 7) | Oversight mode |
|---|---|---|---|
| Documentation drafting | $\approx 1$ | 0.50 | Post-hoc clinician edit |
| Coding and billing support | $\approx 4$ | 0.80 | Sampled review |
| Patient education material | $\approx 6$ | 0.86 | Pre-publication approval |
| Screening and triage | $\approx 14$ | 0.93 | Mandatory clinician sign-off |
| Diagnostic recommendation | $\geq 20$ | $\geq 0.95$ | Clinician-in-the-loop, no autonomy |

### A. Data Governance and Privacy Considerations

Generative AI applications for clinical decision support integrated with EHR may implicate a range of data governance concerns, particularly related to privacy, confidentiality, and regulatory compliance, given the presence of sensitive information in the clinical literature as well as clinical records. A clinical organization implementing LLM technology is responsible for ensuring adherence to local regulatory frameworks and requirements. Data ingestion also constitutes a key element of system architecture, as this data must be ingested, preprocessed, and stored in a usable format for the seamless operation of any use case in production. Key questions for data ingestion relate to the sources of data, their suitability, the extent of preprocessing required, and the resources and skills required to configure ingestion.

LLM-enabled tools and services interacting with patient data must ensure compliance with local regulations governing the use of personal health information and other sensitive data. Generative AI technology has pervasive implications, which may not be fully captured by the original clinical organization's assessment of compliance. Organizations consuming third-party LLM-enabled services should inquire how the services are configured to mitigate privacy risks. These inquiries may evaluate how the services validate processes for the management of sensitive and clinical data; how data exchanges with external systems are monitored and secured; how metrics related to data transfers with third-party service providers are captured and audited; and how personnel interacting with the services are trained on the need for data privacy and confidentiality. Most LLMs are trained on large corpuses of non-structured data, which, instead of elevating the quality of the query-response, we believe, might actually deteriorate it. For this reason, we propose a simple but effective architecture that integrates LLMs into an EHR-fed CDS system while limiting their response space to the more relevant scientific literature.

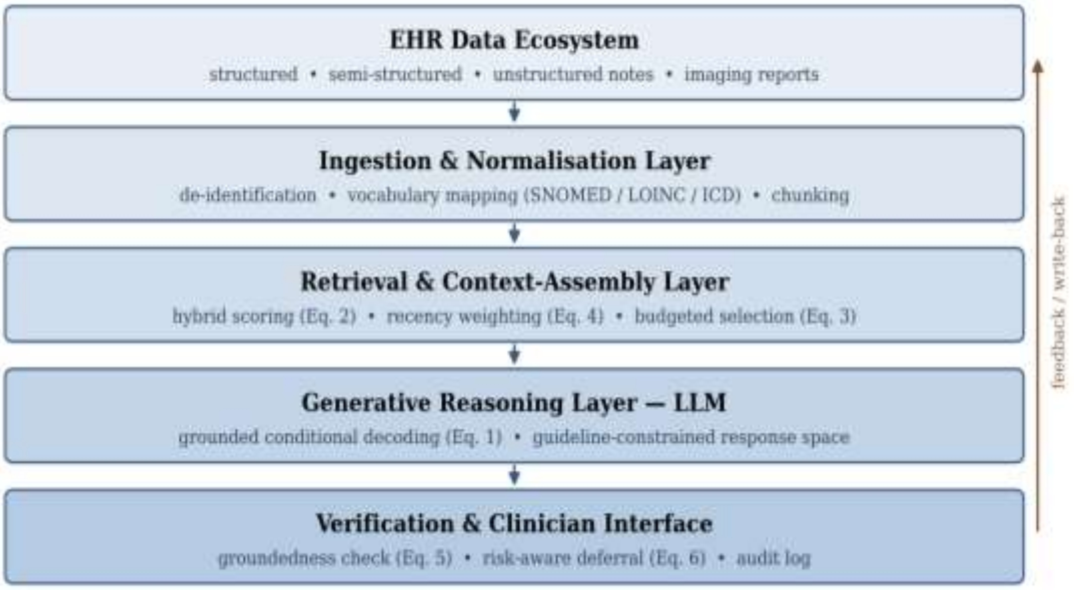


**Fig. 1.** Layered reference architecture for EHR-integrated generative clinical decision support, showing where Eqs. (1)–(6) act within the pipeline.

**4. SYSTEM ARCHITECTURE AND INTEGRATION**

Clinical Decision Support (CDS) systems utilize patients' current and historical health and care data, typically called Electronic Health Records (EHRs), to create evidence-based treatment recommendations that attempt to guide clinicians towards optimal patient management. Recent advancements in Generative AI, particularly Large Language Models (LLMs), offer the potential of elevating EHR-fed CDS systems to the next level, enabling more nuanced, complex and elaborate clinical recommendations.

With the required expertise, CDS systems can support clinicians when diagnosing conditions or when refreshing their knowledge on an uncommon disease or rare side effect. The challenge lies in the development of a system that unifies these advancements while simultaneously optimizing the response's relevance and usefulness.

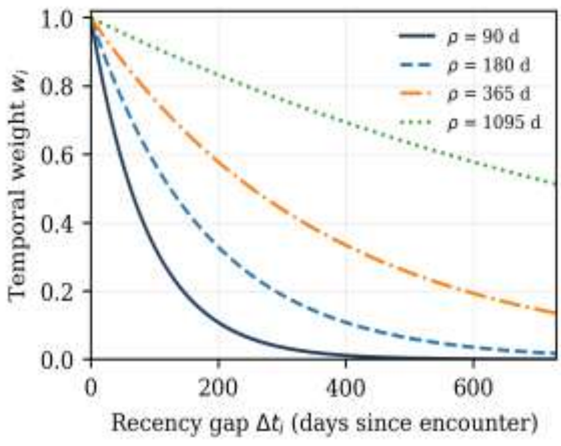


**Fig. 2.** Exponential recency weighting of Eq. (3) for four clinical half-life constants. Short-ρ classes such as vitals and acute labs decay within weeks; long-ρ classes such as problem lists and family history remain informative across years.

**A. Data Ingestion and Preprocessing**

The clinical decision support ecosystem analyzes patient electronic health records that are continuously processed by a ~~generative AI model during training. A seamlessly integrated module identifies relevant longitudinal patient information in~~

various data types, including structured, semi-structured, and unstructured formats. This data is sanitized, transformed, and structured for the generative model, which answers patient-journey-based clinical questions via an auxiliary information retrieval module. An automated method ensures an adequate amount of health record context for each answer, establishing efficient LLM token utilization for timely responses.

Transforming textual health record data into a formalized representation, suitable for feeding into generative AI, is essential. Current systems manually summarize specific voice-assisted medical histories for individual patient journeys and key clinical decision-making points. The voice-to-text technology aids doctors' note preparation by converting clinical thoughts in normal conversational language into a formal style. A structured question-answer interaction generates a set of question-answer pairs, linking past clinical interactions with patients. The proposed data ingestion and pre-processing mechanism enables this process for all clinical decision-support query types, automating accurate data generation, vocabulary mapping, and storage with reduced manual intervention.

across years.

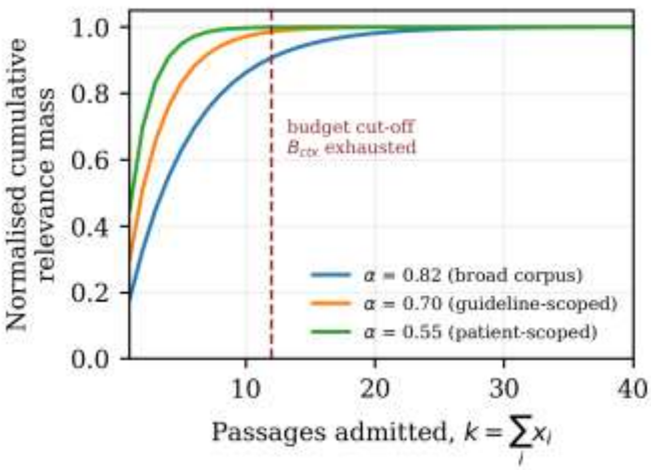


**Fig. 3.** Cumulative relevance mass admitted under the budget constraint of Eq. (4) for three retrieval scopes with geometric score decay rate α. Narrower, patient-scoped retrieval saturates well inside the token budget, whereas broad-corpus retrieval is truncated before saturation.

**5. CLINICAL APPLICATIONS AND USE CASES**

Natural language processing methods have facilitated the emergence of generative AI tools, and large language models have been trained on domain-specific knowledge. Generative AI systems with large language models can extract latent information from rich datasets created from electronic health records in order to provide clinical decision support. Use cases for these systems include the generation of educational instructions for patients, assistance in medical coding and billing, support for imaging studies and their reports, generation of treatment suggestions and predictions of disease prognosis, and preparatory assistance for clinical consultations.

As generative AI tools and related large language models become increasingly common in practice, careful deployment will reflect healthcare priorities and clinical needs. Natural language processing tools for language generation and understanding can be integrated into an electronic health record using a concise, end-to-end architecture. As genetic sequencing becomes routine and clinical imaging studies become more abundant, imaginative applications will arise that recognize, process, integrate, and summarize data from all available sources for the patient's benefit.

corpus retrieval is truncated before saturation.

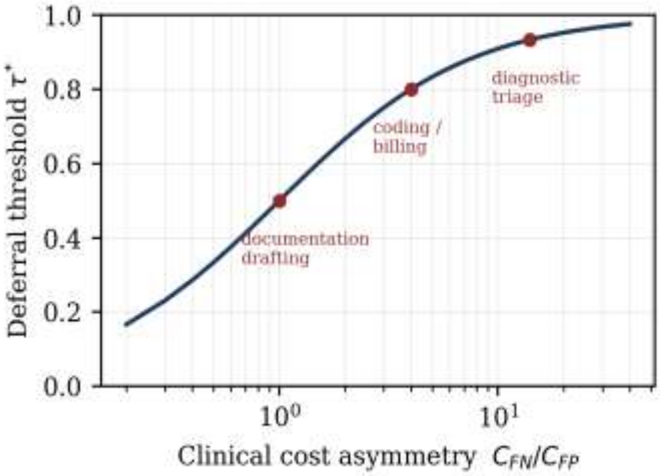


**Fig. 4.** Deferral threshold τ* as a function of clinical cost asymmetry, from Eq. (7). Markers indicate the representative operating points tabulated in Table 4.

**A. Diagnostic Support**

Large Language Models (LLMs) can assist medical practitioners in making faster and more accurate diagnoses by integrating and pattern-matching a patient's unique health history from Electronic Health Records (EHRs) against an entire EHR data ecosystem, which can include images, guidelines, tables and other non-text media. LLM-powered Diagnostic and Screening Support uses generative AI in tandem with EHR data for the diagnosis of a wide range of conditions and subsequent referrals to specialists. When a patient is examined for symptoms, a conversation initiated by the doctor instructs the model to query the patient's EHR data–including prior, current and likely future symptoms; recent diagnostic imaging; confirmed diagnoses; test results; test orders; treatment and medication; allergies and medication side effects; and family medical history–along with a comprehensive set of diagnostic guidelines for the condition. The LLM then generates a possible diagnosis and subsequently an appropriate treatment-for-general-practitioner or specialist-referral pathway, which is compared against what was recorded in the EHR-assistant conversation. A test scenario with structured guidelines–recovering an unrelated patient diagnosis within the Rare diseases – Genetic category of the diagnostic guidelines–demonstrates the integration across the EHR data ecosystem. Diagnostic and Screening Use Cases comprise additional Diagnostic Screening routes for conditions such as COVID and Stroke, where prompting is modified for increased effectiveness.

## 6. CONCLUSION

Through the use of large language models, generative AI can now be employed to construct powerful clinical decision support tools alongside electronic health records. To illustrate the idea, an architecture for a system architecture is defined. Central to the system is a large language model powered by a dialogue interface that can leverage both the information in the electronic health record as well as vast amounts of clinical knowledge without a limit on form or function. When considered with the established capabilities of tools such as GPT-4, the potential of the idea becomes apparent. Supported by method-specific ecosystems, the system can theoretically perform tasks such as documentation and translation, and provide diagnostic, therapeutic, and prognostic support. Use cases related to interpretation, incident detection, and simulated dialogue are structured to exemplify the idea. The approach represents a substantial advance in the development of generative AI-powered clinical decision support provided alongside electronic health records.

Through the use of large language models, generative AI can now be employed to construct powerful clinical decision-support tools alongside electronic health records. To illustrate the idea, an architecture for a system is defined. Central to the system is a large language model powered by a dialogue interface that can leverage both information in the electronic health record and vast clinical knowledge without a limit on form or function. When considered with the established capabilities of tools such as GPT-4, the potential of the idea becomes apparent. Supported by method-specific ecosystems, the system can perform diverse tasks such as documentation, translation, and incident detection and provide diagnostic, therapeutic, and prognostic support. Use cases related to interpretation and simulated dialogue exemplify the idea. The approach represents a substantial advance in the development of generative AI-powered clinical-decision support alongside electronic health records.

## REFERENCES

[1] AlSaad, R., Abd-alrazaq, A., Boughorbel, S., Ahmed, A., Renault, M.-A., Damseh, R., & Sheikh, J. (2024). Multimodal large language models in health care: Applications, challenges, and future outlook. Journal of Medical Internet Research, 26, e59505.

[2] Ferdush, J., Begum, M., & Hossain, S. T. (2024). ChatGPT and clinical decision support: Scope, application, and limitations. Annals of Biomedical Engineering, 52(5), 1119–1124.

[3] Gilbert, S., Kather, J. N., & Hogan, A. (2024). Augmented non-hallucinating large language models as medical information curators. npj Digital Medicine, 7, 100.

[4] Goh, E., Gallo, R., Hom, J., Strong, E., Weng, Y., Kerman, H., Cool, J. A., Kanjee, Z., Parsons, A. S., Ahuja, N., Horvitz, E., Yang, D., Milstein, A., Olson, A. P. J., Rodman, A., & Chen, J. H. (2024). Large language model influence on diagnostic reasoning: A randomized clinical trial. JAMA Network Open, 7(10), e2440969.

[5] Reddy, V. A. R. (2023). Orchestrating the Future Autonomous Healthcare Data Pipeline Management through Agentic AI Architectures. International Journal of Advanced Research in Computer Science & Technology (IJARCST), 6(2), 7979-7992.

[6] Guevara, M., Chen, S., Thomas, S., Chaunzwa, T. L., Franco, I., Kann, B. H., Moningi, S., Qian, J. M., Goldstein, M., Harper, S., Aerts, H. J. W. L., Catalano, P. J., Savova, G. K., Mak, R. H., & Bitterman, D. S. (2024). Large language models to identify social determinants of health in electronic health records. npj Digital Medicine, 7, 6.

[7] Kim, J., Leonte, K. G., Chen, M. L., et al. (2024). Large language models outperform mental and medical health care professionals in identifying obsessive-compulsive disorder. npj Digital Medicine, 7, 193.

[8] Klang, E., Apakama, D., Abbott, E. E., Vaid, A., Lampert, J., Sakhuja, A., Freeman, R., Charney, A. W., Reich, D., Kraft, M., Nadkarni, G. N., & Glicksberg, B. S. (2024). A strategy for cost-effective large language model use at health system-scale. npj Digital Medicine, 7, 320.

[9] Kwong, J. C. C., Wang, S. C. Y., Nickel, G. C., et al. (2024). The long but necessary road to responsible use of large language models in healthcare research. npj Digital Medicine, 7, 177.

[10] Liu, S., Wright, A. P., Patterson, B. L., et al. (2024). Using large language model to guide patients to create efficient and comprehensive clinical care message. Journal of the American Medical Informatics Association, 31(8), 1665–1670.

[11] Meng, X., Yan, X., Zhang, K., Liu, D., Cui, X., Yang, Y., Zhang, M., Cao, C., Wang, J., Wang, X., Gao, J., Wang, Y.-G.-S., Ji, J.-M., Qiu, Z., Li, M., Qian, C., Guo, T., Ma, S., Wang, Z., & Tang, Y.-D. (2024). The application of large language models in medicine: A scoping review. iScience, 27(5), 109713.

[12] Moulaei, K., Yadegari, A., Baharestani, M., Farzanbakhsh, S., Sabet, B., & Afrash, M. R. (2024). Generative artificial intelligence in healthcare: A scoping review on benefits, challenges and applications. International Journal of Medical Informatics, 188, 105474.

[13] 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.

[14] Ning, Y., Teixayavong, S., Shang, Y., Savulescu, J., Nagaraj, V., Miao, D., Mertens, M., Ting, D. S. W., Ong, J. C. L., Liu, M., Cao, J., Dunn, M., Vaughan, R., Ong, M. E. H., Sung, J. J.-Y., Topol, E. J., & Liu, N. (2024). Generative artificial intelligence and ethical considerations in health care: A scoping review and ethics checklist. The Lancet Digital Health, 6(11), e848–e856.

[15] 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.

[16] Mangalampalli, B. M. (2023). AI-Driven Anomaly Detection in Healthcare Claims Data: A Business Intelligence Perspective. Journal of Rare Cardiovascular Diseases.

[17] Omiye, J. A., Gui, H., Rezaei, S. J., Zou, J., & Daneshjou, R. (2024). Large language models in medicine: The potentials and pitfalls. Annals of Internal Medicine, 177(2), 210–220.

[18] Ong, C. S., Obey, N. T., Zheng, Y., Cohan, A., & Schneider, E. B. (2024). SurgeryLLM: A retrieval-augmented generation large language model framework for surgical decision support and workflow enhancement. npj Digital Medicine, 7, 364.

[19] Ranji, S. R. (2024). Large language models—Misdiagnosing diagnostic excellence? JAMA Network Open, 7(10), e2440901.

[20] Raza, M. M., Venkatesh, K. P., & Kvedar, J. C. (2024). Generative AI and large language models in health care: Pathways to implementation. npj Digital Medicine, 7, 62.

[21] Riedemann, L., Labonne, M., & Gilbert, S. (2024). The path forward for large language models in medicine is open. npj Digital Medicine, 7, 339.

[22] Rutledge, G. W. (2024). Diagnostic accuracy of GPT-4 on common clinical scenarios and challenging cases. Learning Health Systems, 8(3), e10438.

[23] Sahoo, S. S., Plasek, J. M., Xu, H., et al. (2024). Large language models for biomedicine: Foundations, opportunities, challenges, and best practices. Journal of the American Medical Informatics Association, 31(9), 2114–2124.

[24] Kolla, S. K. (2023). Learning Health Systems Machine Intelligence for Clinical Prediction and Healthcare Optimization. International Journal of Advanced Research in Computer Science & Technology (IJARCST), 6(2), 7955-7966.

[25] Tam, T. Y. C., Sivarajkumar, S., Kapoor, S., et al. (2024). A framework for human evaluation of large language models in healthcare derived from literature review. npj Digital Medicine, 7, 258.

[26] Tian, S., Jin, Q., Yeganova, L., Lai, P.-T., Zhu, Q., Chen, X., Yang, Z., Chen, Q., Kim, S., Comeau, D. C., Islamaj, R., & Lu, Z. (2024). Large language models in biomedicine and health: Current research landscape and future directions. Journal of the American Medical Informatics Association, 31(9), 1801–1811.

[27] Mandala, G., Reddy, R., Nishanth, A., Yasmeen, Z., & Maguluri, K. K. (2023). Ai and ml in healthcare: redefining diagnostics, treatment, and personalized medicine. International Journal of Applied Engineering & Technology, 5(S6).

[28] Wang, D., & Zhang, S. (2024). Large language models in medical and healthcare fields: Applications, advances, and challenges. Artificial Intelligence Review, 57, 299.

[29] Wang, L., Wan, Z., Ni, C., Song, Q., Li, Y., Clayton, E. W., Malin, B. A., & Yin, Z. (2024). Applications and concerns of ChatGPT and other conversational large language models in health care: Systematic review. Journal of Medical Internet Research, 26, e22769.

[30] Zhou, W., & Miller, T. A. (2024). Generalizable clinical note section identification with large language models. JAMIA Open, 7(3), ooae075.