Global Journal of Engineering and Technology [GJET].
ISSN: **2583-3359** (Online)
Frequency: Monthly
Published By GSAR Publishers
Journal Homepage Link- https://gsarpublishers.com/journal-gjet-home/

Large Language Model–Enabled Health Information Systems: A Systematic Review of Fhir-Based Interoperability, Clinical Decision Support, Privacy, and Trustworthy Ai Governance

By

Miracle A Atianashie[1], Chukwuma Chinaza Adaobi [2], Caleb Boakye Yiadom [3]
[1]Research Fellow, Quality Assurance Directorate, Catholic University of Ghana
[2]General Manager, Metascholar Consult Limited, Ghana
[3]SSIU, Metascholar Consult Limited, Ghana

Article History

Received: 02/07/2026
Accepted: 05/07/2026
Published: 07/07/2026

Vol – 5 Issue – 7

PP: - 01-18

Abstract

Large Language Model (LLM)–enabled Health Information Systems are increasingly transforming how clinical data are captured, exchanged, interpreted, and governed. This systematic review examined recent evidence on the integration of LLMs into Health Information Systems, with emphasis on FHIR-based interoperability, clinical decision support, privacy protection, and trustworthy AI governance. The review adopted a systematic methodology and synthesised studies published from 2020 onward across health informatics, clinical artificial intelligence, digital health, interoperability, privacy-preserving computing, and governance literature. The findings show that LLMs can support health systems by converting unstructured clinical narratives into structured data, improving documentation, assisting medical evidence summarisation, enhancing patient communication, and strengthening decision-support workflows. FHIR and SMART on FHIR emerged as important foundations for standardised and reusable clinical data exchange. However, the review also found that LLM-enabled systems remain limited by hallucination, weak clinical validation, privacy leakage, algorithmic bias, limited transparency, and uncertain accountability. Privacy-preserving approaches such as federated learning, secure computation, encryption, and controlled retrieval offer useful safeguards, but they require strong institutional governance and continuous monitoring. The review concludes that LLM-enabled Health Information Systems should be implemented as human-supervised, standards-based, and governance-driven clinical infrastructures rather than autonomous decision-making tools. Responsible adoption requires FHIR-conformant architecture, clinician oversight, privacy-by-design, fairness evaluation, transparent reporting, regulatory alignment, and lifecycle governance.

Keywords: *Large Language Models, Health Information Systems, FHIR Interoperability, Clinical Decision Support, Trustworthy AI Governance*

1. INTRODUCTION

Health Information Systems are increasingly becoming intelligent, interoperable, and data-driven infrastructures for supporting clinical documentation, diagnosis, treatment planning, patient engagement, administrative coordination, and population health surveillance. Traditionally, these systems have depended on structured electronic health records, rule-based clinical decision support, and manually coded clinical data. However, the rapid development of Large Language Models has introduced a new generation of AI-enabled health information systems capable of interpreting, summarising, generating, and reasoning over complex clinical language. LLMs are particularly significant because much of healthcare knowledge is still stored in unstructured formats, including clinical notes, discharge summaries, referral letters, laboratory interpretations, imaging reports, medication narratives, and patient-provider communication records [1]. Their ability to process free-text clinical data has therefore created new possibilities for improving the usability, responsiveness, and analytical capacity of health information systems [2]. The emergence of LLM-enabled health information systems is closely linked to the broader shift toward foundation models and generalist medical artificial intelligence. Unlike narrow AI models designed for single clinical tasks, foundation models can be adapted across multiple healthcare functions, including question answering, summarisation, patient education, documentation support, and decision assistance [3]. In health information systems, this flexibility is important because clinical workflows are rarely isolated. A single patient encounter may require retrieval of past

*Corresponding Author: Miracle A Atianashie

medical history, interpretation of laboratory results, medication reconciliation, risk stratification, communication with patients, and referral documentation. LLMs may therefore serve as a natural-language interface between clinicians, patients, and complex digital health records.

A central challenge in this transformation is interoperability. Healthcare systems often operate through fragmented databases, incompatible coding structures, and heterogeneous electronic health record platforms. The Fast Healthcare Interoperability Resources standard has become a major response to this problem because it supports modular, web-based exchange of health information across systems [4]. FHIR enables clinical data to be represented as standardised resources such as Patient, Observation, Condition, MedicationRequest, Encounter, Diagnostic Report, and Procedure. In LLM-enabled health information systems, FHIR is important because it provides a structured foundation for transforming unstructured clinical narratives into machine-readable and exchangeable health data. Recent work on FHIR-GPT demonstrates that LLMs can support conversion of clinical narratives into FHIR medication-related resources, suggesting that generative AI can contribute to health data standardisation and interoperability when combined with validation mechanisms [5]. Beyond interoperability, LLMs are gaining attention for clinical decision support. Studies have shown that advanced LLMs can encode substantial medical knowledge and perform strongly on medical question-answering benchmarks [6]. ChatGPT's performance on the United States Medical Licensing Examination also demonstrated the capacity of general-purpose LLMs to generate clinically relevant explanations and reasoning patterns [7]. Similarly, GPT-4 has been discussed as a potentially useful medical chatbot, although its risks remain serious in clinical contexts [8]. In patient-facing communication, AI chatbot responses have been rated highly for quality and empathy when compared with physician responses to public patient questions, suggesting possible value in supporting patient education and response drafting [9]. These developments show that LLMs may enhance health information systems by improving information retrieval, clinical documentation, triage support, and communication.

However, the clinical promise of LLM-enabled systems must be interpreted carefully. LLMs can produce fluent but incorrect outputs, generate unsupported recommendations, and omit clinically important details, or present uncertainty in a misleadingly confident manner. Reviews of LLMs in medicine therefore emphasise that their use must be accompanied by careful validation, domain-specific testing, clinician oversight, and clear boundaries of responsibility [10]. In health information systems, these risks are amplified because the output of a model may influence real clinical action. A wrong summary, inaccurate medication interpretation, or unsafe decision-support recommendation may affect diagnosis, treatment, referral, or patient trust. Privacy and data protection represent another major background issue. Health information systems manage highly sensitive personal data, and LLM integration may expose patient information through prompts, outputs, model memory, third-party processing, or insecure deployment pipelines. Privacy-preserving approaches, including local LLM deployment, controlled information retrieval, de-identification, access control, and structured extraction from clinical text, are increasingly being explored to reduce these risks [11]. Nevertheless, technical privacy measures alone are insufficient. Ethical concerns remain regarding consent, secondary use of data, transparency, patient autonomy, accountability, and the possibility of hidden data leakage [12].

Trustworthy AI governance is therefore essential. The ethics literature on ChatGPT and LLMs in healthcare identifies recurring concerns around fairness, bias, misinformation, opacity, non-maleficence, privacy, and human oversight [13]. Bias is especially important because LLMs can reproduce historical inequities embedded in training data. Evidence that LLMs may propagate race-based medical misconceptions raises serious concerns about their use in clinical decision support and patient-facing systems [14]. For health information systems, this means that LLM outputs must be evaluated not only for technical accuracy but also for fairness across demographic groups, clinical populations, languages, and health system contexts. Regulatory and governance debates have also become more urgent. LLM-enabled health tools differ from earlier AI systems because they are general-purpose, adaptive, conversational, and capable of being used across multiple clinical and non-clinical tasks. These characteristics complicate medical device regulation, liability, cybersecurity, data provenance, intellectual property, and professional accountability [15]. The increasing number of AI-based medical devices and algorithms approved by regulators shows that AI is already entering clinical practice, but LLM-enabled systems require even stronger lifecycle governance because their behaviour may vary across prompts, settings, users, and data contexts [16].

Clinical evaluation standards are also evolving to address AI-specific risks. CONSORT-AI provides reporting guidance for clinical trial reports involving AI interventions [17], while SPIRIT-AI strengthens protocol reporting for clinical trials involving AI systems [18]. DECIDE-AI further supports early-stage clinical evaluation of AI-based decision support systems before wider deployment [19]. For prediction models, TRIPOD+AI provides updated reporting guidance for regression-based and machine-learning-based clinical prediction models [20]. These frameworks are highly relevant to LLM-enabled health information systems because they emphasise transparency, clinical context, human-AI interaction, error analysis, and real-world evaluation.

2. RELATED STUDIES

Recent studies on Large Language Model–enabled Health Information Systems show that the field is developing around four closely connected concerns: clinical usefulness, data interoperability, privacy protection, and trustworthy governance. The literature does not treat LLMs merely as conversational tools; rather, it increasingly examines them as components of digital health infrastructure that may support electronic health records, clinical decision support, documentation, patient communication, and health data standardisation. However, related studies also show that LLM-enabled systems remain uneven in maturity. Some studies demonstrate strong performance in summarisation, evidence synthesis, FHIR-based data exchange, and clinical workflow support, while others warn that hallucination, poor reproducibility, privacy leakage, bias, and weak accountability may

*Corresponding Author: Miracle A Atianashie

undermine safe implementation. The following related studies are organised into four subtopics: LLM-enabled clinical decision support, FHIR-based interoperability, privacy-preserving health data processing, and trustworthy AI governance.

2.2 LLM-Enabled Clinical Decision Support and Clinical Workflow Support

Related studies show that LLM-enabled clinical decision support is one of the most active areas of health information system research. Hager et al. [21] evaluated the limitations of LLMs in clinical decision-making and found that high performance on medical examinations does not necessarily translate into safe clinical reasoning, especially when models must interpret incomplete information, follow clinical guidelines, and operate within realistic workflow constraints. Their study is important because it shifts attention from benchmark scores to practical clinical decision-making conditions. Van Veen et al. [22] also demonstrated the value of adapted LLMs in clinical text summarisation, showing that carefully adapted models could produce summaries comparable to, and sometimes better than, expert-generated summaries. This finding suggests that LLMs may reduce documentation burden in health information systems. Tang et al. [23] examined medical evidence summarisation and showed that LLMs can assist with synthesising clinical evidence, although their outputs require careful factual verification. Reddy [24] proposed a translational value assessment framework, arguing that healthcare LLMs must be evaluated according to clinical value, safety, usability, fairness, and implementation readiness rather than technical performance alone. Cascella et al. [25] similarly explored multiple healthcare scenarios and concluded that ChatGPT-like tools may support education, documentation, research, and clinical communication, but cannot replace professional judgement. Collectively, these studies suggest that LLM-enabled decision support should be designed as an assistive, human-supervised layer within health information systems rather than as an autonomous clinical authority.

2.3 FHIR-Based Interoperability and Health Data Standardisation

FHIR-based interoperability is central to related studies because LLM-enabled health information systems require structured, exchangeable, and semantically meaningful clinical data. Mandl et al. [26] discussed the SMART/HL7 FHIR Bulk Data Access API as a mechanism for population-level extraction of standardised health data from electronic health records. Their work is relevant because LLM-enabled systems require scalable access to structured clinical data if they are to provide context-aware support across patient populations. Jones et al. [27] extended this discussion through a landscape survey of planned SMART/HL7 Bulk FHIR implementations and tools, showing that FHIR-based APIs were becoming increasingly important for analytics, research, and system-wide data exchange. Griffin et al. [28] examined real-world health applications using FHIR and found that FHIR apps vary in clinical focus, technical design, and implementation maturity. This indicates that interoperability standards alone do not guarantee effective deployment; usability, security, workflow fit, and governance remain important. Kapitan et al. [29] further argued that interoperability depends not only on the selection of open standards but also on open-source implementations, institutional context, and the practical ecosystem around each standard. Tsafnat et al. [30] added that the coexistence of multiple open health data standards can create both opportunity and fragmentation. Together, these studies suggest that LLM-enabled health information systems should not rely on free-text prompts alone. They require FHIR-conformant architecture, validated data exchange, terminology alignment, and implementation strategies that connect LLM outputs to standardised clinical resources.

2.4 Privacy-Preserving Health Data Processing in LLM-Enabled Systems

Privacy is a major issue in related studies because LLM-enabled health information systems may process sensitive patient histories, diagnoses, medications, laboratory values, imaging reports, and clinician notes. Rieke et al. [31] argued that federated learning could support digital health by allowing institutions to develop machine learning models without directly sharing patient data. This is relevant to LLM-enabled systems because many hospitals cannot centralise patient data due to privacy, legal, and ethical restrictions. Kaissis et al. [32] also examined secure, privacy-preserving, and federated machine learning in medical imaging, emphasising that privacy protection must include technical safeguards, secure computation, and governance controls. Sheller et al. [33] demonstrated the feasibility of federated learning in medicine by showing how institutions could collaborate on model development without sharing raw patient data. Dayan et al. [34] applied federated learning to prediction of clinical outcomes in patients with COVID-19, showing that privacy-preserving collaboration can produce clinically useful models across institutions. Feretzakis et al. [35] reviewed privacy-preserving techniques in generative AI and LLMs, including differential privacy, federated learning, homomorphic encryption, and secure multi-party computation. Their work is especially relevant because LLM-enabled health information systems may expose sensitive information through prompts, outputs, embeddings, or model updates. These studies collectively show that privacy in LLM-enabled health systems cannot depend only on de-identification. It requires privacy-by-design architecture, local or controlled deployment, access restrictions, secure audit trails, and continuous monitoring of how patient data are processed.

2.5 Trustworthy AI Governance, Fairness, and Accountability

Trustworthy AI governance is another major theme in related studies because LLM-enabled health information systems can affect clinical decisions, patient safety, equity, and institutional accountability. Meskó and Topol [36] argued that LLMs in healthcare require regulatory oversight because their general-purpose nature, conversational flexibility, and uncertain boundaries make them different from traditional medical software. Their work highlights the need for transparency, accountability, and post-deployment monitoring. Liu et al. [37] contributed to this debate by developing a translational perspective on clinical AI fairness, arguing that fairness should be assessed in relation to real clinical impact, patient groups, data quality, and implementation settings. Seyyed-Kalantari et al. [38] showed that AI algorithms applied to chest radiographs may underdiagnose underserved populations, demonstrating how algorithmic systems can reproduce inequities when data and model performance differ across groups. Vyas et al.

*Corresponding Author: Miracle A Atianashie

[39] similarly warned that race correction in clinical algorithms can embed social assumptions into medical decision-making, thereby producing unequal care. Norgeot et al. [40] proposed the MI-CLAIM checklist to improve transparency in clinical AI modelling, including reporting of data sources, model development, validation, and intended use. These studies are highly relevant to LLM-enabled health information systems because LLM outputs may appear persuasive even when biased, incomplete, or clinically unsafe. Therefore, governance should include model documentation, subgroup evaluation, human oversight, clinical accountability, incident reporting, and continuous performance monitoring.

3. METHODOLOGY

This review adopted a systematic review methodology to examine how Large Language Model–enabled Health Information Systems are being studied and implemented in relation to FHIR-based interoperability, clinical decision support, privacy protection, and trustworthy AI governance. The methodology was designed to ensure that the review process was transparent, reproducible, and sufficiently rigorous for synthesising evidence from health informatics, artificial intelligence, clinical decision support, digital health governance, and medical data interoperability literature. Since the topic cuts across technical, clinical, ethical, and regulatory domains, the review was not limited to purely clinical studies but also considered peer-reviewed technical articles, health informatics studies, AI governance papers, and interoperability-focused research. The methodology therefore sought to capture both the practical uses of LLMs in health information systems and the broader governance conditions required for their safe and responsible implementation.

3.1 Review Design

The study employed a systematic review design guided by the principles of structured evidence identification, screening, eligibility assessment, and thematic synthesis. This design was considered appropriate because the study aimed to organise and critically examine existing scholarly evidence on an emerging and multidisciplinary field. A systematic review approach allows the researcher to move beyond a general narrative discussion by applying a clear search strategy, explicit eligibility criteria, defined data extraction procedures, and a transparent synthesis process. In this study, the systematic review design was used to identify how LLM-enabled health information systems are currently represented in the literature, what benefits and risks are reported, and what gaps remain in relation to interoperability, clinical decision support, privacy, and trustworthy AI governance. The review was also interpretive in nature because the included literature covered diverse study designs, including empirical evaluations, system-development studies, conceptual frameworks, governance discussions, clinical AI assessment papers, and interoperability studies. Due to this diversity, a purely statistical meta-analysis was not suitable. Instead, the review adopted a thematic synthesis approach that enabled the findings to be grouped according to the major concerns of the study. This approach was useful because the evidence on LLM-enabled health information systems is still developing, and many studies remain experimental, prototype-based, or conceptual rather than fully implemented in real clinical environments.

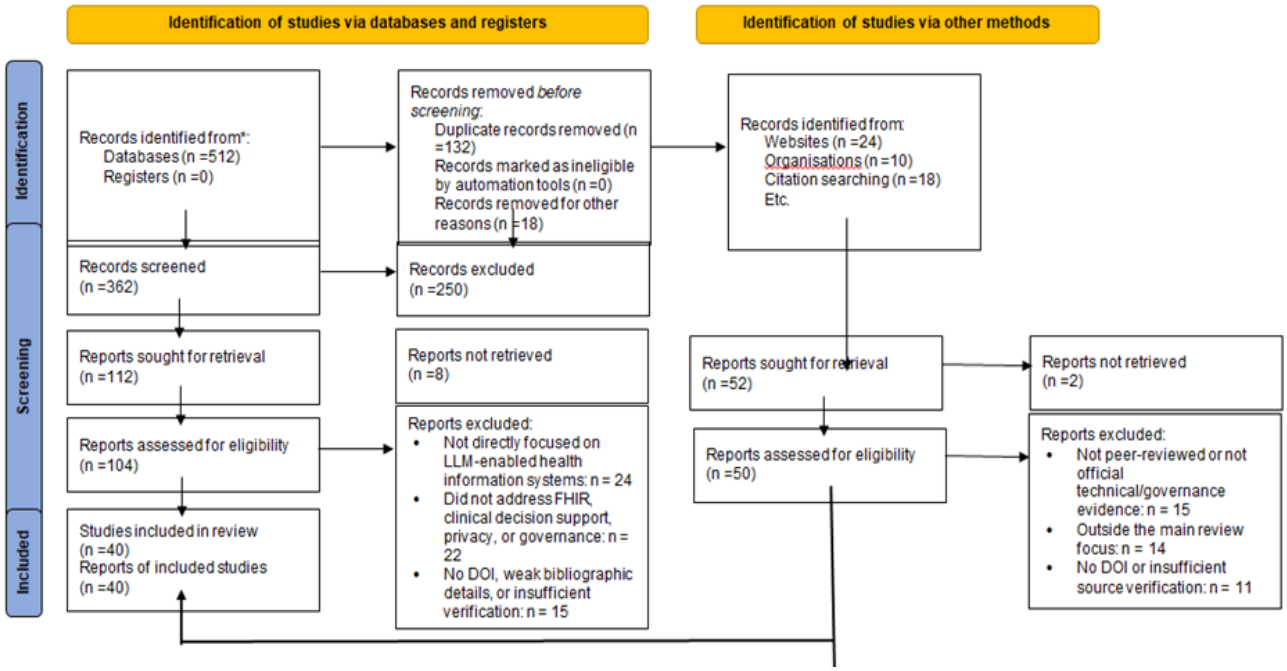
3.2 Scope of the Review

The scope of the review focused on studies that addressed Large Language Models within the context of health information systems, electronic health records, FHIR-based interoperability, clinical decision support, health data privacy, and trustworthy AI governance. The review considered LLMs as advanced AI systems capable of processing, generating, summarising, and reasoning over clinical language and health-related data. Within this scope, health information systems were understood broadly to include electronic health record systems, clinical information systems, patient-facing digital health platforms, decision support tools, documentation systems, health data exchange platforms, and AI-supported clinical workflow systems. The review was limited to literature published from 2020 onward because the rapid growth of LLMs, generative AI, and foundation models in healthcare became especially visible during this period. This timeframe also captures recent developments in FHIR implementation, AI reporting standards, privacy-preserving machine learning, clinical decision support evaluation, and trustworthy AI governance. The review did not focus on general artificial intelligence applications unless they had clear relevance to LLM-enabled systems, health information infrastructure, interoperability, privacy, or governance. This helped to maintain the focus of the review while still allowing the inclusion of relevant AI governance and health informatics literature.

3.3 Search Strategy

The search strategy was developed to capture literature across four major concept areas: Large Language Models, Health Information Systems, FHIR-based interoperability, and trustworthy clinical AI implementation. Search terms were combined using Boolean operators to ensure that the search retrieved studies related to both the technical and clinical dimensions of the topic. Terms relating to LLMs included “large language model,” “LLM,” “generative AI,” “foundation model,” “medical chatbot,” and “clinical language model.” These terms were combined with health information system terms such as “electronic health record,” “EHR,” “clinical information system,” “health information system,” “digital health,” and “clinical decision support.” Additional search terms were included to capture the specific focus areas of the review. For interoperability, the terms included “FHIR,” “HL7 FHIR,” “SMART on FHIR,” “health data exchange,” “interoperability,” and “clinical data standardisation.” For privacy, the search included terms such as “privacy-preserving,” “federated learning,” “data protection,” “health data privacy,” “de-identification,” and “secure clinical AI.” For trustworthy AI governance, the terms included “trustworthy AI,” “AI governance,” “algorithmic fairness,” “clinical AI safety,” “AI accountability,” “bias,” “transparency,” and “human oversight.” These terms were searched in different combinations to ensure comprehensive coverage of the literature relevant to the review objectives.

*Corresponding Author: Miracle A Atianashie



3.4 Information Sources

The review drew evidence from major scholarly databases and digital libraries relevant to medicine, health informatics, computer science, and digital health. These included PubMed, Scopus, Web of Science, IEEE Xplore, ACM Digital Library, ScienceDirect, SpringerLink, JMIR Publications, Nature Portfolio, BMJ, JAMA Network, and The Lancet Digital Health. These sources were selected because they contain peer-reviewed literature on clinical AI, health information systems, interoperability standards, clinical decision support, medical informatics, and digital health governance. In addition to journal databases, selected standards-based and policy-oriented sources were considered where they were directly relevant to the review. These included literature and guidance related to FHIR, SMART on FHIR, AI reporting standards, clinical AI evaluation, and responsible AI governance. However, the primary emphasis of the review remained on scholarly peer-reviewed literature. Preprints were considered only when they provided highly relevant technical evidence on emerging LLM-enabled health information systems, but such sources were treated cautiously because they may not have undergone formal peer review.

3.5 Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Studies published from 2020 onward	Studies published before 2020
Studies focused on LLMs, generative AI, or foundation models in healthcare	Studies on general AI without LLM relevance
Studies addressing EHRs, HIS, FHIR, interoperability, or clinical decision support	Studies unrelated to health information systems
Studies discussing privacy, security, governance, fairness, or trust in clinical AI	Opinion pieces without scholarly or technical contribution
Peer-reviewed articles, reviews, frameworks, and relevant technical studies	Non-healthcare LLM studies
English-language publications	Duplicate or inaccessible records

3.6 Study Selection Process

The study selection process followed a structured screening procedure to ensure that only relevant studies were included in the review. First, all retrieved records were examined by title to remove clearly irrelevant studies. The remaining studies were then screened by abstract to determine whether they addressed LLMs, health information systems, interoperability, clinical decision support, privacy, or AI governance. Studies that appeared relevant after abstract screening were subjected to full-text review to confirm their eligibility. During the full-text review stage, each article was assessed against the inclusion and exclusion criteria. Particular attention was given to whether the study had a direct connection to the review topic rather than only mentioning LLMs or healthcare in passing. Studies were retained when they provided evidence, analysis, framework development, evaluation, or governance insight relevant to LLM-enabled health information systems. Where articles overlapped in focus, the most relevant and methodologically useful studies were prioritised. This process helped to ensure that the final evidence base was both focused and sufficiently diverse to answer the review questions.

3.7 Data Extraction Procedure

Data extraction was conducted using a structured approach to ensure consistency across the included studies. For each selected

*Corresponding Author: Miracle A Atianashie

article, key information was extracted on the author, year of publication, study aim, study design, healthcare context, type of LLM or AI system examined, data source, health information system function, FHIR or interoperability component, clinical decision support role, privacy-related issues, governance considerations, main findings, and reported limitations. This allowed the review to compare studies across technical, clinical, and governance dimensions. The extraction process also focused on the maturity of implementation. Studies were examined to determine whether they were conceptual papers, prototype designs, retrospective evaluations, benchmark studies, simulation studies, or real-world clinical implementations. This distinction was important because many LLM-enabled health information system studies demonstrate technical feasibility but do not yet provide strong evidence of safety, usability, effectiveness, or sustainability in routine healthcare settings. Extracting this information made it possible to interpret the findings in relation to both promise and readiness for practical deployment.

3.8 Quality Appraisal

The quality appraisal process considered the methodological strength and relevance of each included study. Since the review included different types of literature, a flexible appraisal approach was adopted. Empirical studies were assessed based on clarity of objectives, appropriateness of design, quality of data sources, transparency of methods, validity of evaluation metrics, consideration of clinical context, and discussion of limitations. Technical studies were assessed for system description, reproducibility, validation strategy, error analysis, and relevance to health information system implementation. For studies addressing governance, privacy, ethics, or policy, appraisal focused on conceptual clarity, relevance to clinical AI, strength of argument, practical applicability, and alignment with healthcare safety and accountability concerns. This approach was necessary because governance and ethical studies are often not evaluated using the same criteria as experimental studies. Overall, the quality appraisal helped to determine the weight given to each study during synthesis. Studies with stronger methodological clarity, direct relevance, and practical health-system implications were given greater emphasis in the review findings.

3.9 Data Synthesis

The review used thematic synthesis to organise the findings because the included studies were diverse in design, scope, and outcome measures. Thematic synthesis was appropriate because the review was not attempting to calculate a pooled statistical effect but to understand how the literature explains the development, application, risks, and governance of LLM-enabled health information systems. The extracted findings were grouped according to recurring themes that aligned with the objectives of the review. Four major synthesis themes were developed: FHIR-based interoperability, LLM-enabled clinical decision support, privacy and security in health data processing, and trustworthy AI governance. Under each theme, the review examined the reported benefits, limitations, risks, and implementation requirements. This approach made it possible to identify both technical and non-technical issues affecting LLM-enabled health information systems. It also allowed the review to compare areas where the evidence is relatively strong with areas where research remains limited, fragmented, or mainly theoretical.

3.10 Ethical Considerations

Since this study was a systematic review of existing literature, it did not involve direct contact with human participants, patient recruitment, clinical intervention, or collection of primary health data. Therefore, formal ethical approval was not required. However, ethical considerations remained important because the topic itself concerns patient privacy, data governance, clinical safety, fairness, accountability, and responsible use of AI in healthcare. The review therefore treated ethical and governance concerns as substantive components of the analysis rather than as secondary issues. The review also maintained academic integrity by ensuring that selected studies were properly acknowledged and that interpretations were based on the content and relevance of the reviewed literature. Since LLM-enabled health information systems may influence clinical decision-making and patient trust, the methodology was designed to avoid overstating the maturity of the evidence. Studies were interpreted cautiously, especially where findings were based on prototypes, limited datasets, simulated environments, or benchmark evaluations rather than real-world clinical deployment.

3.11 Methodological Limitations

The methodology has some limitations. First, the field of LLM-enabled health information systems is rapidly evolving, meaning that new studies, systems, and governance frameworks may emerge after the review period. Second, the diversity of study designs limited the possibility of conducting a meta-analysis. Many studies used different datasets, evaluation metrics, clinical tasks, model types, and implementation contexts, making direct statistical comparison inappropriate. Third, some relevant technical studies may exist as preprints or industry reports, but these may not have undergone full peer review. Another limitation is that evidence on real-world implementation remains relatively limited. Many studies demonstrate model capability in controlled settings, but fewer examine long-term clinical safety, workflow integration, user acceptance, institutional governance, and post-deployment monitoring. As a result, the review methodology is better suited for mapping and synthesising the current state of evidence than for drawing final conclusions about clinical effectiveness. Despite these limitations, the methodology provides a rigorous and transparent basis for examining how LLM-enabled health information systems are developing across interoperability, clinical decision support, privacy, and trustworthy AI governance.

4. Results

This section presents the findings of the systematic review according to the four objectives of the study. The results are organised into four tables, each addressing a major dimension of Large Language Model–enabled Health Information Systems. The first table focuses on FHIR-based interoperability and health data standardisation. The second table presents studies on clinical decision support and workflow improvement. The third table examines privacy, security, and data protection concerns. The fourth table addresses trustworthy AI governance, fairness, accountability, and reporting standards. The results show that LLM-enabled Health Information Systems are developing rapidly, but the evidence remains uneven across technical, clinical, ethical, and governance domains. While several studies demonstrate that LLMs can support clinical language processing, summarisation, decision assistance, and interoperability, others emphasise the need

*Corresponding Author: Miracle A Atianashie

for human oversight, privacy-preserving design, regulatory control, and fairness evaluation. Therefore, the results should be understood as evidence of both opportunity and caution in the implementation of LLM-enabled systems in healthcare.

Table One: To Examine How LLMs Support FHIR-Based Interoperability and Health Data Standardisation

Author(s) and Year	Study Focus	Key Variable/Concept	LLM-Related Function	FHIR/Interoperability Contribution	Key Finding	Implication for Health Information Systems
Thirunavukarasu et al. (2023) [1]	LLMs in medicine	Clinical language processing	Interpretation and summarisation of clinical text	Supports conversion of unstructured clinical narratives into usable digital health data	LLMs can process clinical language and support the use of health data in clinical and administrative contexts.	LLMs can improve the structuring of clinical information for electronic health records and interoperability workflows.
Clusmann et al. (2023) [2]	Future of LLMs in medicine	Clinical workflow integration	Documentation support, information retrieval, and workflow assistance	Bridges the gap between clinical text and standardised health information systems	LLMs may improve clinical workflow, reduce documentation burden, and support digital health systems.	LLMs can strengthen interoperability when embedded into clinical documentation and EHR systems.
Moor et al. (2023) [3]	Foundation models for medical AI	Generalist medical AI capability	Multi-task medical reasoning and data interpretation	Enables flexible interpretation of diverse medical data across systems	Foundation models can support a wide range of medical AI tasks across different healthcare settings.	Generalist LLMs can support health data standardisation across diverse clinical domains and institutions.
Ayaz et al. (2021) [4]	FHIR implementation review	FHIR adoption and implementation	Not directly LLM-focused; provides interoperability foundation	FHIR improves healthcare data exchange through standardised resources and APIs	FHIR improves healthcare interoperability but requires strong technical and organisational implementation.	FHIR provides the standard structure that LLMs can support through automated mapping and data transformation.
Li et al. (2024) [5]	FHIR-GPT and interoperability	Clinical-to-FHIR data conversion	Mapping clinical information into FHIR-compatible resources	Directly supports automated generation of FHIR resources from clinical text	LLMs can help convert clinical information into FHIR resources.	LLMs can reduce manual coding burden and improve the speed of health data standardisation.
Mandl et al. (2020) [26]	SMART/HL7 Bulk FHIR API	Large-scale data access	Not LLM-focused; supports data extraction and population-level analytics	Bulk FHIR enables large-scale health data exchange across systems	Bulk FHIR supports population health data access for analytics, research, and public health monitoring.	LLMs can be used alongside Bulk FHIR to summarise, query, and interpret large interoperable datasets.

*Corresponding Author: Miracle A Atianashie

Jones et al. (2021) [27]	SMART/HL7 Bulk FHIR tools	Health data analytics	Analytics support through interoperable datasets	Bulk FHIR tools support structured access to healthcare data	Bulk FHIR tools are useful for health data analytics and population health management.	LLMs can enhance analytics by generating explanations, summaries, and insights from standardised FHIR datasets.
Griffin et al. (2022) [28]	Real-world FHIR applications	Clinical system integration	Potential support for app-based interpretation and documentation	FHIR apps support clinical integration but vary in maturity	FHIR-based applications support clinical integration, though implementation quality differs across settings.	LLMs may improve the usability of FHIR applications by supporting natural language interfaces and clinical decision support.
Kapitan et al. (2025) [29]	Open standards and interoperability	Open standards implementation	Possible support for mapping and harmonisation	Open standards require practical implementation support	Open standards are important but must be supported by governance, infrastructure, and institutional readiness.	LLMs can assist implementation by helping users interpret standards, map data elements, and identify inconsistencies.
Tsafnat et al. (2024) [30]	Open health data standards	Standard fragmentation and harmonisation	Cross-standard interpretation and semantic mapping	Multiple standards can support exchange but may also create fragmentation	Open health data standards create opportunities for interoperability but may lead to fragmentation if poorly coordinated.	LLMs can help harmonise different standards by mapping concepts across systems and improving semantic consistency.

Table One shows that LLMs support FHIR-based interoperability mainly through clinical language processing, data mapping, documentation support, semantic interpretation, and automated conversion of clinical information into structured formats. The evidence suggests that FHIR provides the technical foundation for standardised health data exchange, while LLMs can enhance the practical use of FHIR by helping to transform unstructured clinical narratives into machine-readable resources. Studies such as Li et al. (2024) [5] directly demonstrate the potential of LLMs to support clinical-to-FHIR conversion, while studies on SMART/HL7 Bulk FHIR show the importance of structured APIs for large-scale data access and analytics. However, the table also shows that interoperability is not achieved by technology alone. Successful implementation depends on governance, technical infrastructure, institutional readiness, data quality, and coordination among stakeholders. Therefore, LLMs should be understood as supportive technologies that can strengthen FHIR-based interoperability when combined with strong standards, clear governance, and reliable health information system infrastructure.

Table Two: Role of LLMs in Clinical Decision Support and Clinical Workflow Improvement

Author(s) and Year	Study Focus	Key Variable/Concept	LLM-Related Clinical Function	Workflow Contribution	Key Risk or Limitation	Key Finding	Implication for Health Information Systems
Singhal et al. (2023) [6]	Medical knowledge in LLMs	Clinical knowledge encoding	Medical question answering, clinical	Supports faster access to medical knowledge and	Performance on medical benchmarks may not fully	LLMs can encode and apply clinical knowledge	HIS can use LLMs to support knowledge

*Corresponding Author: Miracle A Atianashie

			reasoning, and knowledge retrieval	evidence-based explanations	represent real-world clinical safety	across different medical tasks.	retrieval and clinical information assistance, but outputs must be clinically verified.
Kung et al. (2023) [7]	ChatGPT and USMLE performance	Medical education and reasoning support	Answering examination-style clinical questions and explaining medical concepts	Supports medical learning, training, and educational decision support	Examination performance does not mean readiness for independent clinical decision-making	ChatGPT showed potential for AI-assisted medical education and clinical learning support.	LLM-enabled HIS may support clinician training and decision-support education, but should not replace expert judgement.
Lee et al. (2023) [8]	GPT-4 in medicine	Clinical promise and model limitation	Clinical conversation, reasoning support, and health information explanation	Improves access to clinical explanations and may support complex information interpretation	Risk of inaccurate, incomplete, or overconfident responses	GPT-4 has significant clinical promise but requires careful oversight and responsible use.	HIS should integrate LLMs with clinician supervision, validation controls, and clear boundaries of use.
Ayers et al. (2023) [9]	AI responses to patient questions	Patient communication and empathy	Generating patient-friendly health responses	May improve response drafting, patient education, and communication support	AI responses may still lack full clinical context or patient-specific judgement	AI responses were rated highly for quality and empathy compared with physician responses in public patient questions.	LLM-enabled HIS can support patient communication, but final messages should be reviewed by qualified professionals.
Omiye et al. (2024) [10]	Potentials and pitfalls of LLMs	Clinical usefulness and safety risk	Documentation, summarisation, information retrieval, and decision assistance	Can support multiple clinical and administrative tasks	May produce hallucinations, unsafe advice, bias, or misleading outputs	LLMs are useful in medicine but may generate unsafe or unreliable outputs without adequate safeguards.	HIS must include safety checks, human oversight, error monitoring, and accountability mechanisms.
Hager et al. (2024) [21]	LLMs in clinical decision-making	Clinical decision reliability	Assisting diagnostic reasoning and clinical decision support	Can support clinicians by organising information and suggesting possible reasoning pathways	LLMs may struggle with realistic clinical uncertainty, incomplete information, and complex decision contexts	LLMs require mitigation strategies before being used in clinical decision-making.	LLM-enabled HIS should apply validation, risk controls, clinician review, and restricted deployment in high-risk clinical tasks.

*Corresponding Author: Miracle A Atianashie

Van Veen et al. (2024) [22]	Clinical text summarisation	Documentation efficiency	Summarising clinical notes, reports, and patient records	Reduces documentation burden and improves access to key clinical information	Summaries may omit important clinical details or introduce inaccuracies	Adapted LLMs can produce strong clinical summaries and may outperform experts in some summarisation tasks.	HIS can use adapted LLMs to improve documentation workflows, but summaries should be reviewed for completeness and accuracy.
Tang et al. (2023) [23]	Medical evidence summarisation	Evidence synthesis	Summarising medical literature, guidelines, and clinical evidence	Helps clinicians access condensed evidence for decision support	Evidence summaries may contain factual errors or unsupported conclusions	LLMs can summarise medical evidence but require careful verification.	HIS can support evidence-based practice through LLM summarisation, provided that outputs are checked against trusted sources.
Reddy (2023) [24]	Translational value assessment	Clinical value and implementation readiness	Assessing usefulness, safety, fairness, usability, and workflow fit	Guides responsible movement from technical performance to clinical adoption	Technical accuracy alone is insufficient for healthcare implementation	LLMs must be assessed for clinical value, safety, usability, fairness, and implementation readiness.	HIS adoption should be based on translational value assessment, not model performance alone.
Cascella et al. (2023) [25]	ChatGPT in healthcare scenarios	Multi-purpose clinical and research support	Supporting education, documentation, research writing, and communication	Improves efficiency in selected clinical, academic, and administrative tasks	ChatGPT cannot replace professional expertise or clinical accountability	ChatGPT can support clinical and research-related healthcare tasks when used cautiously.	LLM-enabled HIS should position LLMs as assistive tools for workflow support rather than autonomous healthcare providers.

Table Two shows that LLMs have significant potential to support clinical decision support and workflow improvement in Health Information Systems. The reviewed studies demonstrate that LLMs can encode medical knowledge, answer clinical questions, summarise clinical text, generate patient-friendly responses, assist medical education, and support evidence synthesis [6]–[9], [22], [23]. These functions can improve the efficiency of clinical workflows by reducing documentation burden, improving access to clinical information, and supporting communication between clinicians and patients. However, the evidence also shows that clinical usefulness should not be interpreted as clinical autonomy. Studies by Omiye et al. [10], Hager et al. [21], and Reddy [24] emphasise that LLMs may produce unsafe, incomplete, biased, or overconfident outputs if they are not properly evaluated and supervised. Therefore, LLMs should be integrated into Health Information Systems as assistive technologies rather than independent clinical decision-makers. Their safest role is to support clinicians through summarisation, retrieval, explanation, documentation assistance, and workflow support while leaving final diagnosis, treatment decisions, and patient management to qualified healthcare professionals. Overall, LLM-enabled clinical decision support requires human oversight, validation mechanisms, safety monitoring, and clear accountability structures before routine clinical deployment.

*Corresponding Author: Miracle A Atianashie

Table Three: Privacy, Security, and Data Protection Issues in LLM-Enabled Health Information Systems

Author(s) and Year	Study Focus	Key Privacy/Security Variable	Main Data Protection Concern	Proposed/Relevant Safeguard	Key Finding	Implication for LLM-Enabled Health Information Systems
Wiest et al. (2024) [11]	Privacy-preserving LLMs	Privacy-preserving retrieval	Risk of exposing sensitive patient information during LLM-based search and retrieval	Structured retrieval, access control, and privacy safeguards	LLMs can support structured information retrieval when privacy safeguards are embedded into the system design.	LLM-enabled HIS should use controlled retrieval systems that limit unnecessary exposure of patient data.
Harrer (2023) [12]	Ethical use of LLMs in healthcare	Ethical responsibility and patient privacy	Possible misuse of patient data, unsafe outputs, and lack of transparency	Ethical governance, human oversight, and responsible AI principles	Ethical use of LLMs in healthcare requires privacy, safety, transparency, and accountability.	HIS using LLMs must be guided by clear ethical policies and responsible deployment standards.
Haltaufderheide and Ranisch (2024) [13]	Ethics of ChatGPT in healthcare	Accountability and confidentiality	Difficulty in assigning responsibility when LLM outputs affect patient care	Accountability frameworks, clinical supervision, and data protection protocols	Privacy and accountability are major ethical concerns in the use of ChatGPT and similar LLMs in healthcare.	LLM-enabled HIS should clearly define responsibility among clinicians, developers, institutions, and system users.
Ong et al. (2024) [15]	Ethical and regulatory challenges	Privacy, liability, and safety	Legal uncertainty, patient data exposure, and safety risks from unreliable AI responses	Regulation, clinical validation, and risk assessment	LLMs raise concerns around privacy, liability, clinical safety, and regulatory compliance.	Healthcare institutions must ensure that LLM systems comply with privacy laws and clinical safety standards before deployment.
Rieke et al. (2020) [31]	Federated learning in digital health	Distributed data protection	Direct sharing of patient data across institutions may increase privacy risks	Federated learning	Federated learning reduces the need for direct patient data sharing while enabling collaborative model development.	LLM-enabled HIS can adopt federated approaches to support learning across institutions without centralising sensitive data.
Kaissis et al. (2020) [32]	Privacy-preserving machine learning	Secure medical data processing	Medical data are highly sensitive and vulnerable to misuse when used for AI training	Secure computation, encryption, and privacy-preserving machine learning	Secure learning techniques can help protect medical data during AI development and deployment.	LLM systems should integrate encryption, secure computation, and privacy-aware model training methods.

*Corresponding Author: Miracle A Atianashie

Sheller et al. (2020) [33]	Federated learning in medicine	Multi-institutional collaboration	Hospitals may need to collaborate without transferring raw patient data	Federated learning and decentralised training	Institutions can collaborate on model development without directly sharing raw data.	LLM-enabled HIS can improve collaborative healthcare AI while maintaining institutional control over patient data.
Dayan et al. (2021) [34]	Federated learning for COVID-19 outcomes	Privacy-preserving clinical prediction	Clinical prediction models require large datasets, but data sharing may threaten confidentiality	Federated model training	Privacy-preserving models can support clinical prediction across institutions without exposing patient-level data.	LLM-based clinical decision support can benefit from privacy-preserving training methods when using multi-site health datasets.
Feretzakis et al. (2024) [35]	Privacy techniques in generative AI	Differential privacy, encryption, and secure design	Generative AI may memorise, reproduce, or expose sensitive health information	Differential privacy, encryption, anonymisation, and secure system architecture	LLMs require strong privacy techniques such as differential privacy, encryption, and secure-by-design implementation.	LLM-enabled HIS should be designed to prevent patient data leakage, unauthorised access, and model memorisation risks.
Meskó and Topol (2023) [36]	Regulatory oversight of LLMs	Governance and regulatory supervision	Unregulated LLMs may create risks for privacy, trust, safety, and clinical accountability	Strong regulatory oversight, audit systems, and clinical governance	Health-related LLMs require strong regulatory supervision to ensure safe and trustworthy use.	LLM-enabled HIS should be subject to continuous monitoring, auditing, regulatory review, and institutional governance.

Table Three shows that privacy, security, and data protection are central concerns in LLM-enabled Health Information Systems. The reviewed studies indicate that LLMs may improve clinical documentation, retrieval, prediction, and decision support, but they also introduce risks related to patient confidentiality, unauthorised data exposure, model memorisation, liability, and weak accountability. The evidence further shows that privacy-preserving approaches such as federated learning, encryption, differential privacy, secure computation, access control, and structured retrieval can reduce the risks associated with using sensitive health data. However, technical safeguards alone are not sufficient. Ethical governance, regulatory oversight, clinical supervision, institutional accountability, and continuous auditing are required to ensure safe implementation. Therefore, LLM-enabled HIS must be designed as secure, privacy-aware, and governance-driven systems that protect patient data while still supporting responsible clinical innovation.

Table Four: Trustworthy AI Governance, Fairness, Accountability, and Reporting Standards for LLM-Enabled Health Information Systems

Author(s) and Year	Study Focus	Key Governance Variable	Fairness/Accountability Concern	Reporting or Governance Standard	Key Finding	Implication for LLM-Enabled Health Information Systems
Omiye et al. (2023) [14]	Race-based medicine in LLMs	Algorithmic bias and race-based	LLMs may reproduce biased clinical assumptions if trained on historical or	Bias auditing, fairness evaluation,	LLMs may reproduce race-based clinical bias	LLM-enabled HIS must be tested for racial, ethnic, gender,

*Corresponding Author: Miracle A Atianashie

		reasoning	unbalanced medical data.	and inclusive dataset governance	and generate recommendations that reflect unequal assumptions in healthcare.	and socioeconomic bias before clinical deployment.
Benjamens et al. (2020) [16]	FDA-approved AI medical devices	Regulatory monitoring and post-market surveillance	AI systems may change over time or perform differently in real clinical environments.	Regulatory approval, monitoring, validation, and safety review	Clinical AI requires continuous regulatory monitoring to ensure safety, reliability, and effectiveness.	LLM-enabled HIS should undergo formal regulatory review, real-world monitoring, and periodic performance evaluation.
Liu et al. (2020) [17]	CONSORT-AI reporting guideline	Transparent reporting in AI clinical trials	Poor reporting can make AI clinical trial findings difficult to interpret, reproduce, or evaluate.	CONSORT-AI guideline	AI clinical trials require transparent reporting of intervention details, human-AI interaction, data handling, and evaluation methods.	LLM-enabled HIS used in clinical trials should report model design, data sources, user interaction, outcomes, and limitations clearly.
Cruz Rivera et al. (2020) [18]	SPIRIT-AI protocol guideline	Protocol transparency and trial design	AI study protocols may lack sufficient detail on model integration, data use, and clinical workflow.	SPIRIT-AI guideline	AI trial protocols need clear reporting of study design, AI intervention, and data requirements, and clinical implementation plans.	LLM-enabled HIS studies should describe trial design, model role, target users, safety procedures, and governance arrangements before implementation.
Vasey et al. (2022) [19]	DECIDE-AI guideline	Early-stage clinical evaluation	AI decision-support systems may be deployed before sufficient clinical testing.	DECIDE-AI guideline	Early-stage AI decision-support tools require structured clinical evaluation before wider adoption.	LLM-based decision-support systems should be tested in controlled clinical settings before full integration into HIS.
Collins et al. (2024) [20]	TRIPOD+AI statement	Transparent prediction model reporting	AI prediction models may lack clarity on data sources, model development, validation, and intended use.	TRIPOD+AI statement	AI prediction models need transparent reporting to support trust, reproducibility, and clinical interpretation.	LLM-enabled HIS that support prediction or risk assessment should report model inputs, outputs, validation results, and clinical limitations.
Liu et al. (2023) [37]	Clinical AI fairness	Real-world fairness assessment	AI systems may perform unequally across patient groups when used in actual healthcare settings.	Fairness testing, subgroup analysis, and external validation	Fairness must be evaluated in real clinical settings rather than relying only on laboratory performance.	LLM-enabled HIS should be evaluated across diverse populations, hospital contexts, languages, and patient groups.
Seyyed-Kalantari et al. (2021) [38]	Bias in chest radiograph AI	Health equity and diagnostic fairness	AI systems may underdiagnose or perform poorly for underserved patient populations.	Equity-focused evaluation and subgroup performance reporting	AI may underdiagnose underserved patient groups, creating risks of unequal healthcare outcomes.	LLM-enabled HIS must include equity checks to ensure that clinical recommendations do not disadvantage vulnerable groups.

*Corresponding Author: Miracle A Atianashie

Vyas et al. (2020) [39]	Race correction in clinical algorithms	Fairness, equity, and clinical justice	Race-based corrections in algorithms may reinforce unequal care and limit access to treatment.	Removal of harmful race correction, ethical review, and equity-based redesign	Clinical algorithms can reinforce unequal care when race is used uncritically as a correction factor.	LLM-enabled HIS should avoid uncritical use of race-based assumptions and promote clinically justified, equity-sensitive decision-making.
Norgeot et al. (2020) [40]	MI-CLAIM checklist	Transparent documentation of clinical AI models	Lack of documentation can reduce reproducibility, accountability, and trust in clinical AI systems.	MI-CLAIM checklist	Clinical AI modelling requires transparent documentation of data, model development, validation, and intended clinical use.	LLM-enabled HIS should document training data, model behaviour, validation procedures, limitations, and accountability structures.

Table Four shows that trustworthy AI governance in LLM-enabled Health Information Systems depends on fairness, transparency, accountability, regulatory oversight, and clear reporting standards. The reviewed studies indicate that LLMs and other clinical AI systems may reproduce bias, particularly where training data, clinical assumptions, or historical healthcare practices contain inequalities. Evidence from studies on race-based medicine, diagnostic bias, and race correction shows that AI systems can reinforce unequal care if fairness is not deliberately evaluated. The table also shows that governance frameworks such as CONSORT-AI, SPIRIT-AI, DECIDE-AI, TRIPOD+AI, and MI-CLAIM provide important reporting structures for improving transparency, reproducibility, and clinical accountability. These standards are especially relevant because LLM-enabled HIS may influence clinical documentation, diagnosis, decision support, prediction, and patient communication. Therefore, trustworthy implementation requires more than technical accuracy. It requires continuous bias auditing, subgroup evaluation, regulatory monitoring, transparent reporting, human oversight, and institutional responsibility. Overall, LLM-enabled HIS should be governed as socio-technical systems where fairness, safety, explainability, and accountability are built into the design, evaluation, deployment, and post-deployment monitoring process.

5. DISCUSSION

This section discusses the major findings of the review in relation to existing literature on Large Language Model–enabled Health Information Systems. The discussion is organised around the four central areas of the review: FHIR-based interoperability, clinical decision support, privacy protection, and trustworthy AI governance. Overall, the findings show that LLM-enabled Health Information Systems have strong transformative potential, but their implementation must be approached with technical caution, clinical responsibility, and governance maturity. The reviewed studies suggest that LLMs can improve clinical data processing, summarisation, documentation, decision support, and patient communication. However, the literature also shows that these benefits are accompanied by risks involving hallucination, unsafe clinical reasoning, data leakage, algorithmic bias, weak accountability, and insufficient real-world validation. Therefore, the discussion positions LLM-enabled Health Information Systems as socio-technical infrastructures rather than simple digital tools.

5.1 FHIR-Based Interoperability and Health Data Standardisation

The findings of this review show that FHIR-based interoperability is a critical foundation for the meaningful integration of LLMs into Health Information Systems. This agrees with Ayaz et al. [4], who observed that FHIR improves healthcare data exchange by supporting modular, standardised, and web-based health information sharing. The present review further shows that LLMs can support interoperability by helping to process unstructured clinical text and convert it into structured digital health resources. This finding is consistent with Li et al. [5], whose work on FHIR-GPT demonstrated that LLMs can enhance health interoperability by generating FHIR-compatible clinical resources. Similarly, Mandl et al. [26] and Jones et al. [27] showed that SMART/HL7 Bulk FHIR APIs can support large-scale health data access, population health analytics, and system-wide clinical data exchange. However, the findings also indicate that interoperability cannot depend on LLM generation alone. Although LLMs can interpret clinical language, their outputs may be incomplete, inconsistent, or semantically inaccurate if not validated. This supports Griffin et al. [28], who found that real-world FHIR applications vary in technical maturity and implementation quality. Kapitan et al. [29] and Tsafnat et al. [30] also emphasised that open standards require practical implementation ecosystems. Thus, LLM-enabled interoperability must combine generative intelligence with FHIR validation, terminology alignment, governance, and institutional readiness.

5.2 LLM-Enabled Clinical Decision Support and Workflow Improvement

The review findings indicate that LLMs can support clinical decision support and workflow improvement through summarisation, information retrieval, documentation assistance, patient communication, and medical evidence synthesis. This agrees with Singhal et al. [6], who showed that LLMs can encode substantial clinical knowledge, and Kung et al. [7], who reported that ChatGPT demonstrated potential for medical education and medical reasoning tasks. The results also align with Lee et al. [8], who argued that GPT-4 has important medical potential but must be used with awareness of its limitations. In the same direction,

*Corresponding Author: Miracle A Atianashie

Ayers et al. [9] found that AI-generated responses to patient questions were rated highly for quality and empathy, suggesting usefulness in patient communication. Nevertheless, the review also shows that LLM-enabled decision support must remain clinician-supervised. This is because clinical language fluency does not guarantee clinical safety. Omiye et al. [10] warned that LLMs may produce unsafe or misleading outputs, while Hager et al. [21] showed that LLMs still have limitations in realistic clinical decision-making. Van Veen et al. [22] and Tang et al. [23] demonstrated strong performance in summarisation, but both areas still require factual verification. Therefore, LLMs should be understood as assistive tools that strengthen clinical workflow rather than autonomous systems that replace professional judgement.

5.3 Privacy, Security, and Data Protection in LLM-Enabled Health Information Systems

The findings of this review show that privacy and security are among the most serious concerns in LLM-enabled Health Information Systems. This is because LLMs may process sensitive clinical information through prompts, embeddings, summaries, model outputs, and retrieval pipelines. Wiest et al. [11] demonstrated that privacy-preserving LLMs can support structured medical information retrieval, but their work also reinforces the need for careful data protection mechanisms. Harrer [12] similarly argued that the ethical use of LLMs in healthcare requires serious attention to privacy, safety, responsibility, and patient protection. Haltaufderheide and Ranisch [13] further confirmed that privacy and accountability are major ethical concerns in ChatGPT-related healthcare applications. The review findings also align with studies on privacy-preserving machine learning. Rieke et al. [31], Kaissis et al. [32], Sheller et al. [33], and Dayan et al. [34] showed that federated learning can support multi-institutional collaboration without direct sharing of raw patient data. Feretzakis et al. [35] extended this argument by identifying differential privacy, encryption, federated learning, and secure computation as important protections for generative AI and LLMs. However, technical protection alone is insufficient. Effective privacy protection also requires access control, audit trails, consent governance, local deployment decisions, and clear institutional responsibility. Therefore, privacy-by-design must be embedded into LLM-enabled systems from development to deployment.

5.4 Trustworthy AI Governance, Fairness, and Accountability

The review findings show that trustworthy AI governance is essential for the safe and responsible implementation of LLM-enabled Health Information Systems. This finding is strongly supported by Meskó and Topol [36], who argued that LLMs in healthcare require regulatory oversight because of their general-purpose nature, clinical influence, and unpredictable boundaries of use. The review also confirms that fairness must be treated as a central governance concern. Omiye et al. [14] demonstrated that LLMs may reproduce race-based medical assumptions, while Liu et al. [37] argued that clinical AI fairness must be assessed in real healthcare settings rather than only through abstract technical metrics. Seyyed-Kalantari et al. [38] also showed that AI systems may underdiagnose underserved populations, and Vyas et al. [39] warned that race correction in clinical algorithms can reinforce unequal care. The findings further suggest that trustworthy governance requires transparent reporting and structured evaluation. This agrees with CONSORT-AI [17], SPIRIT-AI [18], DECIDE-AI [19], TRIPOD+AI [20], and MI-CLAIM [40], which emphasise transparency in data sources, model design, clinical evaluation, human-AI interaction, and intended use. Therefore, LLM-enabled Health Information Systems should be governed across their full lifecycle. This includes design review, validation, bias assessment, clinical monitoring, user training, incident reporting, and continuous revision after deployment.

6. CONCLUSION

This review concludes that Large Language Model–enabled Health Information Systems represent an important but complex development in digital healthcare. The findings show that LLMs can improve the capacity of health information systems to process clinical narratives, support documentation, enhance information retrieval, assist clinical decision support, and strengthen patient communication. The review further establishes that FHIR-based interoperability is central to the safe and meaningful integration of LLMs into healthcare systems because it provides a standardised structure for exchanging clinical data across platforms. However, the evidence also shows that LLM-enabled systems are not yet mature enough to function as autonomous clinical decision-making tools. Their outputs may be affected by hallucination, incomplete reasoning, privacy risks, algorithmic bias, weak validation, and uncertain accountability. Therefore, their clinical value depends on careful implementation, human oversight, privacy-preserving design, transparent reporting, and trustworthy governance. LLM-enabled Health Information Systems should be understood as assistive clinical infrastructures that can improve healthcare delivery when they are properly validated, ethically governed, and aligned with clinical standards.

7. RECOMMENDATION

Health institutions should adopt LLM-enabled Health Information Systems cautiously and ensure that such systems are built on FHIR-compliant interoperability standards. Developers should integrate strong validation mechanisms to check the accuracy, completeness, and clinical safety of LLM-generated outputs. Clinicians should remain responsible for final decision-making, while LLMs should serve only as assistive tools for summarisation, documentation, retrieval, and decision support. Healthcare organisations should apply privacy-by-design principles, including access control, audit trails, encryption, controlled retrieval, and data minimisation. Regulators and institutional review bodies should establish clear rules for accountability, fairness, transparency, and post-deployment monitoring. Health systems should also evaluate LLM performance across different patient groups to reduce bias and promote equity. Continuous training should be provided for clinicians, administrators, and technical teams to ensure safe and responsible use.

8. CONTRIBUTION TO KNOWLEDGE

This review contributes to knowledge by bringing together evidence on Large Language Models, Health Information Systems, FHIR-based interoperability, clinical decision support, privacy protection, and trustworthy AI governance within a single scholarly synthesis. It advances understanding by showing that LLM-enabled health systems should not be assessed only by their language-generation ability but by their capacity to operate safely

*Corresponding Author: Miracle A Atianashie

within clinical, technical, ethical, and regulatory environments. The review also highlights FHIR as a critical foundation for responsible LLM integration because it supports structured, standardised, and interoperable clinical data exchange. In addition, the study contributes by clarifying that privacy, fairness, accountability, and lifecycle governance are not secondary concerns but essential conditions for trustworthy implementation. The review therefore provides a useful conceptual and practical foundation for researchers, developers, health managers, clinicians, and policymakers seeking to design, evaluate, or regulate LLM-enabled Health Information Systems.

Ethical Approval

Ethical approval was not required for this study because it was a systematic review based entirely on previously published scholarly literature. The study did not involve direct contact with human participants, patient recruitment, clinical intervention, animal subjects, or collection of primary health data. Nevertheless, ethical standards were observed by ensuring proper acknowledgement of all reviewed sources and by interpreting the included studies responsibly and accurately.

Consent to Participate

Consent to participate was not applicable to this study because no human participants were recruited or directly involved. The review relied only on published scholarly articles, technical reports, and relevant governance literature.

Consent for Publication

Consent for publication was not required because the manuscript does not contain personal, identifiable, confidential, or patient-level information. All authors have reviewed the manuscript and consent to its submission and publication.

Funding Information

The authors received no specific grant, sponsorship, or financial support from any funding agency in the public, commercial, or not-for-profit sectors for the conduct, authorship, or publication of this study.

Conflict of Interest Statement

The authors declare that they have no known financial, personal, professional, institutional, or intellectual conflicts of interest that could have influenced the conduct, interpretation, or publication of this manuscript.

Data Availability Statement

All data supporting the findings of this systematic review are available in the published articles and scholarly sources cited in the reference list. No new datasets were generated or analysed during the conduct of this review.

Author Contributions Statement

Miracle A. Atianashie conceptualised the study, developed the review focus, contributed to the methodology, guided the synthesis of the literature, drafted major sections of the manuscript, and reviewed the final paper. Chukwuma Chinaza Adaobi contributed to the literature review, screening of relevant studies, interpretation of findings, discussion development, and critical revision of the manuscript. Caleb Boakye Yiadom contributed to data extraction, organisation of reviewed studies, formatting of tables, reference preparation, manuscript editing, and final proofreading. All authors reviewed and approved the final version of the manuscript for submission and publication.

Acknowledgements

The authors acknowledge the scholarly contributions of the researchers whose published works formed the basis of this systematic review. The authors also appreciate the institutional and academic support received during the preparation of the manuscript.

Declaration of Originality

The authors declare that this manuscript is an original scholarly work. It has not been published previously and is not currently under consideration for publication elsewhere.

Clinical Trial Registration

Not applicable. This study was a systematic review and did not involve a clinical trial, clinical intervention, or patient recruitment.

Use of Human or Animal Subjects

Not applicable. This study did not involve human participants, animal subjects, biological samples, clinical experiments, or identifiable patient records.

Competing Interests

The authors declare that there are no competing interests associated with this manuscript.

REFERENCES

1. A. J. Thirunavukarasu, D. S. J. Ting, K. Elangovan, L. Gutierrez, T. F. Tan, and D. S. W. Ting, "Large language models in medicine," *Nature Medicine*, vol. 29, no. 8, pp. 1930–1940, 2023, doi: 10.1038/s41591-023-02448-8.
2. J. Clusmann, F. R. Kolbinger, H. S. Muti, Z. I. Carrero, J. N. Eckardt, N. G. Laleh, C. M. L. Löffler, S. C. Schwarzkopf, M. Unger, G. P. Veldhuizen, S. J. Wagner, and J. N. Kather, "The future landscape of large language models in medicine," *Communications Medicine*, vol. 3, no. 1, Art. no. 141, 2023, doi: 10.1038/s43856-023-00370-1.
3. M. Moor, O. Banerjee, Z. S. H. Abad, H. M. Krumholz, J. Leskovec, E. J. Topol, and P. Rajpurkar, "Foundation models for generalist medical artificial intelligence," *Nature*, vol. 616, no. 7956, pp. 259–265, 2023, doi: 10.1038/s41586-023-05881-4.
4. M. Ayaz, M. F. Pasha, M. Y. Alzahrani, R. Budiarto, and D. Stiawan, "The Fast Health Interoperability Resources (FHIR) standard: Systematic literature review of implementations, applications, challenges and opportunities," *JMIR Medical Informatics*, vol. 9, no. 7, e21929, 2021, doi: 10.2196/21929.
5. Y. Li, H. Wang, H. Z. Yerebakan, Y. Shinagawa, and Y. Luo, "FHIR-GPT enhances health interoperability with large language models," *NEJM AI*, 2024, doi: 10.1056/AIcs2300301.
6. K. Singhal et al., "Large language models encode clinical knowledge," *Nature*, vol. 620, no. 7972, pp. 172–180, 2023, doi: 10.1038/s41586-023-06291-2.
7. T. H. Kung et al., "Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models," *PLOS Digital Health*, vol.

*Corresponding Author: Miracle A Atianashie

2, no. 2, e0000198, 2023, doi: 10.1371/journal.pdig.0000198.
8. P. Lee, S. Bubeck, and J. Petro, "Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine," *New England Journal of Medicine*, vol. 388, no. 13, pp. 1233–1239, 2023, doi: 10.1056/NEJMsr2214184.
9. J. W. Ayers et al., "Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum," *JAMA Internal Medicine*, vol. 183, no. 6, pp. 589–596, 2023, doi: 10.1001/jamainternmed.2023.1838.
10. J. A. Omiye, H. Gui, S. J. Rezaei, J. Zou, and R. Daneshjou, "Large language models in medicine: The potentials and pitfalls," *Annals of Internal Medicine*, vol. 177, no. 2, pp. 210–220, 2024, doi: 10.7326/M23-2772.
11. I. C. Wiest et al., "Privacy-preserving large language models for structured medical information retrieval," *npj Digital Medicine*, vol. 7, Art. no. 257, 2024, doi: 10.1038/s41746-024-01233-2.
12. S. Harrer, "Attention is not all you need: The complicated case of ethically using large language models in healthcare and medicine," *eBioMedicine*, vol. 90, Art. no. 104512, 2023, doi: 10.1016/j.ebiom.2023.104512.
13. J. Haltaufderheide and R. Ranisch, "The ethics of ChatGPT in medicine and healthcare: A systematic review on Large Language Models (LLMs)," *npj Digital Medicine*, vol. 7, Art. no. 183, 2024, doi: 10.1038/s41746-024-01157-x.
14. J. A. Omiye, J. C. Lester, S. Spichak, V. Rotemberg, and R. Daneshjou, "Large language models propagate race-based medicine," *npj Digital Medicine*, vol. 6, Art. no. 195, 2023, doi: 10.1038/s41746-023-00939-z.
15. J. C. L. Ong et al., "Ethical and regulatory challenges of large language models in medicine," *The Lancet Digital Health*, vol. 6, no. 6, pp. e428–e432, 2024, doi: 10.1016/S2589-7500(24)00061-X.
16. S. Benjamens, P. Dhunnoo, and B. Meskó, "The state of artificial intelligence-based FDA-approved medical devices and algorithms: An online database," *npj Digital Medicine*, vol. 3, Art. no. 118, 2020, doi: 10.1038/s41746-020-00324-0.
17. X. Liu et al., "Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension," *Nature Medicine*, vol. 26, no. 9, pp. 1364–1374, 2020, doi: 10.1038/s41591-020-1034-x.
18. S. Cruz Rivera et al., "Guidelines for clinical trial protocols for interventions involving artificial intelligence: The SPIRIT-AI extension," *Nature Medicine*, vol. 26, no. 9, pp. 1351–1363, 2020, doi: 10.1038/s41591-020-1037-7.
19. B. Vasey et al., "Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI," *Nature Medicine*, vol. 28, no. 5, pp. 924–933, 2022, doi: 10.1038/s41591-022-01772-9.
20. G. S. Collins et al., "TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods," *BMJ*, vol. 385, e078378, 2024, doi: 10.1136/bmj-2023-078378.
21. P. Hager, F. Jungmann, R. Holland, K. Bhagat, I. Hubrecht, M. Knauer, J. Vielhauer, M. Makowski, R. Braren, G. Kaissis, and D. Rueckert, "Evaluation and mitigation of the limitations of large language models in clinical decision-making," *Nature Medicine*, vol. 30, no. 9, pp. 2613–2622, 2024, doi: 10.1038/s41591-024-03097-1.
22. D. Van Veen, C. Van Uden, L. Blankemeier, J.-B. Delbrouck, A. Aali, C. Bluethgen, A. Pareek, M. Polacin, E. P. Reis, A. Seehofnerová, N. Rohatgi, P. Hosamani, W. Collins, N. Ahuja, C. P. Langlotz, J. Hom, S. Gatidis, J. Pauly, and A. S. Chaudhari, "Adapted large language models can outperform medical experts in clinical text summarization," *Nature Medicine*, vol. 30, no. 4, pp. 1134–1142, 2024, doi: 10.1038/s41591-024-02855-5.
23. L. Tang, Z. Sun, B. Idnay, J. G. Nestor, A. Soroush, P. A. Elias, Z. Xu, Y. Ding, G. Durrett, J. F. Rousseau, C. Weng, and Y. Peng, "Evaluating large language models on medical evidence summarization," *npj Digital Medicine*, vol. 6, Art. no. 158, 2023, doi: 10.1038/s41746-023-00896-7.
24. S. Reddy, "Evaluating large language models for use in healthcare: A framework for translational value assessment," *Informatics in Medicine Unlocked*, vol. 41, Art. no. 101304, 2023, doi: 10.1016/j.imu.2023.101304.
25. M. Cascella, J. Montomoli, V. Bellini, and E. Bignami, "Evaluating the feasibility of ChatGPT in healthcare: An analysis of multiple clinical and research scenarios," *Journal of Medical Systems*, vol. 47, Art. no. 33, 2023, doi: 10.1007/s10916-023-01925-4.
26. K. D. Mandl, D. Gottlieb, J. C. Mandel, J. M. Kohane, R. Ramoni, D. C. Mandl, and J. R. Mandl, "Push button population health: The SMART/HL7 FHIR Bulk Data Access Application Programming Interface," *npj Digital Medicine*, vol. 3, Art. no. 151, 2020, doi: 10.1038/s41746-020-00358-4.
27. J. Jones, D. Gottlieb, J. C. Mandel, V. Ignatov, A. Ellis, W. Kubick, and K. D. Mandl, "A landscape survey of planned SMART/HL7 bulk FHIR data access API implementations and tools," *Journal of the American Medical Informatics Association*, vol. 28, no. 6, pp. 1284–1287, 2021, doi: 10.1093/jamia/ocab028.
28. A. C. Griffin, A. Kukhareva, C. G. S. Del Fiol, and V. L. Patel, "Clinical, technical, and implementation characteristics of real-world health applications using FHIR," *JAMIA Open*, vol. 5, no. 4, Art. no. ooac077, 2022, doi: 10.1093/jamiaopen/ooac077.
29. D. Kapitan, F. Heddema, A. Dekker, M. Sieswerda, B.-J. Verhoeff, and M. Berg, "Data interoperability in context: The importance of open-source implementations when choosing open standards," *Journal of Medical Internet Research*, vol. 27, Art. no. e66616, 2025, doi: 10.2196/66616.
30. G. Tsafnat, R. Dunscombe, D. Gabriel, G. Grieve, and C. Reich, "Converge or collide? Making sense of a plethora of open data standards in health care," *Journal of*

*Corresponding Author: Miracle A Atianashie

Medical Internet Research, vol. 26, Art. no. e55779, 2024, doi: 10.2196/55779.

31. N. Rieke et al., "The future of digital health with federated learning," *npj Digital Medicine*, vol. 3, Art. no. 119, 2020, doi: 10.1038/s41746-020-00323-1.
32. G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nature Machine Intelligence*, vol. 2, no. 6, pp. 305–311, 2020, doi: 10.1038/s42256-020-0186-1.
33. M. J. Sheller et al., "Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data," *Scientific Reports*, vol. 10, Art. no. 12598, 2020, doi: 10.1038/s41598-020-69250-1.
34. I. Dayan et al., "Federated learning for predicting clinical outcomes in patients with COVID-19," *Nature Medicine*, vol. 27, no. 10, pp. 1735–1743, 2021, doi: 10.1038/s41591-021-01506-3.
35. G. Feretzakis, P. Kalles, A. Verykios, and D. Kalles, "Privacy-preserving techniques in generative AI and large language models: A narrative review," *Information*, vol. 15, no. 11, Art. no. 697, 2024, doi: 10.3390/info15110697.
36. B. Meskó and E. J. Topol, "The imperative for regulatory oversight of large language models in healthcare," *npj Digital Medicine*, vol. 6, Art. no. 120, 2023, doi: 10.1038/s41746-023-00873-0.
37. M. Liu et al., "A translational perspective towards clinical AI fairness," *npj Digital Medicine*, vol. 6, Art. no. 172, 2023, doi: 10.1038/s41746-023-00918-4.
38. L. Seyyed-Kalantari, H. Zhang, M. B. A. McDermott, I. Y. Chen, and M. Ghassemi, "Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations," *Nature Medicine*, vol. 27, no. 12, pp. 2176–2182, 2021, doi: 10.1038/s41591-021-01595-0.
39. D. A. Vyas, L. G. Eisenstein, and D. S. Jones, "Hidden in plain sight—Reconsidering the use of race correction in clinical algorithms," *New England Journal of Medicine*, vol. 383, no. 9, pp. 874–882, 2020, doi: 10.1056/NEJMms2004740.
40. B. Norgeot et al., "Minimum information about clinical artificial intelligence modeling: The MI-CLAIM checklist," *Nature Medicine*, vol. 26, no. 9, pp. 1320–1324, 2020, doi: 10.1038/s41591-020-1041-y.

*Corresponding Author: Miracle A Atianashie