Generative Artificial Intelligence Applications in Electronic Medical Records and Health Information Management-An Updated Review

Lafy Yahya Mohammed Aljabri[1], Moayad Mofareh Bariqi[1], Mohammed Jamah Tawhari[1], Abdullah Ali Jaber Hzazi[1], Mohammed Ghalib Bhkali[1], Ali Dhaifallah Madrabaa[1], Abduh Ali Harash[1], Mohsen Hamad Hakami[1], Hassan Abdullah Jawhari[1], Mohammed Yahya Qadri[1], Ayyashi Dahnan Jubran[1]

Abstract:
Background: Electronic Medical Records (EMRs) are central to modern healthcare delivery but generate large volumes of complex, unstructured clinical data that are difficult to process and utilize effectively. Generative Artificial Intelligence (GenAI), particularly large language models, has emerged as a promising technology for enhancing EMR functionality across multiple clinical and administrative domains. **Aim**: This review aims to explore and categorize the applications of GenAI within EMR systems, assess its performance across different healthcare tasks, and identify its benefits, limitations, and implications for clinical practice. **Methods**: A scoping review approach was used to analyze studies examining GenAI applications in EMRs. A total of 55 studies were reviewed and categorized into key thematic areas including data manipulation, patient communication, clinical decision-making, clinical prediction, summarization, and other emerging applications. Findings were synthesized narratively to identify patterns of use, performance outcomes, and reported limitations. **Results**: GenAI demonstrated strong performance in data manipulation, clinical summarization, and patient communication, improving efficiency and reducing clinician workload. In clinical prediction and certain structured tasks, GenAI showed performance comparable to or exceeding traditional machine learning models. However, clinical decision-making applications revealed limitations, including inaccurate recommendations, hallucinations, and unsafe outputs in high-risk settings. Concerns regarding bias, interpretability, data privacy, and legal accountability were consistently reported across studies. **Conclusion**: GenAI shows significant potential to enhance EMR systems by improving efficiency, accessibility, and clinical support functions. However, its current limitations restrict its use as an autonomous clinical tool. Safe integration requires robust governance, continuous validation, and human oversight.

[1] Ministry of Health, National Guard Health Affairs, Saudi Arabia

Keywords: Generative Artificial Intelligence, Electronic Medical Records, Large Language Models, Clinical Decision Support, Health Informatics, Patient Communication, Medical Data Analytics

Introduction

Electronic Medical Records (EMRs) have become a cornerstone of contemporary healthcare systems, serving as comprehensive digital platforms that facilitate the documentation, storage, retrieval, and exchange of patient health information. These systems support a broad spectrum of clinical and administrative functions, including patient record management, clinical decision-making, communication among healthcare professionals, and the continuity of care across healthcare settings. The widespread adoption of EMRs has transformed healthcare delivery by enabling clinicians to access extensive patient information in real time, thereby improving the efficiency, accuracy, and coordination of clinical services. As healthcare organizations continue to generate and accumulate vast amounts of patient-related data, EMRs have evolved into critical repositories containing detailed clinical narratives, diagnostic findings, treatment plans, laboratory results, and other forms of healthcare documentation. Despite the substantial benefits associated with the digitization of healthcare information, the rapid expansion of EMR data has introduced significant challenges. A considerable proportion of information contained within EMRs exists in unstructured textual formats, including physician notes, discharge summaries, consultation reports, and other narrative documentation. Although these records contain valuable clinical insights, their volume and complexity often hinder efficient extraction, interpretation, and utilization. Healthcare professionals frequently face difficulties in reviewing extensive documentation within limited timeframes, increasing the risk of information overload and reducing the efficiency of clinical workflows. Consequently, the healthcare sector has sought innovative technological solutions capable of facilitating the management, interpretation, and application of these increasingly complex datasets [1].

Recent developments in generative artificial intelligence (GenAI) have created new opportunities for addressing many of the challenges associated with the utilization of EMR data. Advances in natural language processing technologies, particularly the emergence of large language models (LLMs), have significantly enhanced the ability of computational systems to understand, generate, summarize, and interpret human language. Public awareness and professional interest in these technologies expanded substantially following the release of ChatGPT by OpenAI in November 2022 [1]. Since then, healthcare researchers, clinicians, administrators, and health information specialists have increasingly explored the potential applications of LLMs in medical environments. These models demonstrate remarkable capabilities in processing large volumes of textual information and generating coherent, contextually relevant outputs that closely resemble human-generated communication. Such capabilities position them as promising tools for improving the accessibility and usability of information embedded within EMRs. Although artificial intelligence has been investigated within healthcare for several decades, earlier AI methodologies were fundamentally different from contemporary generative models. Traditional AI systems typically relied on highly specialized algorithms developed for narrowly defined tasks, such as disease prediction, image classification, or risk stratification. The development of these systems required extensive quantities of curated, high-quality training data and often involved significant investments of time, expertise, and financial resources. Furthermore, separate models frequently had to be designed, validated, and implemented for each individual clinical application. These limitations constrained the scalability and adaptability of traditional AI approaches within complex healthcare environments.

In contrast, LLMs offer a more flexible and cost-efficient paradigm for interacting with healthcare information. Rather than being restricted to a single predefined task, these models possess the ability to perform multiple functions, including information retrieval, text generation, clinical summarization, documentation support, and question answering. Their capacity to rapidly adapt to diverse healthcare scenarios without extensive task-specific retraining represents a significant advancement in the field of medical informatics. Moreover, ongoing improvements in model performance have further strengthened their potential to enhance clinical productivity, reduce administrative burdens, and support healthcare professionals in managing growing volumes of patient data. Nevertheless, the rapid pace of innovation in generative AI has exceeded the ability of many healthcare institutions and stakeholders to comprehensively assess its implications for clinical practice. While enthusiasm regarding the transformative potential of these technologies continues to grow, important questions remain concerning their effectiveness, reliability, safety, ethical implications, and integration into existing healthcare workflows. The accelerated evolution of LLMs has generated substantial scholarly interest; however, the evidence base remains relatively immature, and many proposed applications have yet to undergo rigorous evaluation within real-world clinical settings. Consequently, healthcare professionals, policymakers, technology developers, and researchers face challenges in determining the most appropriate and effective ways to implement these technologies [1].

Given the emerging nature of this field, the available literature remains fragmented, and comprehensive syntheses of current evidence are limited. Much of the existing scholarship focuses on individual applications or early-stage evaluations, making it difficult to establish a clear understanding of the broader role of generative AI within EMR systems. Therefore, there is a critical need to consolidate current knowledge, identify prevailing trends, and evaluate the range of potential use cases associated with these technologies. Addressing this need can provide healthcare stakeholders with a clearer perspective on both the opportunities and challenges presented by generative AI.

Accordingly, this scoping review was undertaken in accordance with the PRISMA-ScR guidelines to examine the existing body of evidence, classify emerging applications, identify areas demonstrating significant promise, and provide an informed overview of the evolving relationship between generative artificial intelligence and electronic medical records. The findings are intended to support clinicians, healthcare administrators, health information specialists, EMR vendors, and the broader healthcare community in understanding the current landscape and future potential of these transformative technologies [1].

GenAI for Data Manipulation

The growing adoption of Electronic Medical Records (EMRs) has resulted in the accumulation of vast quantities of healthcare data, creating unprecedented opportunities for clinical research, healthcare management, and evidence-based decision-making. However, the value of these data is often constrained by their complexity, heterogeneity, and predominantly unstructured nature. Clinical information is frequently recorded in diverse formats, including narrative physician notes, discharge summaries, pathology reports, radiology interpretations, and consultation records. The variability in documentation practices across healthcare providers and institutions further complicates the process of extracting meaningful information from EMRs. Consequently, healthcare organizations face significant challenges in transforming raw clinical data into actionable knowledge. Generative Artificial Intelligence (GenAI), particularly large language models (LLMs), has emerged as a powerful solution for addressing these challenges through advanced data manipulation capabilities. These technologies possess the ability to process large volumes of textual information, identify relevant clinical concepts, and convert unstructured data into structured formats suitable for analysis. By automating tasks that traditionally required substantial manual effort, GenAI enhances the efficiency and scalability of healthcare data management. Within the reviewed literature, data manipulation represented one of the most prominent applications of GenAI, with 24 out of 55 studies focusing on this area. This substantial proportion reflects the growing recognition of GenAI as an effective tool for unlocking the value embedded within EMR systems [2][3]. Several studies demonstrated the effectiveness of GenAI in extracting clinically relevant information for disease phenotyping, a process that involves identifying patient characteristics and disease patterns from healthcare records [4]. Accurate phenotyping is essential for clinical research, personalized medicine, and population health management. Traditional approaches often require extensive manual chart reviews and rule-based algorithms, which can be time-consuming and difficult to generalize across different healthcare settings. GenAI models, by contrast, can rapidly analyze large datasets and identify complex clinical relationships with greater flexibility and efficiency. The reviewed literature also highlighted the application of GenAI in specialized clinical information extraction tasks. For example, researchers successfully employed these technologies to extract lung cancer staging descriptors from clinical documentation, enabling the automated identification of critical diagnostic and prognostic information [5]. Such applications support more efficient clinical data analysis while reducing the burden on healthcare professionals responsible for reviewing extensive patient records. Similarly, GenAI-assisted data mining techniques have been used to facilitate the identification of patients living with HIV by analyzing large collections of clinical records and detecting relevant disease indicators that may otherwise remain difficult to identify through conventional methods [6]. A consistent finding across the majority of studies was the superior performance of GenAI systems compared with manual human data processing, particularly when handling large and complex datasets. These models demonstrated the capacity to perform extraction and classification tasks with greater speed while maintaining high levels of accuracy. The ability to process thousands of records in a relatively short period offers significant advantages for healthcare organizations seeking to improve operational efficiency, conduct large-scale research, or support quality improvement initiatives. As healthcare data volumes continue to expand, automated approaches become increasingly necessary to ensure timely and effective information utilization. An additional strength identified within the literature was the capability of GenAI models to perform zero-shot data extraction [7]. This refers to the ability of a model to extract information or generate outputs in formats that were not explicitly included during its training process. Such flexibility represents a major advancement over traditional machine learning systems, which typically require extensive retraining or customization when introduced to new tasks. The feasibility of zero-shot extraction suggests that GenAI systems can adapt to evolving healthcare requirements and diverse clinical documentation formats with minimal additional development effort. Overall, the evidence indicates that GenAI has considerable potential to transform data manipulation within EMR environments. By facilitating efficient information extraction, supporting complex data mining activities, enabling disease phenotyping, and improving the accessibility of clinical information, these technologies can enhance both healthcare operations and medical research. As the volume and complexity of healthcare data continue to increase, GenAI-driven data manipulation is likely to become an increasingly important component of modern health information management systems.

GenAI for Patient Communication

One of the most promising applications of Generative Artificial Intelligence (GenAI) within Electronic Medical Record (EMR) systems is its ability to enhance patient communication. Effective communication between healthcare providers and patients is a fundamental component of high-quality healthcare, influencing patient satisfaction, treatment adherence, trust, and health outcomes. However, increasing clinical workloads and growing volumes of electronic patient messages have created substantial communication burdens for healthcare professionals. In response to these challenges, GenAI technologies, particularly large language models (LLMs), have emerged as valuable tools capable of assisting clinicians in generating patient-centered communications that are both efficient and informative [8]. The reviewed literature demonstrates growing interest in this application area, with 9 of the 55 included studies investigating the use of GenAI for patient communication. These studies explored the capacity of AI systems to draft responses to patient inquiries, facilitate electronic messaging, and improve the overall quality of interactions between healthcare providers and patients. The findings suggest that GenAI can support clinicians by reducing the administrative effort associated with managing large volumes of patient correspondence while maintaining acceptable standards of communication quality. Evidence from a study conducted within a urology setting revealed that nearly half of the responses generated by ChatGPT to patient inquiries were considered acceptable for use in patient messaging [8]. This finding highlights the potential of GenAI to function as a supportive communication tool capable of producing clinically relevant and understandable responses. By assisting healthcare professionals in drafting replies to routine patient questions, these technologies may reduce the time and effort required for electronic communication while allowing clinicians to focus on more complex aspects of patient care.

Similarly, research performed in breast reconstruction surgery clinics demonstrated that ChatGPT exhibited notable strengths in both empathy and informational accuracy when responding to patient concerns [9]. Empathy represents a critical component of effective healthcare communication, as patients often seek not only medical information but also reassurance, understanding, and emotional support. The ability of GenAI systems to generate responses perceived as compassionate and supportive suggests that these technologies may contribute to improving the patient experience when appropriately implemented and supervised. Beyond the direct enhancement of patient interactions, the integration of GenAI into communication workflows has also shown potential benefits for healthcare professionals. Several studies included in the review reported reductions in physician burnout and perceived workload following the implementation of AI-assisted communication systems [8,10,11]. Although some investigations found that objective measures of time expenditure remained relatively unchanged, clinicians frequently reported that AI support reduced the cognitive burden associated with repetitive messaging tasks. This finding is particularly significant given the growing concerns regarding clinician burnout, which has become a major challenge across healthcare systems worldwide.

The comparative quality of AI-generated and physician-generated communications remains an area of active investigation. Some studies found that messages produced by GenAI were perceived as more empathetic, concise, direct, and easier for patients to understand than those written by healthcare providers [9,12]. These findings suggest that AI systems may help standardize communication quality and ensure that important information is delivered clearly and consistently. However, the evidence is not entirely uniform. Other researchers reported that AI-generated responses could occasionally be more difficult to interpret, less personalized, or less aligned with patient expectations. In these studies, messages written directly by physicians were often preferred by both patients and healthcare professionals [9,13]. Despite the encouraging results, several important limitations continue to restrict the widespread adoption of GenAI for patient communication. Concerns regarding factual inaccuracies, inappropriate recommendations, and the potential for misleading information remain significant challenges [8,11,13]. Additionally, excessive reliance on AI-generated content may reduce the individualized nature of patient-provider interactions, potentially affecting patient trust and satisfaction. The need for personalization is particularly important in healthcare settings, where communication must account for individual patient circumstances, preferences, and clinical conditions. Overall, the current evidence indicates that GenAI possesses substantial potential to improve the efficiency and quality of patient communication within EMR environments. By assisting with message generation, enhancing empathetic communication, and reducing administrative burdens, these technologies can support both patients and healthcare professionals. Nevertheless, ongoing human oversight, careful validation, and appropriate governance frameworks remain essential to ensure the accuracy, safety, and personalization of AI-assisted patient communications as healthcare organizations continue to explore their broader implementation.

GenAI for Clinical Decision Making

Clinical decision making represents one of the most important and potentially transformative applications of Generative Artificial Intelligence (GenAI) within healthcare. By leveraging large language models (LLMs) to analyze clinical information, interpret patient data, and generate diagnostic or therapeutic recommendations, healthcare organizations aim to improve the quality, efficiency, and consistency of clinical care. Consequently, considerable attention has been directed toward evaluating the ability of GenAI systems to support healthcare professionals in making informed clinical decisions. Within the reviewed literature, 8 of the 55 included studies examined the role of GenAI in clinical decision-making processes, reflecting growing interest in this emerging area of application. Several studies reported encouraging findings regarding the diagnostic capabilities of GenAI models. In certain clinical scenarios, these systems demonstrated an ability to generate diagnostic assessments and treatment recommendations that were comparable to those provided by healthcare professionals. Such findings suggest that GenAI may serve as a valuable decision-support tool by assisting clinicians in interpreting patient information, identifying possible diagnoses, and considering appropriate management strategies. The capacity of these models to rapidly process large amounts of clinical information further highlights their potential value in complex healthcare environments where timely decision making is essential. Despite these promising results, the evidence also reveals significant limitations that currently restrict the safe and independent use of GenAI in clinical decision making. Several studies identified concerns regarding the accuracy, reliability, and safety of AI-generated recommendations. For example, evaluations conducted within emergency department settings found that GenAI systems demonstrated suboptimal diagnostic accuracy when applied to acute clinical cases [14]. Given the high-risk nature of emergency medicine, even minor inaccuracies may have serious implications for patient safety and clinical outcomes. Additional concerns emerged from studies examining triage decision making. Research findings indicated that GenAI generated unsafe triage recommendations in a considerable proportion of evaluated cases [15]. Accurate triage is essential for prioritizing patients according to clinical urgency and ensuring appropriate allocation of healthcare resources. Errors in this process could result in delayed treatment, inappropriate care pathways, or adverse patient outcomes, highlighting the need for cautious implementation of AI-supported triage systems.

Similarly, investigations assessing the use of GenAI for medication-related decision support revealed notable performance limitations. One study found that these models performed poorly when guiding renal dose adjustments for hospitalized patients [16]. Medication dosing often requires careful consideration of patient-specific clinical factors, including kidney function, comorbidities, and concurrent therapies. Inaccurate recommendations in such contexts may increase the risk of medication errors and patient harm. Another frequently reported limitation involves the phenomenon known as hallucination, whereby GenAI systems generate information that appears plausible but is factually incorrect. Several studies noted that these models occasionally fabricated diagnoses, clinical facts, or supporting references that lacked any basis in established evidence [17]. This tendency presents a substantial challenge in healthcare environments, where clinical decisions must be based on accurate, verifiable, and evidence-based information. Overall, while GenAI demonstrates considerable promise as a clinical decision-support technology, current evidence suggests that its role should remain supportive rather than autonomous. Continued validation, rigorous testing, and human oversight are essential to ensure patient safety and maintain clinical accuracy. As these technologies continue to evolve, their integration into clinical decision-making processes will require careful governance, ongoing evaluation, and adherence to established medical standards.

GenAI for Clinical Prediction

Clinical prediction represents a key area where Generative Artificial Intelligence (GenAI) is increasingly being evaluated for its ability to support healthcare decision-making through data-driven forecasting of patient outcomes. Within the scope of the reviewed literature, 8 out of 55 included studies examined the use of GenAI for clinical prediction tasks, reflecting a growing interest in its capacity to anticipate clinical events, stratify risk, and support proactive care planning. These studies collectively highlight both the strengths and current limitations of GenAI when applied to predictive healthcare analytics. A number of investigations demonstrate that GenAI models, particularly when carefully designed and appropriately integrated into clinical workflows, can perform effectively in specific prediction scenarios. In perioperative care settings, for example, GenAI has been successfully applied to classify patients according to their likelihood of requiring hospital admission versus intensive care unit (ICU) admission [18]. This type of prediction is clinically significant, as early identification of patients requiring higher levels of care enables better resource allocation, improved surgical planning, and enhanced patient safety outcomes. The ability of GenAI to synthesize complex clinical variables and generate structured predictive outputs highlights its potential utility in perioperative risk assessment.

Further evidence supports the effectiveness of GenAI in neurological prediction tasks. In studies focused on seizure recurrence, large language models demonstrated higher predictive accuracy compared to traditional models that relied exclusively on structured clinical data [19]. This finding suggests that GenAI may offer added value by incorporating unstructured clinical narratives alongside structured variables, thereby capturing a more comprehensive representation of patient health status. The integration of diverse data sources appears to enhance model performance, particularly in conditions characterized by complex and multifactorial disease progression. Overall, these findings indicate that GenAI has considerable potential as a clinical prediction tool across a variety of healthcare domains. Its ability to process large volumes of heterogeneous data, including both structured and unstructured information, enables it to generate predictive insights that may complement existing clinical decision-support systems. In its current form, GenAI demonstrates particular promise in scenarios where traditional predictive models may be limited by data constraints or insufficient contextual understanding. However, despite these encouraging results, the application of GenAI in clinical prediction remains an evolving field. Model performance is highly dependent on data quality, training design, and clinical context, and variability in these factors can significantly influence predictive accuracy. As such, while GenAI shows strong potential to enhance clinical forecasting and support proactive healthcare interventions, its use must be guided by rigorous validation, continuous monitoring, and integration with clinical expertise.

GenAI for Summarization

Generative Artificial Intelligence (GenAI) has expanded its utility in healthcare beyond interactive communication to include the critical task of medical text summarization. This function is particularly important within Electronic Medical Record (EMR) systems, where large volumes of complex and often unstructured clinical documentation can create barriers to efficient information use and patient understanding. Summarization tools powered by large language models (LLMs) aim to transform dense medical content into concise, structured, and more accessible formats, thereby improving usability for both clinicians and patients. Within the reviewed literature, 4 out of 55 studies examined the application of GenAI for summarization tasks. Although this represents a smaller proportion compared to other use cases, the findings highlight meaningful potential for improving communication and health literacy. One key application involves the transformation of inpatient discharge summaries into clearer and more readable formats [20]. Discharge summaries typically contain highly technical language, complex clinical terminology, and multiple instructions that patients are expected to follow after hospital discharge. Misinterpretation of this information can lead to medication errors, poor adherence to treatment plans, and increased risk of hospital readmission. GenAI-driven summarization offers a method to simplify this information while preserving clinical accuracy, making it more accessible to patients with varying levels of health literacy.

Another important application identified in the literature is the simplification of radiology reports [21]. Radiology documentation is often highly specialized and written in technical language intended for healthcare professionals. While this format is essential for clinical precision, it can be difficult for patients to understand their diagnostic results. GenAI systems have been used to “translate” these reports into lower reading-level summaries, enabling patients to better comprehend their medical conditions and participate more actively in shared decision-making processes. This capability supports patient-centered care by bridging the communication gap between clinical expertise and patient understanding. The findings across these studies suggest that GenAI has significant potential to enhance both clinical efficiency and patient empowerment through improved communication. By converting complex medical narratives into simplified and structured summaries, these systems can reduce cognitive burden for both healthcare providers and patients. For clinicians, summarization tools may streamline documentation review and support faster clinical decision-making. For patients, improved clarity in medical information can lead to better understanding of diagnoses, treatment plans, and follow-up instructions. Despite these advantages, the use of GenAI for summarization also requires careful consideration of accuracy, context preservation, and potential oversimplification of critical clinical details. Ensuring that essential medical information is not lost or misrepresented during the summarization process remains a key challenge. Therefore, while GenAI demonstrates strong promise in this domain, its implementation must be guided by clinical oversight and continuous validation to ensure safe and effective use in healthcare environments.

GenAI for Other Use Cases

Beyond the primary categories of application identified in this review, a small number of studies demonstrated more diverse and exploratory uses of Generative Artificial Intelligence (GenAI) within healthcare contexts. These applications do not align neatly with the predefined thematic classifications, yet they provide important insight into the expanding scope of large language models (LLMs) in clinical informatics and health services research. Among the 55 included studies, 2 were categorized under “other” due to their broad or non-conventional applications of GenAI. The first study presented a multifaceted application of LLMs that encompassed several interrelated clinical documentation and information management tasks [22]. Specifically, the model was evaluated for its ability to summarize clinical notes, extract relevant patient data, and generate structured documents such as referral letters and clinical memos. This combined functionality reflects the versatility of GenAI systems in handling multiple components of clinical workflow automation simultaneously. Rather than being limited to a single discrete task, the model demonstrated the capacity to operate across several stages of health information processing, from raw documentation interpretation to the production of clinically useful outputs. Such capabilities highlight the potential for GenAI to serve as an integrated support tool within EMR systems, reducing administrative burden while improving the consistency and efficiency of clinical documentation practices. The second study extended the application of GenAI beyond direct EMR functionalities and into the domain of health equity research [23]. This investigation explored how LLMs could be utilized to assess disparities in seizure outcomes across different patient populations. Although not directly focused on EMR operations, the study illustrated how GenAI can be leveraged to analyze clinical data patterns in a way that supports broader healthcare analytics and decision-making processes. By enabling more efficient identification and interpretation of outcome disparities, such applications may contribute to a deeper understanding of inequities in healthcare delivery and inform targeted interventions aimed at improving population-level outcomes. Collectively, these “other” use cases underscore the adaptability of GenAI technologies and their potential to extend beyond traditional clinical documentation or predictive functions. The ability of LLMs to integrate multiple tasks such as summarization, data extraction, and document generation suggests a pathway toward more unified and intelligent health information systems. At the same time, their application in health equity research highlights an emerging role in population health analysis and policy-relevant investigations. While these studies remain exploratory in nature, they indicate that the future integration of GenAI into healthcare systems may not be limited to isolated applications. Instead, it may evolve toward multifunctional platforms capable of supporting both clinical workflows and advanced health analytics. However, as with other applications, careful validation and governance will be essential to ensure accuracy, transparency, and ethical use in real-world healthcare environments.

Discussion

Artificial intelligence in medicine has been an expanding field of research for more than a decade, with consistent interest in its potential to improve clinical outcomes, operational efficiency, and decision-making processes. However, despite substantial academic progress, large-scale integration of AI systems into routine clinical practice has remained limited. Early advancements in medical AI were primarily concentrated in domains characterized by clearly defined tasks and the availability of large, high-quality labeled datasets, particularly radiology and pathology [24]. These fields provided structured environments where traditional machine learning models could be trained effectively to perform classification and detection tasks with measurable outcomes. The emergence of transformer-based architectures and commercially available large language models (LLMs) has marked a significant shift in this trajectory. These models demonstrate an enhanced ability to process and generate coherent, contextually appropriate textual outputs with increasing accuracy. Unlike earlier AI systems that required task-specific design and extensive feature engineering, GenAI models exhibit greater flexibility and adaptability across a wide range of clinical applications. When integrated appropriately into healthcare environments, these systems have the potential to improve workflow efficiency, enhance scalability of information processing, ensure greater consistency in clinical documentation, and reduce the administrative burden placed on healthcare professionals.

Electronic Medical Records (EMRs) have long served as a central platform for innovation in healthcare informatics. Early applications of artificial intelligence within EMRs were largely dependent on conventional machine learning techniques, which required carefully curated datasets and predefined variables. These approaches, while effective in specific contexts, were often limited by their reliance on structured data and lack of adaptability to unstructured clinical narratives. In contrast, findings from this review indicate that modern generative models are capable of matching or surpassing the performance of traditional methods in several key applications. For example, a large-scale institutional model known as NYUTron demonstrated superior performance compared to conventional risk stratification models in predicting outcomes such as mortality, hospital readmission, and insurance denial [25]. This highlights the ability of LLMs to leverage diverse clinical inputs to generate more accurate predictive insights. Similarly, other studies have shown that LLM-based approaches can enhance clinical forecasting by incorporating unstructured textual data from EMRs, thereby improving predictive accuracy beyond what is achievable using structured data alone. In neurological applications, LLMs have demonstrated improved performance in predicting seizure recurrence when compared to classical predictive algorithms [19]. These findings underscore the value of integrating narrative clinical documentation into predictive modeling frameworks, enabling a more comprehensive understanding of patient trajectories.

Hybrid methodologies have also emerged as an important area of development. For instance, combining dictionary-based natural language processing techniques with LLM-based phenotyping approaches has been shown to improve rare disease identification, outperforming traditional ontological or rule-based systems [4]. This suggests that integrating structured clinical knowledge with generative modeling techniques may yield more robust and accurate outcomes. Additionally, the use of zero-shot and few-shot prompting strategies has demonstrated promising results in tasks such as data extraction [7] and cancer symptom identification [2], often requiring minimal annotated data while still exceeding the performance of older machine learning frameworks. Collectively, these findings suggest a gradual transition in medical artificial intelligence from highly specialized, narrowly focused models toward more generalist and adaptable systems capable of handling multiple clinical tasks. The traditional paradigm of "gold-standard" machine learning, which depends heavily on static training datasets and extensive manual feature engineering, may be increasingly supplemented or replaced by flexible generative models that can adapt to diverse clinical scenarios with limited additional training. While specialized machine learning approaches will likely remain relevant in well-defined applications, the broader capability of LLMs to process unstructured clinical text at scale introduces a new dimension of functionality for EMR systems. Despite these advantages, important challenges remain, particularly regarding model reliability, interpretability, and the risk of hallucinated outputs. Nevertheless, the integration of GenAI into EMR systems holds significant promise for transforming healthcare delivery by enabling more dynamic, efficient, and comprehensive utilization of clinical data across a wide range of applications.

GenAI Integration in EMRs: Clinical Impact, Opportunities, and Challenges

Generative Artificial Intelligence (GenAI) integration within Electronic Medical Records (EMRs) holds substantial potential to transform healthcare delivery by influencing both direct and indirect patient outcomes. At a direct clinical level, GenAI can function as a decision-support mechanism that assists healthcare professionals in identifying appropriate diagnostic pathways, suggesting treatment options, and reducing the likelihood of medical errors. By analyzing vast quantities of patient data in real time, these systems can support clinicians in making more informed and timely decisions. This is particularly relevant in high-pressure clinical environments where rapid decision-making is essential for patient survival and optimal outcomes. For example, the automation of acute respiratory distress syndrome (ARDS) diagnosis based on the Berlin criteria illustrates how GenAI tools can enhance diagnostic efficiency and support structured clinical reasoning [26]. Such applications demonstrate how computational assistance can improve adherence to standardized diagnostic frameworks while simultaneously reducing cognitive workload for clinicians. Beyond direct clinical decision support, GenAI also contributes to improved patient outcomes through indirect mechanisms. Applications such as data manipulation, predictive modeling, and summarization can significantly enhance physician efficiency by reducing the administrative burden associated with EMR use. A considerable portion of clinicians' time is currently dedicated to documentation, chart review, and data retrieval tasks, all of which limit direct patient interaction. By automating or streamlining these processes, GenAI allows healthcare professionals to reallocate time toward patient-centered care, complex clinical problem-solving, and shared decision-making. Improved efficiency in information processing also enables better prioritization of cases, ensuring that high-risk patients receive timely attention. In this sense, GenAI does not only enhance the technical aspects of healthcare delivery but also contributes to improving the overall structure and allocation of clinical time.

The growing momentum behind GenAI integration is also reflected in the increasing commitment of both healthcare organizations and technology providers to embed AI-driven tools into EMR systems. Large-scale collaborations between healthcare institutions and technology companies are actively working toward developing real-time, cloud-based solutions for clinical analytics and workflow automation. For instance, Epic has introduced new generative AI features aimed at enhancing clinical documentation and enabling intelligent chart review processes [27]. These developments indicate that GenAI is transitioning from experimental applications into structured enterprise-level deployment. In parallel, both emerging startups and established technology companies are investing heavily in healthcare-focused GenAI systems [28], signaling a broader industry-wide shift toward AI-enabled healthcare ecosystems. However, the integration of GenAI into EMRs extends far beyond technical implementation. Ethical considerations play a central role in determining how these technologies should be deployed in clinical settings. Key concerns include accountability for AI-generated recommendations, informed patient consent for AI-assisted care, and the protection of sensitive health information. Ensuring patient privacy is particularly critical in cloud-based systems where large-scale data processing is involved. Additionally, education for both patients and healthcare professionals is essential to ensure a clear understanding of GenAI capabilities, limitations, and appropriate use cases. Without adequate training and awareness, there is a risk of misuse or overdependence on automated systems, which may compromise patient safety. Despite rapid advancements, GenAI remains in an early stage of clinical integration, and trust among healthcare professionals is still developing. Concerns persist regarding the reliability of AI-generated outputs, particularly in high-stakes clinical decision-making contexts. Issues such as data breaches, inaccuracies in generated responses, and lack of legal clarity surrounding AI-induced medical errors contribute to hesitation among clinicians. Liability remains an unresolved issue, with ongoing debate regarding whether responsibility should lie with clinicians, healthcare institutions, or AI developers. These uncertainties highlight the urgent need for comprehensive regulatory frameworks that clearly define roles, responsibilities, and standards for safe AI deployment in healthcare environments.

Importantly, current evidence indicates that GenAI should be viewed as a tool for augmenting clinical expertise rather than replacing it. Studies consistently demonstrate that while GenAI can enhance efficiency and support clinical workflows, it lacks the autonomy and reliability required for independent decision-making. For example, research evaluating ChatGPT's ability to respond to clinical decision support alerts for patients with renal dysfunction found that it performed inconsistently, achieving correct and identical responses in fewer than 20% of cases [16]. Such findings reinforce the conclusion that GenAI is not yet suitable for fully autonomous integration into EMR-based clinical decision systems. Several limitations further constrain the safe deployment of GenAI in healthcare. Overreliance on these systems may lead to diminished clinical vigilance, particularly given the known risks of hallucinations, factual inaccuracies, and contextually inappropriate outputs. In addition, since many GenAI models are trained on large and diverse internet-based datasets, they may inadvertently reproduce embedded biases, potentially exacerbating existing healthcare disparities and inequities, especially among underserved populations. This raises significant concerns about fairness and equitable access to high-quality care. Furthermore, the "black box" nature of many GenAI systems presents a major challenge for clinical adoption. Limited interpretability makes it difficult for clinicians to understand how specific outputs or recommendations are generated, reducing transparency and potentially undermining trust in AI-assisted decision-making. Nearly all studies reviewed in this field acknowledge at least one of these limitations, reinforcing the importance of maintaining human oversight in all AI-supported clinical processes. Ultimately, the successful integration of GenAI into EMRs will depend not only on technological advancement but also on careful attention to ethical, legal, and clinical governance frameworks. A balanced approach that emphasizes regulation, bias mitigation, transparency, and clinician involvement will be essential to ensure that GenAI enhances healthcare delivery without compromising safety, equity, or professional accountability.

Conclusion

Generative Artificial Intelligence within Electronic Medical Records represents a major shift in healthcare informatics. The evidence reviewed shows consistent utility across multiple domains including data manipulation, clinical prediction, decision support, communication, and summarization. These applications improve efficiency, reduce documentation burden, and enhance access to clinical information. Some models demonstrate performance comparable to or exceeding traditional machine learning approaches, particularly when unstructured clinical text is leveraged. However, limitations remain significant. Issues related to hallucination, bias, lack of interpretability, variable clinical accuracy, and legal uncertainty restrict safe autonomous use. Clinical decision-making tasks show the highest risk, while administrative and supportive tasks show stronger reliability. Current findings support GenAI as an assistive technology rather than a replacement for clinicians. Its future success depends on regulatory frameworks, clinical validation, and responsible integration into workflows. Overall, GenAI has strong potential to reshape EMR systems, but its adoption must remain cautious, structured, and continuously evaluated to ensure patient safety and healthcare equity.

References

1. Introducing ChatGPT. OpenAI Blog. Available online: https://openai.com/index/chatgpt/
2. Zeinali, N.; Albashayreh, A.; Fan, W.; White, S.G. Symptom-BERT: Enhancing Cancer Symptom Detection in EHR Clinical Notes. *J. Pain Symptom Manag.* 2024, *68*, 190–198.e1.
3. Pathak, A.; Yu, Z.; Paredes, D.; Monsour, E.P.; Rocha, A.O.; Brito, J.P.; Ospina, N.S.; Wu, Y. Extracting Thyroid Nodules Characteristics from Ultrasound Reports Using Transformer-based Natural Language Processing Methods. *AMIA Annu. Symp. Proc.* 2023, *2023*, 1193–1200.
4. Wu, J.; Dong, H.; Li, Z.; Wang, H.; Li, R.; Patra, A.; Dai, C.; Ali, W.; Scordis, P.; Wu, H. A hybrid framework with large language models for rare disease phenotyping. *BMC Med. Inform. Decis. Mak.* 2024, *24*, 289.
5. Cho, H.; Yoo, S.; Kim, B.; Jang, S.; Sunwoo, L.; Kim, S.; Lee, D.; Kim, S.; Nam, S.; Chung, J.-H. Extracting lung cancer staging descriptors from pathology reports: A generative language model approach. *J. Biomed. Inform.* 2024, *157*, 104720.
6. Morales-Sánchez, R.; Montalvo, S.; Riaño, A.; Martínez, R.; Velasco, M. Early diagnosis of HIV cases by means of text mining and machine learning models on clinical notes. *Comput. Biol. Med.* 2024, *179*, 108830.
7. Hu, D.; Liu, B.; Zhu, X.; Lu, X.; Wu, N. Zero-shot information extraction from radiological reports using ChatGPT. *Int. J. Med. Inform.* 2024, *183*, 105321.
8. Scott, M.; Muncey, W.; Seranio, N.; Belladelli, F.; Del Giudice, F.; Li, S.; Ha, A.; Glover, F.; Zhang, C.A.; Eisenberg, M.L. Assessing Artificial Intelligence–Generated Responses to Urology Patient In-Basket Messages. *Urol. Pract.* 2024, *11*, 793–798.
9. Soroudi, D.B.; Gozali, A.M.; Knox, J.A.; Parmeshwar, N.; Sadjadi, R.; Wilson, J.C.B.; Lee, S.A.; Piper, M.L. Comparing Provider and ChatGPT Responses to Breast Reconstruction Patient Questions in the Electronic Health Record. *Ann. Plast. Surg.* 2024, *93*, 541–545.
10. Liu, S.; McCoy, A.B.; Wright, A.P.; Carew, B.; Genkins, J.Z.; Huang, S.S.; Peterson, J.F.; Steitz, B.; Wright, A. Leveraging large language models for generating responses to patient messages—A subjective analysis. *J. Am. Med. Inform. Assoc.* 2024, *31*, 1367–1379.
11. Garcia, P.; Ma, S.P.; Shah, S.; Smith, M.; Jeong, Y.; Devon-Sand, A.; Tai-Seale, M.; Takazawa, K.; Clutter, D.; Vogt, K.; et al. Artificial Intelligence–Generated Draft Replies to Patient Inbox Messages. *JAMA Netw. Open* 2024, *7*, e243201.
12. Small, W.R.; Wiesenfeld, B.; Brandfield-Harvey, B.; Jonassen, Z.; Mandal, S.; Stevens, E.R.; Major, V.J.; Lostraglio, E.; Szerencsy, A.; Jones, S.; et al. Large Language Model–Based Responses to Patients' In-Basket Messages. *JAMA Netw. Open* 2024, *7*, e2422399.
13. Reynolds, K.; Nadelman, D.; Durgin, J.; Ansah-Addo, S.; Cole, D.; Fayne, R.; Harrell, J.; Ratycz, M.; Runge, M.; Shepard-Hayes, A.; et al. Comparing the quality of ChatGPT- and physician-generated responses to patients' dermatology questions in the electronic medical record. *Clin. Exp. Dermatol.* 2024, *49*, 715–718.
14. Williams, C.Y.K.; Miao, B.Y.; Kornblith, A.E.; Butte, A.J. Evaluating the use of large language models to provide clinical recommendations in the Emergency Department. *Nat. Commun.* 2024, *15*, 8236.
15. Fraser, H.; Crossland, D.; Bacher, I.; Ranney, M.; Madsen, T.; Hilliard, R. Comparison of Diagnostic and Triage Accuracy of Ada Health and WebMD Symptom Checkers, ChatGPT, and Physicians for Patients in an Emergency Department: Clinical Data Analysis Study. *JMIR mHealth uHealth* 2023, *11*, e49995.
16. van Nuland, M.; Snoep, J.D.; Egberts, T.; Erdogan, A.; Wassink, R.; van der Linden, P.D. Poor performance of ChatGPT in clinical rule-guided dose interventions in hospitalized patients with renal dysfunction. *Eur. J. Clin. Pharmacol.* 2024, *80*, 1133–1140.
17. Bridges, J.M. Computerized diagnostic decision support systems—A comparative performance study of Isabel Pro vs. ChatGPT4. *Diagnosis* 2024, *11*, 250–258.

18. Chung, P.; Fong, C.T.; Walters, A.M.; Aghaeepour, N.; Yetisgen, M.; O'reilly-Shah, V.N. Large Language Model Capabilities in Perioperative Risk Prediction and Prognostication. *JAMA Surg.* 2024, *159*, 928–937.
19. Beaulieu-Jones, B.K.; Villamar, M.F.; Scordis, P.; Bartmann, A.P.; Ali, W.; Wissel, B.D.; Alsentzer, E.; de Jong, J.; Patra, A.; Kohane, I. Predicting seizure recurrence after an initial seizure-like episode from routine clinical notes using large language models: A retrospective cohort study. *Lancet Digit. Health* 2023, *5*, e882–e894.
20. Zaretsky, J.; Kim, J.M.; Baskharoun, S.; Zhao, Y.; Austrian, J.; Aphinyanaphongs, Y.; Gupta, R.; Blecker, S.B.; Feldman, J. Generative Artificial Intelligence to Transform Inpatient Discharge Summaries to Patient-Friendly Language and Format. *JAMA Netw. Open* 2024, *7*, e240357.
21. Doshi, R.; Amin, K.S.; Khosla, P.; Bajaj, S.; Chheang, S.; Forman, H.P. Quantitative Evaluation of Large Language Models to Streamline Radiology Report Impressions: A Multimodal Retrospective Analysis. *Radiology* 2024, *310*, e231593.
22. Tan, R.E.S.; Teo, W.Z.W.; Puhaindran, M.E. Artificial Intelligence in Hand Surgery–How Generative AI is Transforming the Hand Surgery Landscape. *J. Hand Surg.* 2024, *29*, 81–87.
23. Xie, K.; Ojemann, W.K.S.; Gallagher, R.S.; Shinohara, R.T.; Lucas, A.; Hill, C.E.; Hamilton, R.H.; Johnson, K.B.; Roth, D.; Litt, B.; et al. Disparities in seizure outcomes revealed by large language models. *J. Am. Med. Inform. Assoc.* 2023, *31*, 1348–1355.
24. Alowais, S.A.; Alghamdi, S.S.; Alsuhebany, N.; Alqahtani, T.; Alshaya, A.I.; Almohareb, S.N.; Aldairem, A.; Alrashed, M.; Bin Saleh, K.; Badreldin, H.A.; et al. Revolutionizing healthcare: The role of artificial intelligence in clinical practice. *BMC Med. Educ.* 2023, *23*, 689.
25. Jiang, L.Y.; Liu, X.C.; Pour Nejatian, N.; Nasir-Moin, M.; Wang, D.; Abidin, A.; Eaton, K.; Riina, H.A.; Laufer, I.; Punjabi, P.; et al. Health system-scale language models are all-purpose prediction engines. *Nature* 2023, *619*, 357–362.
26. Gandomi, A.; Wu, P.; Clement, D.R.; Xing, J.; Aviv, R.; Federbush, M.; Yuan, Z.; Jing, Y.; Wei, G.; Hajizadeh, N. ARDSFlag: An NLP/machine learning algorithm to visualize and detect high-probability ARDS admissions independent of provider recognition and billing codes. *BMC Med. Inform. Decis. Mak.* 2024, *24*, 195.
27. Capoot, A. Epic Systems is building more than 100 new AI features for doctors and patients. Here's what's coming. *CNBC*, 21 August 2024. Available online: https://www.cnbc.com/2024/08/21/epic-systems-ugm-2024-ai-tools-in-mychart-cosmos-.html
28. Milliard, M. New Oracle EHR Promises AI-Enabled Reinvention. *Healthcare IT News*, 30 October 2024. Available online: https://www.healthcareitnews.com/news/new-oracle-ehr-promises-ai-enabled-reinvention