

Integrating LLMs into IoT-Driven Smart Healthcare Systems: A Systematic Literature Review and Future Agenda

Prithvi Raju Mekala ^{*}, Yonas Kassa ^{*} and Sushma Mishra 

Computer and Information Systems, Robert Morris University, Pittsburgh, PA 15108, USA; mishra@rmu.edu

^{*} Correspondence: pxmst124@mail.rmu.edu (P.R.M.); kassa@rmu.edu (Y.K.)

Abstract

The convergence of Large Language Models (LLMs) with the Internet of Things (IoT) is driving a transformative shift toward a continuous, context-aware smart healthcare ecosystem. Due to its novelty, existing research in this domain remains fragmented, leaving a critical gap in unified frameworks that synthesize domain applications, functional AI deployment roles, network architectures, and security boundaries. Following PRISMA 2020 guidelines, this paper presents a systematic literature review and quantitative analysis evaluating a selected corpus of 61 peer-reviewed and 14 preprint papers in this domain. Methodologically, we assess a novel hybrid article discovery strategy, finding that an AI-powered prompt-based literature search strategy achieves higher precision than traditional keyword-based Boolean queries (86% vs. 42%) on the evaluated search sample, which may reduce screening workloads. We found that the major limitation of AI-based literature search is non-determinism, which is also an inherent property of LLM-powered applications. To address this, we propose methodological guidelines for using an AI-assisted hybrid literature search strategy. Based on the selected literature, we establish a multi-layer taxonomy organizing the IoT-LLM advances in the healthcare domain across four pillars: application domain, LLM role, IoT device type, and architectural deployment pattern. Quantitative synthesis reveals a heavy research concentration in remote patient monitoring and personal health management (representing 59% of the corpus combined), primarily driven by the data accessibility of wearable sensors (64%). Cross-tabulation uncovers a distinct capability–constraint spectrum: cloud-based deployments lean on heavyweight state-of-the-art models (mainly GPT-family models) for complex semantic reasoning, whereas edge, federated, and blockchain-based hybrid systems leverage localized models (BERT and LLaMA families). Patient data privacy and reduced communication overhead were among the main reasons for choosing localized models. Crucially, our assessment reveals a pervasive neglect of LLM-specific vulnerabilities such as prompt injection and jailbreak attacks and a tendency to treat regulatory frameworks (e.g., HIPAA, GDPR) as design features rather than empirically validated compliance metrics. Finally, we propose an actionable future research agenda prioritizing multi-device system orchestration, emergency care integration, privacy-preserving LLMs, and deployment-scale clinical validation.



Academic Editors: Mahmoud Abouyoussef and Mohamed Ibrahim

Received: 17 July 2026

Revised: 2 September 2026

Accepted: 3 September 2026

Published: 8 September 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: Large Language Models (LLMs); Internet of Things (IoT); Internet of Medical Things (IoMT); smart healthcare; remote patient monitoring; systematic literature review; edge AI; data privacy; system orchestration

1. Introduction

The convergence of Large Language Models (LLMs) with the Internet of Things (IoT) in healthcare represents a transformative shift in modern healthcare systems. IoT enables continuous, real-time collection of physiological and environmental data through interconnected sensors and smart medical devices, while LLMs and other artificial intelligence techniques provide advanced capabilities for interpreting complex, multimodal clinical data. Together, they enable intelligent, data-driven healthcare ecosystems that support remote monitoring, early diagnosis, and personalized treatment.

Remote patient monitoring (RPM), one of the most prominent applications of IoT in healthcare, allows clinicians to monitor patients outside traditional clinical settings using wearable devices and sensors [1]. These systems improve patient outcomes by enabling early detection of health deterioration, reducing hospital visits, and supporting continuous care. IoT technologies further enhance healthcare efficiency by enabling real-time data collection, automation of routine processes, and improved patient engagement [2].

The integration of AI, particularly advanced models such as LLMs, extends these capabilities by enabling contextual understanding of medical data, natural language interaction, and intelligent decision support. As healthcare systems face increasing pressure from aging populations and rising chronic disease cases, such intelligent and scalable solutions are becoming essential.

The existing literature highlights the growing role of AI and IoT integration in healthcare. AI-enabled RPM systems have demonstrated significant improvements in early disease detection, personalized monitoring, and predictive analytics, particularly through machine learning and deep learning techniques [1]. Similarly, IoT-based healthcare systems have been widely adopted for applications such as chronic disease management, emergency monitoring, and personalized treatment strategies, supported by continuous data collection from wearable sensors [2].

Recent studies also emphasize the architectural evolution of AIoT healthcare systems, including cloud-based, edge-based, and hybrid models. For example, phase-driven AIoT architectures for remote patient monitoring integrate IoT for data acquisition and AI for real-time analytics and decision-making, enabling scalable and structured healthcare solutions [3]. Additionally, the emergence of edge and fog computing has addressed latency and reliability challenges by enabling localized data processing closer to patients [4].

Although the integration of AI and IoT has been extensively explored, several critical gaps remain. First, many existing systems emphasize data collection and transmission, with limited focus on real-time analytics and predictive intelligence [3]. Second, current research predominantly focuses on conventional AI models (e.g., CNNs, random forests) rather than exploring the emerging role of LLMs in healthcare IoT ecosystems.

Furthermore, challenges such as interoperability, data privacy, scalability, and seamless integration across heterogeneous IoT devices persist across the literature [2]. There is limited systematic analysis that jointly examines AI methodologies (LLMs in particular), IoT integration strategies, application domains, architectural patterns, and security implications in a unified framework.

These gaps highlight the need for a comprehensive synthesis that not only categorizes existing research but also analyzes cross-dimensional relationships and emerging trends in LLM–IoT healthcare systems.

To address this gap, this study presents a systematic review and quantitative analysis of the integration of Large Language Models and IoT in healthcare. The aim of our study is to develop a structured taxonomy that provides a comprehensive understanding of how LLMs are being integrated with IoT systems across different healthcare applications and

architectural designs. In this paper, we also studied how the state-of-the-art literature addresses security, privacy, ethics, and regulatory compliance.

This study focuses on the following research questions:

- RQ1: Which healthcare domains are being targeted and supported by IoT-LLM systems?
- RQ2: Which LLMs are used within IoT-based healthcare systems and for what roles?
- RQ3: How are LLMs integrated with IoT systems in healthcare environments?
- RQ4: How are security, privacy, and ethical challenges addressed in LLM-IoT healthcare systems?

As AI is increasingly being adopted across different fields and research processes, emerging AI-assisted scholarly search tools such as Scholar Labs offer a new approach to literature discovery; therefore, one contribution of this review is to evaluate how such an AI-based search method compares with the established keyword-based Google Scholar approach, and how to integrate such emerging systems with the established methodologies.

This review makes five contributions. (1) We present the first comparison of the emerging Scholar Labs and traditional Google Scholar keyword-search-based methods for literature survey. (2) We provide a comprehensive taxonomy of 75 LLM-IoT studies in healthcare systems, classifying them by healthcare application domain, IoT devices, LLM role/model, and architecture. (3) We synthesize cross-dimensional insights via quantitative cross-tabulation of categories. (4) We critically assess security, privacy, and ethical challenges, highlighting common strategies and neglected issues while integrating LLMs for healthcare. (5) We identify key research gaps (e.g., emergency care applications, autonomous orchestration, robust real-world evaluation) and outline future directions. Together, these offer the **first structured overview of LLM-IoT convergence in healthcare**, guiding researchers and practitioners toward effective, secure, and privacy-oriented development of healthcare systems.

The rest of this paper is organized as follows. Section 2 presents the related work. Section 3 describes the review methodology, including the literature search, study selection, data extraction, and comparison of the AI-assisted and traditional search approaches. Section 4 presents the proposed taxonomy and classification framework. Section 5 reports the results and analysis addressing the research questions. Section 6 discusses the main findings, system design trade-offs, research gaps, and limitations. Finally, Section 7 concludes this paper and summarizes the key contributions and future research directions.

2. Related Work

Prior surveys on AI-enabled IoT healthcare mainly synthesized connected medical devices, AIoMT architectures, and deployment challenges in smart healthcare, but they were still centered on traditional AI/ML rather than LLM-driven reasoning or orchestration [5,6]. Subsequent surveys on LLMs in healthcare mapped broader clinical and biomedical applications, evaluation practices, and interoperable data workflows, yet they generally lacked an explicit healthcare IoT device and architecture-centric perspective [7–9]. More recent reviews have started to examine parts of the LLM-IoMT intersection, for example in wearable-sensor interpretation, cloud-enabled medical imaging with IoMT, and edge cardiovascular intelligence, but these studies remain tied to particular modalities or application domains rather than offering a unified healthcare-wide synthesis [10–12]. Accordingly, the main gap is no longer the total absence of related work, but the lack of a focused systematic review that jointly analyzes healthcare domains, IoT device ecosystems, LLM functional roles, deployment architectures, and trustworthiness concerns in one framework. This paper fills this gap by providing a healthcare-focused systematic synthesis of LLM-IoMT studies through a multi-dimensional taxonomy, cross-tabulation

across application–device–role–architecture dimensions, and a dedicated review of security, privacy, ethics, and LLM-specific risks.

3. Methods

3.1. Search Strategy

The search strategy was designed to identify all relevant research articles addressing the integration of LLM and IoT technologies in healthcare applications. Two distinct search methodologies were employed and compared to evaluate their respective effectiveness in systematic literature reviews. The review protocol was registered and published at PROSPERO (registration number CRD420261485115). Literature search was initially conducted on 12 March 2026. Multiple search iterations and iterative quality assessment were performed to refine the Scholar Labs prompt and ensure comparability with the Google Scholar search. The final dataset of search results included in this review was obtained on 25 March 2026. The literature for related work was finalized on 20 April 2026.

3.1.1. Database Selection

Google Scholar was selected as the primary database for both search approaches due to its comprehensive coverage of scholarly literature across multiple publishers, including IEEE, Elsevier (ScienceDirect), MDPI, Springer, Nature, ACM, and Frontiers [13]. Google Scholar provides broader coverage of grey literature and conference proceedings compared to subscription databases, making it particularly valuable for emerging technology domains [14].

3.1.2. Method 1: Scholar Labs (AI-Powered Prompt-Based Search)

Scholar Labs represents an innovative and emerging approach to literature search that leverages artificial intelligence to interpret natural language prompts. Rather than requiring users to construct complex Boolean queries, this method allows researchers to describe their information needs in conversational language.

Search Prompt Construction:

The search prompt encoded the core topical scope and the exclusion of review/survey articles in natural language format. Specifically, the following prompt was used:

Prompt: “Find research papers on integrating IoT and LLMs focused on healthcare domain. Exclude review and survey articles.”

Operational Mechanism:

Scholar Labs operates by scanning articles from its index and applying AI-based filtering based on the prompt criteria. Upon submission of the search prompt, the system processes the natural language requirements and returns relevant articles progressively. The system displays results in batches of 10 papers per page, with the user able to navigate through additional results by clicking the “More” button. It is observed that Scholar Labs’ maximum screening capacity for a single prompt is 300 papers because during preliminary testing with multiple prompts, the largest result index observed was approximately 300 articles.

Search Parameters:

The following parameters were applied to the Scholar Labs search through the prompt:

- Language: English
- Article Type: Research articles

3.1.3. Method 2: Traditional Google Scholar (Keyword-Based Boolean Search)

Traditional keyword-based Boolean search represents the conventional approach to systematic literature reviews, requiring researchers to manually construct Boolean queries using AND, OR, and NOT operators to specify inclusion and exclusion criteria [15].

Search Query Construction:

The Boolean search query was constructed following established guidelines for systematic literature reviews [15].

Search Query: “LLM” AND “IoT” AND “healthcare” -intitle:“Review” -intitle:“Survey”

Query Design Rationale:

The Boolean query was designed with three distinct components:

1. Primary Term Group: “IoT” which represents IoT-related technologies (IoT, IoMT, AIoT, etc).
2. Technology Group: “LLM” which represents Large Language Model technologies.
3. Application Group: “healthcare” which represents healthcare applications (healthcare, health monitoring, wearable sensors, smart healthcare, etc.).

Search Parameters:

The following parameters were applied to the traditional Google Scholar search:

- Language: English.
- Sort by: Relevance.
- Citations: Included.

The Boolean query was intentionally designed to prioritize relevance while maintaining sufficient coverage. During preliminary testing, alternative terms and synonyms, including “Large Language Model,” “Generative AI,” “IoMT,” and “Internet of Medical Things,” were evaluated. These broader combinations increased the retrieval of false-positive and out-of-scope studies that barely mention the keywords. The core term “IoT” was retained because it is broader and can capture studies using related variants such as IoMT, HIoT, and AIoT, while searches restricted to these narrower terms may not retrieve studies described only as IoT. Similarly, “LLM” and “healthcare” provided the most relevant results for the specific scope of this review. Additionally, using the same core concepts as the Scholar Labs prompt also improved comparability between the AI-powered and traditional search approaches.

3.2. Inclusion and Exclusion Criteria

To ensure methodological rigor and relevance of selected studies, predefined inclusion and exclusion criteria were established following PRISMA 2020 guidelines (Table S1) [16].

3.2.1. Inclusion Criteria

The predefined inclusion criteria used to determine study eligibility for this review are summarized in Table 1.

Table 1. Studies were included if they met all of the following 4 criteria.

Criterion	Description
IC1: Domain	Studies addressing LLM and IoT/IoMT integration in the healthcare domain.
IC2: Originality	Research articles with original technical contributions.
IC3: Validation	Studies presenting experimental validation with novel methodologies or frameworks.
IC4: Language	Written in English.

3.2.2. Exclusion Criteria

The predefined exclusion criteria used to remove studies that did not meet the requirements of this review are summarized in Table 2.

Table 2. Studies were excluded if they met any of the following 6 criteria.

Criterion	Description
EC1: Publication Type	Systematic literature reviews, review articles, survey papers.
EC2: Document Type	Position papers, conference abstracts, editorials, opinion pieces.
EC3: Originality	Duplicate publications or redundant papers.
EC4: Methodology	Papers lacking experimental validation.
EC5: Scope	Studies were excluded when they focused solely on conventional AI/ML or IoT without substantive use of an LLM or qualifying language-model component, or when the language-model component was not actually used in the proposed or evaluated healthcare application.
EC6: Domain	Papers outside the healthcare domain.

For this review, the term LLM primarily refers to mainstream large generative or foundation language models capable of general-purpose language understanding, generation, reasoning, or conversational interaction, including GPT-family, LLaMA-family, DeepSeek, Gemini, and comparable models. BERT, DistilBERT, MiniLM, and related encoder-oriented models are not treated as equivalent to full-capability generative LLMs. However, compact or specialized transformer language models were retained when the reviewed study explicitly used them as a functional language-model component for tasks such as information extraction, summarization, classification, anomaly interpretation, or other language-oriented processing within an IoT/IoMT healthcare workflow. T5-family models were similarly considered according to their functional use and implementation rather than assumed to represent full-capability LLMs solely by model family. Conventional ML/DL models without such language-model functionality were excluded.

3.2.3. Application of Criteria

The inclusion and exclusion criteria were applied differently by each search method:

In Scholar Labs, the AI-powered system interpreted and applied the topical scope and the criterion excluding review and survey articles through natural language processing, attempting to filter results according to the prompt specifications. In traditional Google Scholar, Boolean operators were used to partially implement criteria; however, manual screening was required for final classification due to limitations in keyword-based filtering.

3.3. Selection Process

The selection process followed a structured approach adapted from the PRISMA 2020 flow diagram methodology [16]. The entire PRISMA process is displayed in Figure 1.

3.3.1. Scholar Labs Selection Process

The Scholar Labs search was conducted using the constructed natural language prompt. The system returned results progressively, displaying a total of 36 papers that were filtered by AI from an observed candidate set of ~250 articles in the system index. For each result, the user clicked to access full article information and determine relevance based on the inclusion criteria.

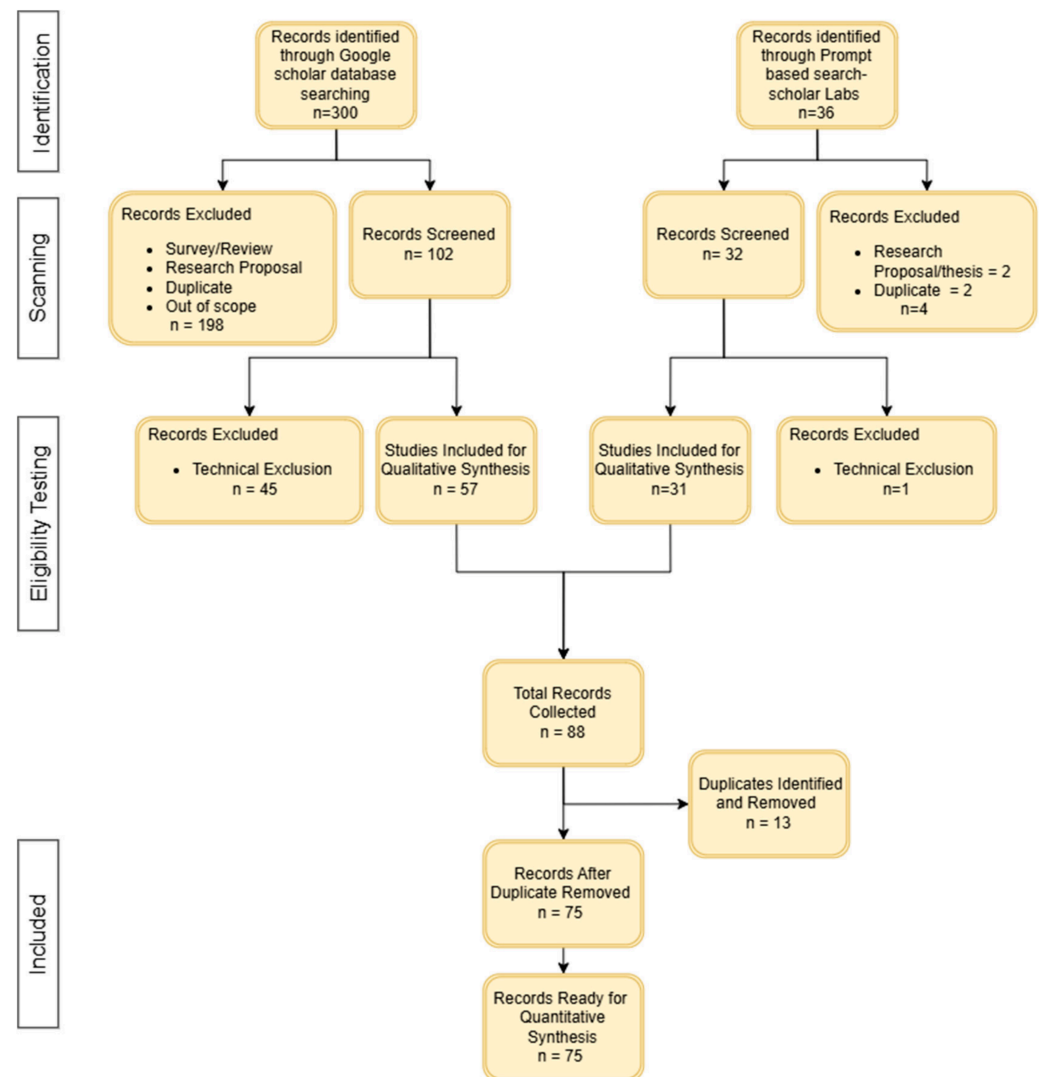


Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) 2020 workflow diagram.

The following is the process flow:

1. Submit natural language prompt to Scholar Labs.
2. System scans top articles from index.
3. AI-based filtering applied according to prompt criteria.
4. Results returned progressively (36 papers).
5. Each paper evaluated against inclusion/exclusion criteria.
6. Relevant papers identified.

3.3.2. Traditional Google Scholar Selection Process

The traditional Google Scholar search was conducted using the constructed Boolean query. Results were sorted by relevance and examined sequentially. The first 300 results were evaluated from 30 pages of Google Scholar to maintain comparability with the Scholar Labs approach.

The following is the process flow:

1. Enter Boolean search query in traditional Google Scholar.
2. Results displayed sorted by relevance (10 per page).
3. Manual review of 300 results from first 30 pages.
4. Each paper evaluated against inclusion/exclusion criteria.

5. Relevant papers identified.

Figure 1 represents the identification, screening, eligibility testing, and inclusion stages of the literature survey. It initially shows the process of both Scholar Labs and Google Scholar records, which are combined in the later stages to yield the final set of papers included for quantitative analysis.

All retrieved records were screened using the inclusion and exclusion criteria presented in Tables 1 and 2. Screening was conducted in two stages. First, the title and abstract of each record were reviewed to determine whether the study appeared to address the integration of LLM methods with IoT/IoMT technologies in healthcare. Records that appeared potentially relevant were then assessed in greater detail by reviewing the full text, methodology, system architecture, evaluation design, and reported contributions. Studies were included only if they satisfied all inclusion criteria and did not meet any exclusion criteria. The term LLM–IoT *integration* is used broadly in this review to describe studies in which language-model methods and IoT/IoMT technologies contribute to the same healthcare application or system. The degree of coupling varies across the literature. In some studies, IoT-generated data are directly processed or interpreted by the language model, or the model influences downstream IoT decisions or workflows. In others, the language-model component performs a more limited function, such as NLP, summarization, extraction, or classification, alongside IoT components within the same healthcare application. Studies were retained when both technologies were actually used in the proposed or evaluated system, even if the LLM was not the primary contribution. Studies were excluded when they lacked LLM relevance, did not include IoT/IoMT integration, were outside the healthcare domain, were review/survey papers, lacked technical validation, or were duplicates. Screening and eligibility assessment were initially performed by a junior researcher using the predefined inclusion and exclusion criteria. The screening decisions and final study selection were iteratively validated with a senior researcher, and any uncertainties were resolved through discussion and revision. Because the quality assessment was not conducted independently by multiple coders throughout the entire process, inter-reviewer agreement statistics such as Cohen’s kappa were not calculated.

3.4. Data Extraction

After the final article list was prepared from both search methods, data were extracted into a structured Excel spreadsheet. The extraction table included the following variables: paper ID/source, publication year, healthcare application, IoT device type, data type, data details, LLM role, LLM model, architecture type, evaluation method, evaluation source/dataset, security/privacy/ethical challenges, key findings, and reported gaps or limitations.

These extracted variables were used to construct the taxonomy across four major dimensions: healthcare application domain, IoT device category, LLM role category, and deployment architecture category. When a study belonged to more than one category, multi-label coding was applied because some of the papers present more than one approach or a combination approach. Therefore, some cross-tabulation totals represent category occurrences rather than unique papers. Missing or unclear information was coded as “not reported” or “unspecified” when the original paper did not provide enough detail.

3.5. Comparative Analysis of the Search Methodologies

For the comparison of the two approaches, we selected the first 36 results from Google Scholar because Scholar Labs returned only 36 results. The performance difference in both methodologies and distribution of the search results is displayed in Table 3 and Figure 2.

Table 3. Comparison of the two methodology approaches.

AI (Scholar Labs)	Manual (Google Scholar)
Total Results Retrieved: 36	Total Screened = 36
Papers Included = 31	Papers Included = 15
Papers Excluded = 5 • Survey/Review = 0 • Out of Scope = 3 • Duplicate = 2	Papers Excluded = 21 • Survey/Review = 8 • Out of Scope = 12 • Duplicate = 1
True Positives = $31/36 = 86\%$	True Positives = $15/36 = 42\%$
False Positives = $5/36 = 14\%$	False Positives = $21/36 = 58\%$

Note: Percentages are calculated using the first 36 results from each search approach to enable direct comparison. The exclusion categories shown here correspond only to this comparative 36-result sample and therefore differ from the complete screening and eligibility exclusions reported in the PRISMA flow diagram.

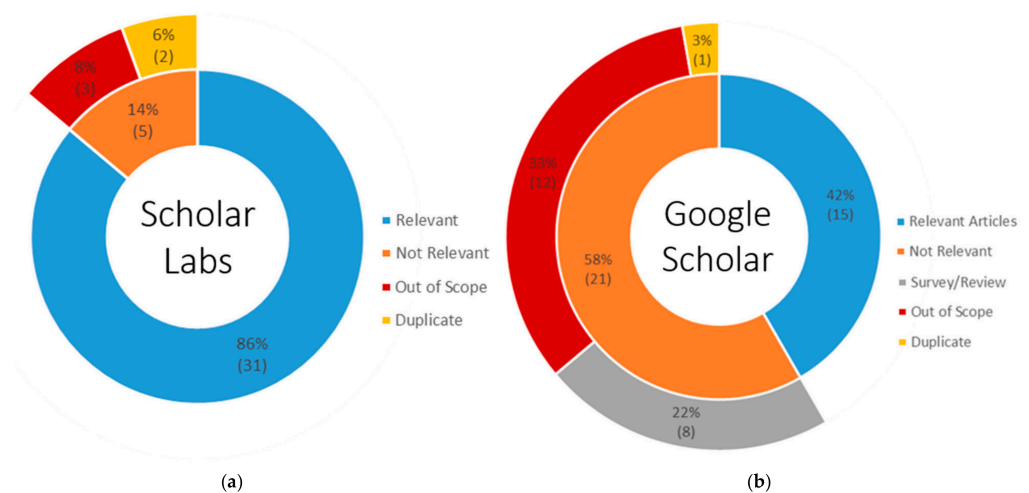


Figure 2. Search results distribution of Scholar Labs and Google Scholar (first 36 results): (a) Scholar Labs results showing relevant articles and excluded articles, including out-of-scope and duplicate papers; (b) Google Scholar results showing relevant articles and excluded articles, including survey/review articles, out-of-scope papers, and duplicates.

3.5.1. Overlap Analysis

Following individual selection processes, an overlap analysis was conducted to identify common papers retrieved by both search methods. This analysis is limited to the first 36 results of Google Scholar for comparison and employed the following approach:

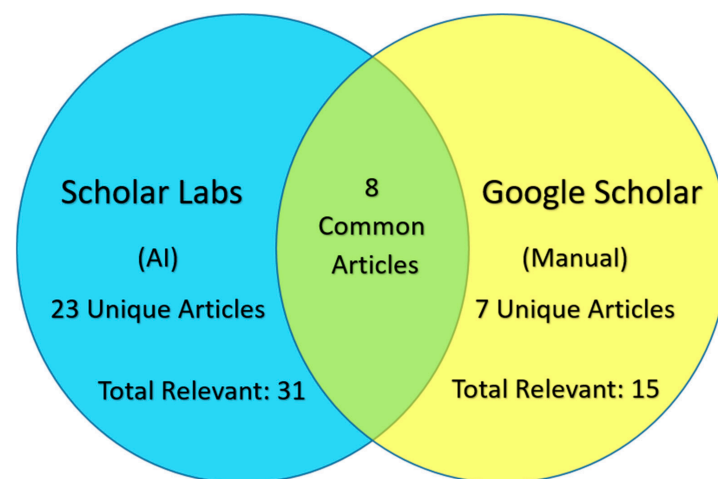
1. Extract unique identifiers (DOI, title) from both result sets.
2. Match papers across datasets using exact title matching and DOI comparison.
3. Classify papers as: Scholar Labs Unique, Google Scholar Unique, or Common.

The overlap analysis revealed 8 papers common to both search methods, with 23 unique papers from Scholar Labs and 7 unique papers from traditional Google Scholar.

Table 4 shows the number of common and unique articles found by both methods and their proportion of common articles among the relevant articles, while Figure 3 visually illustrates the distribution of unique and common papers between the two search methods through a Venn diagram. Note: Overlap percentages are calculated relative to the relevant articles retrieved by each search method.

Table 4. Overlap analysis summary.

Category	Value
Scholar Labs Unique Articles	23
Google Scholar Unique Articles	7
Common Articles	8
Overlap (% of Keyword-Based)	53%
Overlap (% of Prompt-Based)	26%
Total Unique Articles	38

**Figure 3.** Overlap analysis Venn diagram displaying distribution of unique and common papers in both methods.

3.5.2. Corpus-Relative Recall

To evaluate the retrieval effectiveness of each discovery strategy, we calculated corpus-relative recall against the final verified reference corpus of $Q = 75$ unique studies included in the quantitative analysis. In this context, Q represents the total number of studies in the final reference corpus, while true positives (TP) represent the number of studies from this corpus retrieved by each search method within the comparative screening window. Unlike conventional recall calculated against an independently established gold-standard dataset, this measure is intended to simulate an initial literature search and determine what proportion of the final known corpus each method would recover. Accordingly, corpus-relative recall is calculated as the proportion of the 75-study reference corpus retrieved by each method.

The corpus-relative recall for the AI-powered prompt-based strategy is formulated as:

$$R_{AI} = \frac{TP_{AI}}{Q} = \frac{31}{75} \approx 41.33\%$$

The corpus-relative recall for the traditional manual keyword-based strategy is formulated as:

$$R_{Manual} = \frac{TP_{Manual}}{Q} = \frac{15}{75} \approx 20.00\%$$

Because the 75-study reference corpus was itself constructed through the broader hybrid discovery and screening process, these values should be interpreted as corpus-relative retrieval performance rather than independent estimates of absolute search recall.

3.5.3. Strengths and Weaknesses Analysis

This section presents a detailed comparison between the two search methodologies based on the extracted results, as summarized in Table 5.

Table 5. Scholar Labs and Google Scholar strengths and weaknesses matrix.

	Scholar Labs	Google Scholar
Strengths	<u>Higher Precision</u> : Achieved 86% precision compared to 42% for traditional search, indicating an improvement in the result relevance.	<u>Methodological Transparency</u> : Well-established, reproducible methodology that can be precisely documented.
	<u>Effective Exclusion</u> : Automatically excluded all survey and review papers (0% vs. 22%), reducing manual screening workload.	<u>Complete Control</u> : Researchers have full control over search logic through Boolean operators.
	<u>User-Friendly Interface</u> : Natural language interface reduces technical barriers for researchers unfamiliar with Boolean query construction.	<u>Academic Acceptance</u> : Widely accepted and commonly used in academic research.
		<u>Systematic Documentation</u> : Easier to document and replicate in systematic reviews.
Weaknesses	<u>Filtering Limitations</u> : Two papers were dissertations/theses, two duplicate papers and one paper with only partial relevance.	<u>Low Precision</u> : Only 42% precision for original research articles.
	<u>Algorithm Transparency</u> : Limited transparency in AI interpretation of prompts and ranking algorithm.	<u>Inadequate Filtering</u> : High retrieval rate of survey/review papers (22%) and out-of-scope papers (33%) and duplicates (3%) that require manual exclusion.
	<u>Reproducibility Concerns</u> : Due to the non-deterministic nature of AI-based search, repeated use of the same prompt may produce differences in result ordering or ranking, limiting exact reproducibility.	<u>Technical Requirements</u> : Requires knowledge of Boolean query construction and search syntax.
		<u>High Workload & Time-Consuming</u> : The task is cumbersome and relatively takes longer time.

3.5.4. Combined Analysis

The combination of both search methods yielded 38 unique articles, demonstrating the value of a hybrid approach. While Scholar Labs provided 23 unique papers, traditional Google Scholar contributed 7 unique papers, with 8 papers overlapping between both methods.

The comparative analysis reveals that Scholar Labs achieves better performance across most metrics based on the evaluated search sample, particularly in terms of precision (86% vs. 42%) and exclusion of irrelevant paper types.

However, the traditional Boolean approach identified 7 unique papers not found by Scholar Labs, indicating its value as a supplementary search strategy.

3.6. Interpretation

The comparative analysis also reveals differences between the two search methodologies. Scholar Labs' AI-powered approach appears to offer higher precision than traditional Boolean search. This finding aligns with previous studies showing that AI-powered search systems can better match research papers to relevant topics by analyzing the meaning of the text rather than relying only on keywords [17].

One of the most significant advantages observed from Scholar Labs was its ability to exclude pure survey and review articles and articles loosely related to the prompt at the search stage. In traditional Boolean database searching, review articles frequently appear among retrieved results and require manual exclusion during screening, increasing reviewer workload. The AI-powered approach helped address these limitations; for example, it automatically excluded review articles within the evaluated sample, suggesting potential to streamline title/abstract screening workflows, although net time reduction was not directly quantified.

3.6.1. Retrieval Methodological Challenges

1. Scope Filtering Issues: Both approaches retrieved papers that were not directly related to AI/LLM or IoT or clinical healthcare, requiring manual exclusion.
2. Access Restrictions: In both approaches, some papers were only partially accessible or available as abstracts only, limiting full-text evaluation and requiring exclusion.
3. Reproducibility and Non-determinism: A further limitation of Scholar Labs emanates from the non-deterministic nature of LLM-based systems [18]. Repeated execution of the same search prompt may produce differences in the ordering or ranking of retrieved results. To assess this behavior, the search was repeated independently by one author and a co-author using the same prompt. Differences in result ordering were observed across these executions, indicating that the Scholar Labs search may not be exactly reproducible even when the search prompt remains unchanged. We want to emphasize that this observed non-determinism warrants further investigation. This variability represents an important methodological consideration when using AI-powered search tools in systematic literature reviews.

3.6.2. Implications for Systematic Literature Reviews

The findings have significant implications for systematic literature review methodology:

1. Search Strategy Selection: Researchers may find AI-powered search tools a better fit for initial literature discovery due to their simpler natural-language interface.
2. Hybrid Approaches: Combining AI-powered and traditional search methods may provide optimal coverage leveraging the complementary strengths of both approaches.
3. Manual Screening: Regardless of search method, manual title/abstract screening remains essential to assess relevance and ensure consistency with the predefined inclusion and exclusion criteria.

3.7. Summary and Recommended Strategy

3.7.1. Summary of Comparative Study

This methodology section has presented a comprehensive comparison of two search strategies for systematic literature reviews: Scholar Labs (AI-powered prompt-based search) and traditional Google Scholar (keyword-based Boolean search). The evaluation was conducted using a real-world systematic literature review on AI/LLM-IoT integration in healthcare.

Results of our comparative study indicate that Scholar Labs demonstrated a higher precision than traditional Google Scholar (86% vs. 42%) and was more effective in limiting the retrieval of irrelevant paper types at the search stage. The AI-powered search retrieved fewer out-of-scope papers (8% vs. 33%) and no survey/review papers (0% vs. 22%), suggesting that it may reduce the amount of manual screening required. However, the two approaches were complementary, as traditional Google Scholar identified seven unique papers that were not retrieved through Scholar Labs.

Based on the comparative analysis, Scholar Labs (AI-powered prompt-based search) demonstrated potential advantages for systematic literature reviews in terms of:

- Higher precision in the evaluated search (86% vs. 42%).
- More effective exclusion of irrelevant paper types.
- Potential reduction in researcher screening time.

These findings should not be interpreted as evidence that AI-powered prompt-based search, particularly Scholar Labs, is universally more effective than traditional Google Scholar. Systematic reviews require transparency, repeatability, reproducibility, and comprehensive literature coverage. A key limitation of Scholar Labs is the non-deterministic nature of AI-powered search systems, where the same prompt may not consistently gener-

ate identical results or rankings across repeated searches or different users, a characteristic also observed in LLM-based systems [18]. This variability may affect the repeatability and reproducibility expected in systematic literature reviews. In addition, traditional Google Scholar retrieved seven unique relevant papers that were not identified through Scholar Labs, highlighting the importance of using complementary search approaches to improve literature coverage. Therefore, while Scholar Labs showed potential advantages in precision and screening efficiency in this study, traditional Google Scholar remains an important complementary search strategy for supporting broader coverage and more reproducible search procedures.

3.7.2. Recommendations for AI-Assisted Hybrid Systematic Literature Reviews

For researchers conducting systematic literature reviews, we recommend a hybrid strategy that effectively combines both methodologies.

Figure 4 illustrates our proposed hybrid literature search strategy process in a step-by-step manner for an effective systematic literature review. This hybrid approach leverages the strengths of both methodologies while mitigating their individual limitations, resulting in a more thorough and efficient systematic literature review process.

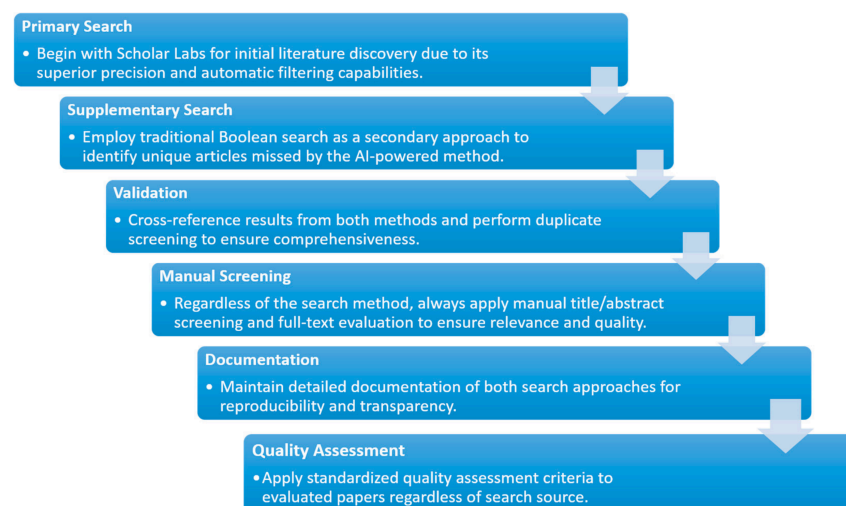


Figure 4. Our proposed hybrid literature search strategy.

4. Taxonomy and Classification

4.1. Taxonomy Overview

Based on the literature survey, a total of 75 papers were identified, as shown in Figure 1. To systematically analyze the current state-of-the-art and identify major research patterns, these studies were classified according to four key dimensions: healthcare application domain, LLM role, IoT devices, and architecture type.

Figure 5 illustrates the proposed multi-layer taxonomy, which organizes the selected studies into four major dimensions and their corresponding subcategories. Among these, the healthcare application domain consists of five categories, while both the LLM role and architecture type contain four categories each. The IoT devices dimension is divided into three categories.

This taxonomy provides a structured understanding of how system design choices are influenced by domain-specific requirements, data sources, deployment environments, and operational constraints. Furthermore, it highlights the major areas of research concentration as well as underexplored areas that offer opportunities for future research.

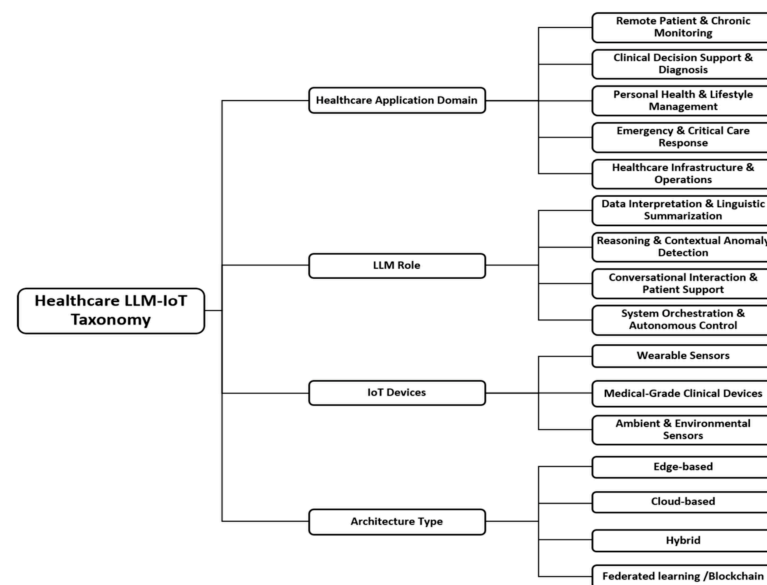


Figure 5. Taxonomy of LLM-IoT healthcare systems.

4.1.1. Taxonomy Development and Classification Procedure

The taxonomy construction followed an iterative, data-driven approach grounded in the selected literature, as opposed to a fixed, deductive framework. An initial taxonomy was created based on an overall assessment of the corpus, healthcare domain knowledge, and AI-assisted suggestions. The preliminary categories included Remote Monitoring, Smart Hospitals, Elderly Care, Clinical Support, and Mental Health for healthcare applications; Reasoning, Interaction, Interpretation, Optimization, Summarization, and Security/Verification for LLM roles; and Edge-based, Cloud-based, Hybrid, Decentralized, and Federated architectures.

During data extraction, information corresponding to each taxonomy dimension was recorded in separate columns for every included study. As additional studies were examined, overlapping or closely related categories were merged, ambiguous categories were refined, and category labels were revised to better reflect recurring patterns across the corpus. This iterative process resulted in the four final taxonomy dimensions and subcategories presented in Figure 5. Although categories like federated learning and blockchain are technically distinct, they were grouped within a single architecture category because, across the reviewed studies, both were primarily employed as decentralized strategies addressing privacy, security, trust, or data-governance requirements rather than because of technical similarity.

The classification followed a multi-label approach in which categories were not mutually exclusive. Studies were assigned to more than one category within a dimension when they explicitly exhibited multiple functions or system characteristics. For example, an LLM could support both conversational interaction and data interpretation, or a system could incorporate more than one IoT device type or architectural component. This approach preserved the multifunctional nature of the reviewed systems and avoided forcing studies into a single category when their reported design spanned multiple taxonomy classes.

4.1.2. Taxonomy Category Definitions

To ensure clarity and consistency in classification, the taxonomy dimensions used in this study are defined as follows.

Healthcare Application Categories:

1. **Remote Patient and Chronic Monitoring:** Refers to systems for long-term tracking of physiological signals outside clinical settings to manage chronic conditions or post-operative recovery (e.g., continuous heart rate monitoring for cardiac patients and mental health patients).
2. **Clinical Decision Support and Diagnosis:** Includes tools that assist healthcare professionals in identifying diseases or recommending treatments by interpreting complex data (e.g., LLM-augmented analysis of ECG data for rare arrhythmia detection).
3. **Personal Health and Lifestyle Management:** Represents consumer-facing applications focused on wellness, fitness, and nutrition through behavioral tracking (e.g., LLM-based coaching for protein intake based on activity logs).
4. **Emergency and Critical Care Response:** Includes systems optimized for the immediate detection of acute health crises or mass-casualty events (e.g., automated fall detection or early-warning systems for sepsis in ICUs).
5. **Healthcare Infrastructure and Operations:** Focuses on backend systems that support reliability, security, and efficiency of healthcare IoT ecosystems.

IoT Device Categories:

1. **Wearable Sensors:** Refers to body-worn or mobile devices that collect physiological data.
2. **Ambient and Environmental Sensors:** Includes stationary sensors that monitor surroundings or user behavior.
3. **Medical-Grade Clinical Devices:** Refers to regulated diagnostic or therapeutic hardware used in clinical environments.

LLM Role Categories:

1. **Data Interpretation and Linguistic Summarization:** Involves transforming sensor data into human-readable outputs.
2. **Conversational Interaction and Patient Support:** Includes natural language interfaces for patient engagement.
3. **Reasoning and Contextual Anomaly Detection:** Refers to identifying risks or anomalies using contextual understanding.
4. **System Orchestration and Autonomous Control:** Involves managing system workflows, data routing, or decision coordination.

Architecture Type Categories:

1. **Edge-Based/Local:** Where computing and LLM inference occur entirely on the local device or at the network edge (e.g., gateway, local server).
2. **Cloud-Based:** Where data is transmitted to a centralized remote server or a large-scale cloud provider for high-performance processing and heavyweight LLM inference.
3. **Federated Learning/Blockchain:** Consists of papers that employed federated learning and blockchain technologies. In the reviewed studies, federated learning is primarily used to support distributed model training or analysis without centrally pooling raw healthcare data, offering privacy advantages. Blockchain-based systems, in contrast, are used to support data integrity, auditability, and secure decentralized health-data management.
4. **Hybrid:** Represents implementations that combine multiple architectural layers (e.g., edge + cloud, edge + fog + cloud).

In what follows, we provide demonstrative examples of systems across taxonomy categories; the ten representative papers presented in Table 6 are grouped by healthcare domain and application context.

Table 6. Representative studies integrating LLMs and IoT in healthcare.

Paper ID	Healthcare Application Domain	IoT Type	LLM Role	Architecture	Key Contributions
[19]	Remote Patient & Chronic Monitoring	Wearable Sensors	- System Orchestration & Autonomous Control - Conversational Interaction & Patient Support - Data Interpretation & Linguistic Summarization	Hybrid (edge + cloud)	Integrates LLM-guided DRL, diffusion denoising, and conversational feedback to generate personalized health alerts.
[20]	Remote Patient & Chronic Monitoring	Medical-Grade Clinical Devices	Reasoning & Contextual Anomaly Detection	Cloud-based	LLaMA-based anomaly detection on IoT medical telemetry with efficient training using LoRA.
[21]	Clinical Decision Support & Diagnosis	Medical-Grade Clinical Devices	Reasoning & Contextual Anomaly Detection Data Interpretation & Linguistic Summarization	Federated learning/Blockchain	Combines LLMs with lightweight blockchain, IPFS in secure Healthcare 5.0 frameworks.
[22]	Clinical Decision Support & Diagnosis	Wearable Sensors	Reasoning & Contextual Anomaly Detection Conversational Interaction & Patient Support	Cloud-based	Integrates continuous sensor data with retrieval-augmented LLM for disease probability estimation.
[23]	Personal Health & Lifestyle Management	Wearable Sensors	Data Interpretation & Linguistic Summarization	Hybrid (edge + fog + cloud)	Dynamically allocates tasks based on data sensitivity and complexity via modular layers.
[24]	Personal Health & Lifestyle Management	Wearable Sensors	Data Interpretation & Linguistic Summarization Conversational Interaction & Patient Support	Cloud-based	PhysioLLM improves actionable health insights from wearable data.
[25]	Emergency & Critical Care Response	Medical-Grade Clinical Devices	Data Interpretation & Linguistic Summarization Reasoning & Contextual Anomaly Detection	Edge-based	Lightweight BERT deployed on edge IoT achieves superior performance over SVM, CNN, LSTM, and GRU on MIMIC-III.
[26]	Emergency & Critical Care Response	Medical-Grade Clinical Devices	Data Interpretation & Linguistic Summarization	Federated learning	Fine-tuned LLM outperforms baseline models and typical medical student on USMLE-style questions in the MIMIC III infectious disease cohort.
[27]	Healthcare Infrastructure & Operations	Medical-Grade Clinical Devices	Reasoning & Contextual Anomaly Detection	Hybrid (cloud +edge+ blockchain)	Integrates Deepseek-R1 LLM with blockchain and IoT for secure real-time fault detection, for offline secure deployment.
[28]	Healthcare Infrastructure & Operations	Medical-Grade Clinical Devices	System Orchestration & Autonomous Control	Hybrid (edge + cloud)	MEDIotWALL demonstrates a low-cost, Raspberry Pi-based, two-tier IoMT firewall with AI-assisted rule generation to provide effective protection of medical devices.

5. Results and Analysis

5.1. RQ1: Healthcare Domain Analysis

The distribution of healthcare applications across the taxonomy in Table 7 reveals a strong concentration in remote patient and chronic monitoring systems, followed by personal health and lifestyle management and clinical decision support and diagnosis, while emergency and critical care and infrastructure-focused applications appear less frequently. This distribution reflects the natural alignment between IoT capabilities and long-term physiological monitoring, where continuous data streams from wearable and medical-grade devices are readily available for LLM-driven interpretation. The frequencies reported in this review represent the level of research attention within the selected literature and should not be interpreted as evidence of greater scientific importance, clinical effectiveness, or technological superiority of one application, model, device, or architecture over another.

Table 7. Healthcare application category distribution.

Healthcare Application Category	Count	Percentage (%)
Remote Patient & Chronic Monitoring	24	32%
Personal Health & Lifestyle Management	20	27%
Clinical Decision Support & Diagnosis	14	19%
Emergency & Critical Care Response	10	13%
Healthcare Infrastructure & Operations	7	9%
Total	75	100%

Within the remote patient and chronic monitoring cohort, research is primarily focused on elderly care and remote monitoring, telemedicine and diagnosis, followed by generic chronic-condition monitoring, diabetes management, and cardiovascular conditions. The remaining studies address mental health, Alzheimer's disease, and pediatric health. In clinical decision support, decision support and diagnostic reasoning were the most frequent topics, followed by mental health and predictive analytics and anomaly detection. The remaining studies focused on disease-specific applications involving the eye, ear, heart, and tuberculosis. In personal health and lifestyle management, the largest number of papers concerned sleep and stress, followed by behavioral analytics, lifestyle management, training, tracking, and coaching. In emergency and critical care, approximately half of the papers addressed emergency alerts, medical diagnosis, context awareness, and fall risk. The remaining studies examined post-surgery/ICU monitoring, stroke prediction, and remote emergency care. Finally, healthcare infrastructure and operations showed no clear dominant trend.

Remote monitoring systems are observed to leverage LLMs for continuous signal interpretation and anomaly detection, transforming structured sensor data into meaningful insights for both clinicians and patients [19,20,29–45]. In contrast, clinical decision support applications utilize LLMs for reasoning and contextual understanding [21,22], often integrating multiple data modalities such as patient history, sensor readings, and conversational inputs to support diagnostic workflows [46–55]. Personal health and lifestyle management systems represent a distinct category where LLMs are used in interactive and behavioral contexts, enabling personalized feedback, coaching, and engagement through conversational interfaces [24,56–66]. These systems emphasize usability and personalization over clinical rigor, but increasingly incorporate trust mechanisms such as grounding and guardrails.

Emergency and critical care applications, while less frequent, highlight low-latency, edge-based and hybrid decision-making, where real-time responsiveness is critical [25,67–72].

These systems seem to prioritize performance and reliability, sometimes at the expense of complex security mechanisms.

The domain distribution indicates that LLM–IoT healthcare research is heavily driven by data availability and application feasibility, with remote monitoring dominating due to continuous data streams, while more complex domains such as emergency care and infrastructure remain relatively underexplored.

Healthcare Application × LLM Role Cross-Tabulation

A cross-tabulation between healthcare application category and LLM role category, presented in Table 8, reveals that domain characteristics strongly influence LLM functionality. Because LLM roles were coded using a multi-label approach, percentage values were calculated by dividing the LLM-role paper count within each healthcare application category in Table 8 by the total paper count for that healthcare application category reported in Table 7. As shown in the cross-tab analysis, remote patient monitoring and personal health applications are dominated by data interpretation and conversational roles, reflecting the need for continuous summarization and user interaction. In contrast, clinical decision support systems exhibit a significantly higher concentration of reasoning-based models (79%), highlighting their reliance on context-aware inference for diagnostic support. System orchestration remains largely confined to healthcare infrastructure (57%) [28,73–75], indicating that autonomous control capabilities are not yet widely adopted in patient-facing domains.

Table 8. Healthcare application LLM-role distribution.

Healthcare Application Category	LLM Role			
	Reasoning & Contextual Anomaly Detection	Data Interpretation & Linguistic Summarization	Conversational Interaction & Patient Support	System Orchestration & Autonomous Control
Remote Patient & Chronic Monitoring	8	14	12	1
Personal Health & Lifestyle Management	7	15	12	1
Clinical Decision Support & Diagnosis	11	8	4	0
Emergency & Critical Care Response	4	5	3	2
Healthcare Infrastructure & Operations	3	1	0	4

Note: LLM roles were coded using a multi-label approach; therefore, one paper may appear under more than one LLM role category and the sum of papers of healthcare application category may exceed the actual number of papers presented in Table 7.

5.2. RQ2: LLM Models and Usage Analysis

Analysis of LLM usage across the taxonomy shown in Table 9 reveals that models are primarily employed for three core functional roles: data interpretation, contextual reasoning, and conversational interaction. Among these, data interpretation and summarization is the most frequently studied LLM role in the reviewed literature, reflecting the need to convert raw IoT-generated data into structured and actionable insights.

Reasoning-based applications form the second dominant category, where LLMs are used to identify anomalies, detect risks, and support clinical decision-making by contextualizing heterogeneous data inputs [20–22]. These systems often rely on fine-tuned or domain-adapted models [20,25,35] to improve accuracy in healthcare-specific tasks [51,54,76–78].

Table 9. LLM role category frequency distribution.

LLM Role Category	Count	Percentage (%)
Reasoning & Contextual Anomaly Detection	33	43%
Data Interpretation & Linguistic Summarization	43	57%
Conversational Interaction & Patient Support	31	41%
System Orchestration & Autonomous Control	8	12%

Note: Percentages are calculated based on the total number of reviewed papers ($n = 75$). Since papers may include multiple LLM roles, percentages do not sum to 100%.

Conversational LLM applications, while less frequent, play a critical role in patient-facing systems, including virtual assistants and digital health coaches [24,60,79], as well as chronic-pain guidance systems [80]. These systems enable natural language interaction, enhancing accessibility and user engagement [33,56,57,61].

A smaller but important category of research articles includes system orchestration and autonomous control, where LLMs are integrated into broader workflows to coordinate IoT systems [28,69] and automate decision pipelines [73,74]. This category represents a shift toward more autonomous healthcare systems.

LLM Model × Architecture Type × Healthcare Application

Model-level cross-tabulation in Table 10 shows that GPT-family models are the most frequently used model group and seem to be primarily associated with cloud-based or hybrid deployments, reflecting their suitability for complex reasoning, summarization, and conversational support. LLaMA-family models appear across all architecture types and show stronger relevance in privacy-sensitive or locally controlled settings in the studies, including hybrid and federated deployments. BERT-family models appeared comparatively more in edge-oriented, supporting resource-constrained tasks such as local classification, short-text interpretation, and lightweight clinical language processing. This pattern suggests that model selection is not arbitrary; rather, it is shaped by the interaction between task complexity, deployment constraints, privacy requirements, and healthcare domain.

Table 10. LLM model healthcare-application-architecture frequency distribution.

LLM Model	Total Papers	Architecture			
		Edge-Based	Cloud-Based	Hybrid	Federated Learning/Blockchain
GPT	33	4	19	9	1
LLaMA	24	7	8	7	3
BERT	10	5	2	2	1
Unspecified LLM	12	3	3	6	0
Other Proprietary Models (Gemini, Claude, Grok, Qwen-max, ERNIE Bot, LAM)	12	2	7	3	0
Other Open Models (Mistral & Mixtral, Deepseek, Qwen2.5, BART MNLI, Starling-LM)	12	4	3	5	0
Compact Models (Gemma, Phi, MiniLM, T5)	10	3	4	3	0

Note: Model categories were coded using a multi-label approach when studies used or compared multiple LLM families; therefore, model counts may exceed the total number of reviewed papers.

High-capability GPT-family models were commonly positioned as reasoning or orchestration layers in cloud and hybrid healthcare IoT systems [22,48–50], where they

support interpretation of heterogeneous telemetry and coordination of downstream decision workflows [53,55,81,82]. In contrast, open-source models such as LLaMA, especially served via Ollama, seemed more suitable for privacy-sensitive decision-support contexts, particularly when systems require local, institutional deployment [21,29], or federated deployment [26,54]. Lightweight transformer models such as BERT variants appeared better suited for resource-constrained edge environments [25,33,38], where real-time processing and data locality are more important than maximum reasoning depth [82]. Local deployment is also used to protect sensitive participant data, as Ref. [83] employed Phi-3 locally without uploading the collected data to the cloud. This reflects a broader architectural trade-off. Cloud deployments provide access to the computational resources required for more intensive LLM processing [84]. Edge deployments are used where local or low-latency inference is important [25,38], while federated approaches are adopted in privacy-sensitive settings where raw healthcare data should remain distributed [26,47]. Blockchain-based approaches also support secure healthcare data sharing, as Ref. [85] uses a hierarchical cross-chain architecture to connect distributed hospitals with a health center.

Although GPT-based models are generally considered cloud-oriented due to their large computational requirements, a few studies classified them under edge-based architectures because the GPT models were mainly used during offline training and preparation stages rather than continuous cloud inference. In these studies, GPT-4 and GPT-4o were used as secondary LLM for synthetic data generation, summarization, and teacher-based supervision, while GPT-2 and GPT-NeoX-20B were distilled or fine-tuned into lightweight student models suitable for deployment on edge devices [20,75]. This enabled the final healthcare application to operate in an edge environment while still benefiting from the capabilities of GPT-based models.

5.3. RQ3: Integration of LLM and IoT

The IoT category distribution in Table 11 shows that wearable sensors are the most frequently represented IoT device category in the reviewed studies, accounting for 48 papers (64%), followed by medical-grade clinical devices with 21 papers (28%). Ambient and environmental sensors represent six papers (8%). This indicates that LLM-IoT healthcare research is primarily centered on continuously connected, patient-facing devices, especially wearables, which naturally support remote monitoring [19,39,42–44], lifestyle management [56,59,61–63], and real-time health tracking [68,69,71,72]. In addition, wearables are widely available and generally easier to deploy and evaluate than other IoT device types, which makes the testing less complicated.

Table 11. IoT-category frequency distribution.

IoT Category	Count	Percentage (%)
Wearable Sensors	48	64%
Medical-Grade Clinical Devices	21	28%
Ambient & Environmental Sensors	6	8%
Total	75	100%

5.3.1. Healthcare Application × IoT Category Cross-Tabulation

As shown in Table 12, cross-tabulation between healthcare applications and IoT categories device selection is strongly shaped by application context. Personal health and lifestyle management relies most heavily on wearable sensors, reflecting the accessibility and continuous data collection needs of consumer-facing systems. Remote patient monitoring also depends primarily on wearables but includes medical-grade and ambient sensors, indicating a broader sensing ecosystem for chronic and longitudinal care. In contrast,

healthcare infrastructure and operations is dominated by medical-grade devices [27,28,77], suggesting that operational applications are more concerned with clinical equipment, device reliability, and regulated system performance than with consumer sensing [72,74,84]. Ambient and environmental sensors appear mainly in a small set of remote-monitoring studies [31,32,86] and in other personal/emergency contexts [60,70,81], indicating that environmental context is still underused in LLM–IoT healthcare research.

Table 12. Healthcare application IoT-category frequency distribution.

Healthcare Application Category	IoT Category			Total
	Wearable Sensors	Medical-Grade Devices	Ambient & Environmental	
Remote Patient & Chronic Monitoring	15	6	3	24
Personal Health & Lifestyle Management	18	0	2	20
Clinical Decision Support & Diagnosis	8	6	0	14
Emergency & Critical Care Response	6	3	1	10
Healthcare Infrastructure & Operations	1	6	0	7
Total	48	21	6	75

The integration of LLMs with IoT systems follows distinct architectural patterns, primarily categorized as cloud-based, edge standalone/local-based, hybrid, and federated approaches. These architectures reflect trade-offs between computational capability, latency, and data privacy.

From the Table 13 architecture distribution, cloud-based architectures are frequently reported in the reviewed literature, including remote-monitoring systems [20,36,37,39] and clinical decision-support systems [22,48–51,87]. Cloud-hosted GPT-3.5/ChatGPT is also used in emergency remote healthcare and chronic-pain monitoring [80,88]. Their use is consistent with the computational demands of intensive LLM inference and reasoning [84]. However, these approaches introduce concerns related to data centralization and privacy risks [84].

Table 13. Architecture-type frequency distribution.

Architecture Type	Count	Percentage (%)
Cloud-based	30	40%
Edge Standalone/Local-based	18	24%
Federated Learning/Blockchain	5	6%
Hybrid	24	32%

Note: Architecture categories include studies with combined or multi-layer deployments; therefore, total architecture occurrences may exceed the number of reviewed papers.

Edge standalone/local-based architectures address some limitations of centralized processing by moving inference closer to the data source. In the reviewed studies, lightweight edge implementations support resource-constrained and low-latency healthcare inference [25,38]. Local deployment is also used to reduce reliance on external cloud services and limit exposure of sensitive healthcare information [29,30]. Lightweight or compact models are therefore commonly favored in edge-oriented and compute-constrained environments.

Hybrid architectures combine both approaches, allowing initial processing at the edge followed by advanced analysis in the cloud [56,67]. This pattern is increasingly explored in the reviewed literature as an approach for balancing performance and privacy.

Federated learning and blockchain-based systems represent a decentralized integration strategy in which models can be trained or analyzed across distributed devices or institutions without centrally pooling raw data. Several reviewed studies combine federated learning with privacy-enhancing mechanisms such as differential privacy and secure

aggregation [26,47]. Other implementations combine federated processing with encryption and authentication to protect distributed model updates [73]. However, these approaches also introduce communication, scalability, and implementation complexity.

5.3.2. Healthcare Application $\times$ Architecture Type Cross-Tabulation

The cross-tabulation in Table 14 further highlights a strong alignment between healthcare applications and architectural choices. These cross-tab percentages were calculated using the same method as mentioned under Table 8 by dividing each architecture-specific count in Table 14 by the corresponding total Healthcare Application Category count reported in Table 7. Clinical decision support systems are predominantly cloud-based (57%) [22,48–51,87], reflecting their dependence on high-capacity reasoning models [53,55]. In contrast, emergency and critical care systems largely show higher hybrid adoption (40%), followed by edge adoption (30%), while still emphasizing the importance of low-latency, real-time decision-making [25,68–70]. Remote patient monitoring demonstrates the highest hybrid adoption (38%) [33–35,37,40], indicating the need to balance continuous monitoring with centralized analysis [39,88–90].

Table 14. Healthcare application architecture-type frequency distribution.

Healthcare Application Category	Architecture			
	Edge-Based	Cloud-Based	Hybrid	Federated Learning/Blockchain
Remote Patient & Chronic Monitoring	6	8	9	2
Personal Health & Lifestyle Management	6	10	5	0
Clinical Decision Support & Diagnosis	2	8	3	1
Emergency & Critical Care Response	3	2	4	1
Healthcare Infrastructure & Operations	1	2	3	1

Note: Architecture categories were coded based on reported deployment patterns. Studies using combined architectures may contribute to more than one architecture category.

5.3.3. LLM Role $\times$ Architecture Type Cross-Tabulation

Cross-tabulation of LLM role categories and architecture types in Table 15 shows variation in deployment patterns across functional roles. Because both LLM roles and architecture types were coded using a multi-label approach, the percentages discussed here were calculated using a similar method as mentioned under Table 8 by dividing each architecture-specific count in Table 15 by the corresponding total LLM role count reported in Table 9. Therefore, these percentages represent the proportion of papers within each LLM role category that reported a given architecture type, and they may exceed 100% when summed across architectures. Reasoning and contextual anomaly detection was most often associated with cloud-based architectures (39%), followed by hybrid (30%) and edge-based (24%) deployments. This pattern suggests that reasoning-oriented systems in the reviewed corpus often use cloud resources, while some studies also explore edge or hybrid designs for contexts where local processing or reduced latency may be relevant. Data interpretation and linguistic summarization show a comparatively distributed pattern across cloud-based (39%), hybrid (30%), and edge-based (25%) architectures, suggesting that interpretation tasks are implemented across multiple deployment models. Conversational interaction and patient support follows a similar but slightly more cloud-oriented pattern, with cloud-based architectures accounting for 45%, followed by hybrid (35%) and edge-based (23%) deployments. System orchestration and autonomous control was more frequently observed in edge-based architectures (50%), which may reflect the role of local processing in workflow coordination and real-time system responsiveness in some studies [19,68,69,75].

Table 15. LLM role architecture-type frequency distribution.

LLM Role	Architecture			
	Edge-Based	Cloud-Based	Hybrid	Federated Learning/Blockchain
Reasoning & Contextual Anomaly Detection	8	13	10	2
Data Interpretation & Linguistic Summarization	11	17	13	4
Conversational Interaction & Patient Support	7	14	11	1
System Orchestration & Autonomous Control	4	2	3	0

Note: Both LLM role and architecture were coded using multi-label classification; therefore, totals represent category occurrences rather than unique papers.

5.3.4. IoT Category × Architecture Type Cross-Tabulation

The relationship between IoT categories and architecture in Table 16 reveals that wearable-based systems primarily rely on cloud deployments (48%), while edge and hybrid deployments each account for 27%, reflecting a trade-off between usability and responsiveness. In contrast, medical-grade clinical devices show the highest representation of the combined federated learning/blockchain category (19%). Within these systems, federated learning is primarily associated with privacy-preserving distributed processing of healthcare data [26], whereas blockchain-based implementations emphasize integrity, auditability, and secure data management [21,45]. The percentages discussed here were calculated using a similar method to the one mentioned under Table 8 by dividing each architecture-specific count in Table 16 by the corresponding total IoT category count reported in Table 11.

Table 16. IoT category architecture-type frequency distribution.

IoT Category	Architecture			
	Edge-Based	Cloud-Based	Hybrid	Federated Learning/Blockchain
Wearable Sensors	13	23	13	1
Medical-Grade Clinical Devices	4	6	7	4
Ambient & Environmental Sensors	1	1	4	0

Note: Counts represent architecture occurrences within each IoT device category; studies using hybrid or multiple deployment layers may contribute to more than one architecture category.

A key observation across the taxonomy is that many studies explicitly avoid fully centralized architectures due to security and privacy concerns. Instead, they adopt hybrid or edge-centric designs, where only processed or summarized data is transmitted [23,38].

5.3.5. Evaluation Setting and Validation Evidence

To distinguish the level of empirical validation across the reviewed studies, evaluation approaches were grouped according to whether they relied primarily on simulation or synthetic data, pre-existing real-world datasets, or direct evaluation involving real-world environments, participants, or healthcare professionals. Of the 75 reviewed studies, 23 (31%) relied primarily on simulation or synthetic evaluation, while 30 (40%) evaluated their approaches using pre-existing real clinical, physiological, wearable, or healthcare-related datasets without direct deployment in the intended healthcare environment. In contrast, 21 studies (28%) included some form of direct real-world evaluation, such as deployment in healthcare or daily-life environments, data collection from participants or patients, or evaluation involving healthcare professionals. One study primarily presented architectural analysis. These findings indicate that although many studies use realistic or clinically

derived data, direct prospective validation of LLM–IoT systems in operational healthcare settings remains comparatively limited.

Practical deployment metrics are evaluated in a subset of the reviewed studies, although reporting is inconsistent across the corpus. Inference and response latency receive the clearest attention, with studies comparing edge, fog, and cloud inference [23,38] or evaluating real-time healthcare responsiveness [26,68]. Resource-oriented evaluation is less common; for example, Ref. [38] reports memory/CPU utilization and power consumption during edge deployment, while [81] evaluates response latency and memory footprint under IoT anomaly-detection workloads. Communication efficiency is examined in federated and wireless healthcare systems through model-update, network, or operational overhead [47,69]. Privacy is generally addressed architecturally, although [47] additionally estimates privacy-leakage risk using membership-inference attacks. Security-oriented studies evaluate conventional IoT attack scenarios [20,77,81], whereas systematic robustness testing against LLM-specific adversarial attacks such as prompt injection and jailbreaks remains largely absent.

5.4. RQ4: Cybersecurity, Privacy, and Ethical Challenges

Security, privacy, and ethical concerns are common in LLM–IoT healthcare research, but they are addressed with varying levels of depth across the literature. Among the 75 reviewed papers, some do not discuss these concerns, while others acknowledge them without implementing specific safeguards and instead identify them as areas for future work.

About 57 papers explicitly address at least one of these dimensions. Their approaches include implemented security or privacy mechanisms, privacy-preserving architectural choices, regulatory or governance considerations, ethical oversight, and discussion of related risks and limitations. However, the level of engagement varies considerably, ranging from empirically evaluated security or privacy mechanisms to more limited measures such as local inference, de-identification, IRB approval, human oversight, or acknowledgement of risks including bias, hallucination, and data leakage.

5.4.1. Compliance Guidelines Mentioned

Regarding compliance, only five regulatory frameworks are explicitly named in the corpus. As shown in Table 17, HIPAA appears in seven papers and GDPR in six. HITRUST CSF, EU MDR, and TFDA Part 11 each appear once.

Table 17. Compliance and regulatory frameworks explicitly cited in the reviewed corpus.

Regulatory Framework	<i>n</i>	Papers
HIPAA	7	[21,23,28,38,44,49,91]
GDPR	6	[19,21,23,26,44,91]
HITRUST CSF	1	[28]
EU MDR	1	[28]
TFDA Part 11 (Taiwan)	1	[86]

Note: Regulatory framework counts are not mutually exclusive because a single paper may cite more than one framework.

The way these frameworks are invoked varies across studies. Ref. [38] discusses HIPAA as a rationale for local edge processing, while [23] describes its edge–fog architecture as supporting HIPAA/GDPR-aligned handling by limiting the transmission of identifiable health data. Ref. [44] describes its proposed tiered framework as “HIPAA/GDPR-compliant” and incorporates tokenization, de-identified embedding, and end-to-end encryption. Ref. [26] provides a detailed GDPR-oriented framework, incorporating DPIA support, rights-management workflows, continuous privacy KPI monitoring, and audit

logging. Ref. [49] describes its system as designed for HIPAA-compliant workloads, supported by PII/PHI quarantine tooling and least-privilege access controls. Ref. [28] explicitly aligns its IoMT network security architecture with HIPAA, HITRUST CSF, and EU MDR. However, in several studies, regulatory frameworks primarily serve as motivation or design guidance rather than evidence of demonstrated compliance, and no paper in the corpus reports an external or formal compliance audit, certification, or regulatory submission.

5.4.2. How Security, Privacy, and Ethics Were Addressed

To examine how security, privacy, and ethical concerns are addressed across the reviewed LLM–IoT healthcare studies, the identified safeguards were grouped into major thematic approaches. Table 18 summarizes these approaches across the three dimensions of security, privacy, and ethics, and lists the corresponding studies that adopted or discussed each mechanism.

Table 18. Approaches to security, privacy, and ethics mentioned in the corpus.

Dimension	Approach	Papers
Security	Cryptographic protection (AES-256, TLS, Triple DES, SHA-256)	[21,27,28,44,67,73,77,86,89]
	Blockchain integrity and audit	[21,27,45,73,85]
	Access control/RBAC/IAM	[27,28,45,49,77]
	LLM as intrusion/anomaly detection engine	[20,73,77,81]
	Human approval of AI-generated security rules	[28]
Privacy	Edge/local processing (no cloud transmission)	[23,29–32,34,38,41,46,51,56,59,78,83,86,89]
	Federated learning	[21,26,40,42,47,54,72,73,85,91]
	Differential privacy/secure aggregation	[26,47,54,72,91]
	De-identification/anonymization	[24,26,36,41,60,83,92]
	Tiered data-sensitivity routing (edge–fog–cloud)	[23,32]
Ethics	IRB/ethics committee approval	[24,37,50,60,61,64,65,83]
	Hallucination risk acknowledged; RAG as mitigation	[24,29,59,61,66]
	Bias/fairness concerns raised	[19,29,36,37,76,91,93]
	Human-in-the-loop/clinical guardrails	[24,28–30,49,53,63,74]
	Explainability (SHAP, LIME, traceable reasoning)	[53,91]

Note: Papers may appear under multiple approaches and dimensions because security, privacy, and ethical safeguards are not mutually exclusive.

Security Issues:

Security mechanisms mainly focus on cryptographic protection, access control, blockchain-based integrity preservation, and AI-driven threat detection. Encryption and authentication remain the most commonly adopted strategies for securing IoMT and EHR data.

Several studies implement strong cryptographic and access-control frameworks. For example, Ref. [27] combines SHA-256, TLS 1.2, and AES-256 encryption with offline-capable LLM deployment to maintain fully on-premises processing. Similarly, Ref. [28] employs SSH-based mutual authentication, Vault-managed secret storage, and micro-segmentation for secure device communication. More advanced IAM-oriented protection appears in [77], which integrates HTTPS, RBAC through Keycloak, and a fine-tuned DistilBERT model to detect credential misuse, replay attacks, privilege escalation, and lateral movement, alongside automated containment actions such as token revocation and IP blocking.

Federated healthcare systems introduce additional risks related to poisoned or manipulated updates. To address this, Ref. [73] combines Triple DES encryption with permutation-

based authentication and LLM-based semantic validation to detect malicious federated updates before aggregation.

Within the reviewed studies, blockchain architectures are primarily used to support auditability, tamper resistance, and decentralized trust. Ref. [21] integrates lightweight Ethereum with IPFS for tamper-proof medical report storage, while [45] applies NFT-governed digital twin ownership for pacemaker systems. Ref. [85] further proposes a cross-chain smart-contract framework for trustworthy inter-institutional data sharing. However, several studies note that blockchain integration increases computational overhead and scalability challenges when combined with LLM pipelines.

A smaller group of studies deploys LLMs directly for cybersecurity tasks. Ref. [20] fine-tunes lightweight models such as LLaMA 3.2-1B, GPT-2, Phi-1, and Gemma-2B for intrusion detection targeting firmware tampering, replay attacks, device impersonation, and unsafe insulin dosage manipulation. Similarly, Ref. [81] evaluates LLM-based contextual reasoning for DoS/DDoS and spoofing detection. Importantly, Ref. [28] introduces a human-in-the-loop safeguard requiring administrator approval before AI-generated firewall rules can be deployed.

Privacy Issues:

The studied papers primarily attempt to achieve privacy preservation through edge processing, federated learning, de-identification, and data minimization. Local and edge-based deployment emerges as the dominant strategy for reducing PHI exposure and dependence on external cloud services.

Edge or on-device deployment appears in at least 16 studies. Ref. [38] deploys quantized DistilBERT models directly on edge devices for HIPAA-oriented clinical note summarization, while Ref. [29] self-hosts open-source LLMs served with Ollama and Matrix end-to-end encryption to eliminate third-party cloud exposure. Similarly, Ref. [30] deploys Llama 3 8B entirely on local CPU infrastructure with sensitivity-aware routing for low, medium, and high-risk inputs.

Several studies further minimize exposure by transmitting only abstracted or feature-level representations rather than raw patient data. For example, Ref. [86] transmits only numeric feature vectors extracted from video streams while protecting stored records using AES-256 encryption and TFDA Part 11 audit logging. Ref. [83] similarly demonstrates privacy-aware local deployment by running Phi-3 locally for nursing stress monitoring.

Federated learning represents the second most common privacy-preserving strategy. For instance, Ref. [47] combines federated learning with differential privacy, secure aggregation, and protection against membership inference and model inversion attacks. On the other hand, Ref. [26] extends this architecture using SGX-isolated inference, hash-based de-identification, and continuous privacy KPI monitoring, while [72] uses frozen embedding encoders to prevent memorization of user content within model weights. However, studies such as [40,62,91] acknowledge that many proposed privacy-preserving mechanisms remain insufficiently validated in real-world deployments.

De-identification and data-minimization strategies are also widely adopted. Ref. [41] restricts LLM inputs to only non-identifiable CGM attributes, while Ref. [24] stores raw Fitbit data exclusively with experimenters and provides only summarized outputs to external APIs while prohibiting diagnostic generation. Ref. [56] uses pseudonymous device identifiers throughout BLE and backend communication, whereas Ref. [26] applies hash-based de-identification prior to federated processing. More adaptive privacy-management approaches appear in [23,32], which implement sensitivity-aware edge-fog-cloud routing based on data sensitivity levels.

The reviewed studies demonstrate a strong preference for architectural privacy preservation through local processing, federated collaboration, and minimized data exposure.

However, most systems remain prototype-level with limited evaluation of long-term privacy guarantees and adversarial robustness, particularly with respect to emerging LLM-specific attack vectors.

Ethical safeguards:

Ethical considerations mainly focus on hallucination risk, fairness and bias, human oversight, explainability, and responsible deployment practices. Although ethical concerns are increasingly acknowledged, ethics is often treated as a supporting design consideration rather than a formally evaluated principal dimension.

The most commonly reported governance mechanism is IRB or ethics committee approval, although these approvals primarily address research-conduct compliance rather than comprehensive clinical deployment safety.

Hallucination and unreliable clinical reasoning are among the most frequently discussed LLM-specific risks in the corpus. Five studies [24,29,59,61,66] explicitly identify fabricated, confabulated, or clinically incorrect outputs. Several mitigation strategies are reported, including data grounding and prompt restrictions [24,29], while Ref. [59] constrains reasoning through a curated medical knowledge graph. In particular, Ref. [29] restricts recommendations to retrieved evidence, whereas Ref. [24] uses prompts that prohibit diagnostic generation. Studies [61,66] further identify hallucination or confabulation as limitations requiring additional safety evaluation. However, several studies acknowledge that these safeguards have not been stress-tested at deployment scale.

Bias and fairness concerns are discussed in seven studies, although empirical evaluation remains limited. Ref. [91] provides the most comprehensive framework by incorporating bias audits, subgroup performance monitoring, diverse datasets, and explainability tools such as SHAP and LIME. In contrast, Ref. [76] highlights demographic and cultural unfairness risks without experimentally validating mitigation strategies. Ref. [37] further raises concerns regarding surveillance and demographic bias in wearable depression-monitoring systems and recommends safeguards including two-factor authentication-secured servers and suicidality escalation alerts. Ref. [19] additionally reports behavioral inconsistency and limited interpretability in medical Mixture of Experts and DRL systems.

Human-in-the-loop safeguards are among the most consistently implemented ethical protection mechanisms. Ref. [49] mandates human override and immutable audit logging for all LLM-driven pipeline actions, Ref. [28] requires administrator approval before deploying AI-generated firewall rules, and Ref. [74] enforces human oversight for all LLM-driven EHR modifications. Ref. [63] further combines OpenAI moderation guardrails with participant check-ins during deployment.

Explainability remains comparatively underexplored. Ref. [91] applies SHAP and LIME for interpretable predictions, while Ref. [53] uses knowledge-graph-grounded reasoning chains to provide traceable inference pathways. Despite these efforts, most systems still function as partially opaque decision-support tools with limited transparency into reasoning processes.

In general, the findings indicate growing ethical awareness regarding hallucination risk, fairness, and human oversight. However, rigorous ethical validation, explainability benchmarking, and accountability mechanisms remain insufficiently developed, highlighting a substantial gap between prototype innovation and clinically trustworthy deployment.

6. Discussion

The integrated analysis of the taxonomy and cross-tabulation results reveals that LLM-IoT healthcare systems are not designed in isolation but are shaped by tightly coupled relationships between application domains, device ecosystems, model roles, and deploy-

ment architectures. Rather than a uniform adoption pattern, the field demonstrates clear specialization based on the operational context of each healthcare application.

6.1. Application-Level Patterns: Why Remote Monitoring and Personal Health Are Widely Researched

Remote patient monitoring and personal health applications are among the most frequently studied application areas in the reviewed literature. This research concentration may reflect the natural compatibility between IoT systems and continuous health-data collection. Remote monitoring systems are already widely used for chronic disease management, older adult care, post-acute recovery, and reducing patient load in hospitals and primary care, and rely on wearable devices and sensors that continuously collect physiological signals, making them a practical foundation for LLM-based interpretation and feedback [94].

Advances in wearable device technologies also explain why monitoring-oriented studies appear more frequently than emergency or infrastructure-focused systems. Wearables collect heart rate, oxygen saturation, blood pressure, body temperature, respiratory rate, posture, and physical activity, and are relatively accessible, non-invasive, and suitable for daily use, which gives researchers easier opportunities to test LLM-IoT systems than regulated medical-grade devices [94]. As a result, the distribution of healthcare applications is shaped by both clinical importance and implementation feasibility: remote monitoring and personal health systems provide continuous data streams, lower deployment barriers, and clear patient-facing use cases, whereas emergency care and clinical infrastructure systems require higher reliability, stronger validation, and stricter regulatory control, which may explain their lower representation.

6.2. LLM Functional Roles: Interpretation, Reasoning, Communication, and Limited Orchestration

The literature indicates LLMs in IoT-driven healthcare are mainly used for data interpretation, reasoning, and conversational support, which is consistent with their documented strengths in patient communication, report generation, clinical decision support, medical education, and extracting information from unstructured health records. Their value lies in synthesizing complex information into coherent, context-aware explanations [95].

In the taxonomy, IoT devices generate raw physiological and behavioral data, and LLMs translate that data into human-readable summaries, alerts, recommendations, and interactive explanations. Therefore, current systems position LLMs as interpretive and assistive layers rather than fully autonomous decision-makers. The limited use of system orchestration and autonomous control reflects the risk-sensitive nature of healthcare, where services are required to remain reliable and resilient by functioning correctly and continuously even under adverse conditions such as faults, cyberattacks, degraded communication, excessive load, or hardware/software failures because failures in healthcare systems may directly affect patient safety, and autonomous LLM-driven control requires a much higher level of trust, explainability, and validation than summarization or communication tasks [96].

6.3. Model Choices: High-Capability Cloud-Based Systems vs. Emerging Open-Source LLMs

Model selection patterns can be explained by architectural differences among LLMs. Mainstream generative LLMs, particularly GPT-family models, are observed to be better suited for semantic and communicative tasks such as report generation and patient interaction, while BERT-based models are often used for classification, knowledge discovery, and low-level task-specific medical language understanding. This aligns with the distinc-

tion that GPT-style models specialize in text generation, whereas BERT-style models use bidirectional encoding that supports deeper language understanding [97].

Consequently, GPT-family models appear more often in cloud-based conversational and reasoning-heavy systems, whereas BERT-family models seem to be more suitable for edge or resource-constrained tasks such as classification, anomaly detection, and short-text interpretation. Similarly, open-source models such as LLaMA seem increasingly relevant in privacy-sensitive contexts because they enable local or institution-controlled deployments. Overall, model choice in LLM-IoT healthcare systems is shaped by the relationship between task type, model capability, deployment environment, privacy expectations, and available computational resources, rather than arbitrary preference.

6.4. Deployment Architectures: Cloud, Edge, Hybrid, and Core Trade-Offs

The architectural distribution reflects a core trade-off between computational capability and healthcare-specific constraints. Cloud-based deployments remain common because large LLMs require substantial computational resources, memory, and storage. However, reliance on centralized cloud infrastructure can introduce concerns related to latency, bandwidth, availability, and transmission of sensitive healthcare data beyond direct institutional control [84].

Edge-based systems address a different set of requirements by moving processing closer to the healthcare data source. Within the reviewed studies, edge implementations demonstrate resource-efficient and low-latency inference in healthcare environments [25,38]. Local deployment is also used to limit dependence on external cloud services and reduce exposure of sensitive healthcare information [29,30]. These findings are consistent with broader edge-LLM literature identifying latency reduction and data locality as important motivations for edge deployment [84].

Hybrid architectures provide an intermediate approach by distributing processing across local and centralized resources. For example, Ref. [23] dynamically allocates medical workloads across edge, fog, and cloud layers according to data sensitivity and computational complexity, while Ref. [32] similarly employs sensitivity-aware tiered processing. Such architectures allow time-critical or sensitive operations to remain closer to the data source while more computationally intensive processing can be transferred to higher-capacity infrastructure. However, this flexibility introduces additional coordination and system complexity.

6.5. Device Ecosystem: Wearable-Centric Designs and Underused Ambient/Clinical Sensing

The dominance of wearable sensors can be explained by the maturity and accessibility of the wearable health ecosystem. IoT healthcare literature shows that sensors and wearable devices are already used for remote monitoring, chronic disease management, personalized care, and patient engagement. These devices can continuously collect vital signs and behavioral data, which creates a strong technical foundation for LLM-based summarization, coaching, and anomaly explanation [94].

In comparison, ambient/environmental sensors and medical-grade clinical devices present greater integration challenges. Medical-grade devices often require stricter certification, clinical workflows, and regulatory compliance. Ambient sensors add contextual value but require additional data fusion, synchronization, and interpretation. The prior smart healthcare literature notes that multimodal medical signal fusion creates challenges such as synchronization, buffering, feature normalization, and classification fusion [98]. Therefore, wearable dominance reflects both practical feasibility and technical alignment with LLM capabilities. Wearables provide data that is continuous, personal, and suitable for explanation, while other device categories require more complex infrastructure and validation.

6.6. Privacy, Security, and Ethical Challenges: Architectural Responses and LLM-Specific Gaps

Privacy and security strongly shape LLM–IoT healthcare system design because these systems process sensitive medical information. The broader IoT healthcare literature identifies privacy, security, interoperability, and standardization as important adoption challenges [2], and similar concerns are visible within the reviewed corpus.

Edge and local architectures address data-exposure concerns by keeping some processing close to the data source. For example, local LLM deployment is used to reduce dependence on third-party cloud services in [29], while Ref. [30] applies sensitivity-aware local processing according to the risk level of the input. Hybrid architectures extend this approach by routing information across edge, fog, and cloud layers according to data sensitivity [23,32].

Federated learning addresses privacy through distributed model training or analysis without requiring institutions or devices to centrally pool raw data. In the reviewed studies, federated systems are combined with differential privacy and secure aggregation to provide additional protection [26,47]. These implementations are consistent with the broader federated-learning literature describing decentralized learning as a privacy-oriented approach for healthcare and edge environments [99].

Blockchain-based systems address a different security objective. Rather than primarily reducing data transmission, reviewed blockchain implementations use distributed ledgers and decentralized storage to support integrity, tamper resistance, and auditability [21,45]. Cross-chain and smart-contract approaches are also explored for secure inter-institutional health-data exchange [85].

6.7. System Design Trade-Offs

The observed deployment patterns across the taxonomy can be interpreted through a set of fundamental system design trade-offs that govern LLM–IoT healthcare integration. Table 19 synthesizes recurring architectural and functional trade-offs observed across the reviewed studies and interpreted in relation to the broader literature.

Table 19. System design trade-offs in LLM–IoT healthcare integration.

Trade-Off	Design Tension	System Implication
Privacy vs. Capability	Cloud LLMs may offer stronger reasoning and interaction but require transmission of sensitive health data; edge/federated models can support data locality and privacy preservation but may reduce model capacity.	Systems must balance privacy protection with access to high-performing LLM capabilities.
Latency vs. Analytical Depth	Edge deployment supports immediate inference for emergency response and real-time monitoring; cloud deployment enables deeper reasoning for latency-tolerant tasks such as clinical decision support.	Deployment choice should align with the urgency and complexity of the healthcare task.
Security vs. Accessibility	Federated learning supports distributed privacy-preserving processing, while blockchain-based safeguards can support integrity and auditability in regulated medical systems but add complexity and overhead; consumer-facing systems prioritize usability and scalability.	Strong security controls may reduce accessibility and ease of adoption.
Complexity vs. Scalability	Hybrid architectures balance privacy and performance but require distributed pipelines, synchronization, and model coordination.	Increased engineering complexity may limit real-world scalability.
Domain Specialization vs. Generalization	Clinical and infrastructure systems favor task-specific LLM roles; consumer applications need conversational and flexible interpretation capabilities.	LLM–IoT systems must balance domain expertise with adaptability across use cases.

Together, these trade-offs illustrate that LLM–IoT healthcare systems are inherently multi-objective, requiring careful balancing of competing priorities rather than optimization along a single dimension.

6.8. Structural Research Gaps

6.8.1. Clinical Validation and Emergency Deployment

The research gaps identified in the taxonomy stem from structural limitations in current LLM–IoT healthcare research. Emergency and critical care systems remain under-represented because they demand low latency, high reliability, real-time decision-making, and strong clinical validation; edge computing is promising here by reducing dependence on cloud connectivity, but deploying LLMs at the edge is constrained by compute and memory limitations [4]. Future studies should therefore evaluate end-to-end response time under realistic emergency conditions and establish whether inference, communication, and alert-generation delays satisfy the time sensitivity of the intended clinical use case.

6.8.2. Autonomous Orchestration and Multimodal Sensing

Autonomous orchestration remains limited because healthcare systems must be reliable and resilient under normal and adverse conditions. Any failure in automated control, monitoring, or response may compromise patient safety. This makes fully autonomous LLM-driven healthcare workflows harder to deploy than assistive systems that keep clinicians or users in the loop [96].

Ambient and environmental sensing is also underused due to multimodal integration complexity: combining physiological, behavioral, and environmental signals requires synchronization, fusion, and context-aware reasoning, which is more demanding than wearable-based physiological monitoring.

6.8.3. Real-World Validation and Deployment Benchmarking

Real-world validation is another major gap. The reviewed studies show varying levels of evaluation, ranging from simulation and synthetic-data testing to evaluation using existing clinical or physiological datasets and, less frequently, direct real-world deployments involving participants or healthcare professionals. Although the use of real clinical data strengthens evaluation, dataset-based testing alone does not necessarily demonstrate reliability under actual clinical workflow conditions. Therefore, given the safety-critical nature of healthcare, LLM–IoT systems should undergo robust evaluation in real-world healthcare environments before clinical deployment, including testing with patients, healthcare professionals, and realistic operational conditions. This validation gap is also visible in privacy, security, and ethics. Although many papers propose safeguards such as federated learning, differential privacy, secure aggregation, local inference, and human oversight, these mechanisms are often not fully implemented or systematically evaluated.

A related deployment-evaluation gap concerns the lack of consistent benchmarking across practical system metrics. Although individual studies evaluate metrics such as inference latency, resource utilization [26,38,68], communication overhead [69], energy consumption, or privacy leakage [47], these dimensions are not assessed consistently across the reviewed systems. Future LLM–IoT healthcare research should therefore report standardized deployment-oriented measures, including end-to-end inference latency, memory and computational footprint, energy consumption, communication cost, quantified privacy leakage and privacy budgets, hallucination and task-error rates, and robustness under both conventional cyberattacks and LLM-specific adversarial attacks. Where privacy-enhancing mechanisms such as differential privacy are used, studies should report the applied privacy parameters and associated utility trade-offs rather than describing privacy preservation

only qualitatively. Such evaluation would enable more meaningful comparison of cloud, edge, hybrid, and federated architectures under realistic healthcare operating conditions.

6.8.4. LLM-Specific Security and Adversarial Robustness

Current studies still rely mainly on conventional IoT security controls such as encryption, access control, authentication, and secure communication, but the introduction of language-model components creates additional attack surfaces that are not fully addressed by these mechanisms. Risks such as prompt injection, jailbreak attacks, hallucination, model extraction, poisoning, membership inference, malicious RAG content, adversarial input manipulation, unsafe model responses, and unsafe autonomous actions remain insufficiently or inconsistently evaluated across the reviewed literature. Although a small number of studies examine related risks such as poisoning or membership inference, systematic empirical validation across these LLM-specific attack vectors remains limited. This is especially important because LLM–IoT healthcare systems may influence alerts, recommendations, patient communication, device interpretation, or workflow actions.

6.8.5. Ethics, Fairness, and Regulation

Ethical and regulatory validation also remains underdeveloped. Ethics is often addressed through IRB approval, bias discussion, hallucination awareness, or human-in-the-loop design, but few studies provide systematic fairness evaluation, safety auditing, subgroup-specific performance measures across demographic groups, explainability assessment, or formal accountability mechanisms. Similarly, regulatory discussion is narrow and mostly limited to frameworks such as HIPAA, GDPR, HITRUST CSF, EU MDR, and TFDA Part 11. These frameworks are mostly cited as motivations for design choices rather than evidence of demonstrated compliance, leaving broader international regulatory obligations and deployment-readiness requirements insufficiently addressed.

6.8.6. Design Implications

LLM–IoT healthcare systems are evolving along a capability–constraint spectrum, while the adoption of rapidly advancing LLM capabilities within healthcare IoT remains uneven. Emerging advances in efficient edge inference, stronger guardrails against LLM-specific vulnerabilities, privacy-preserving adaptation, and reliable autonomous reasoning should be evaluated and translated into healthcare applications more rapidly, while maintaining rigorous safety and clinical-validation requirements. Cloud-based systems offer greater computational capacity for complex reasoning and communication but can introduce latency and data-control concerns [84]. Edge implementations are used to support local and responsive processing in resource-constrained settings [25,38], while federated systems emphasize distributed privacy-preserving processing [26,47]. Hybrid architectures provide a middle ground by combining local processing with access to higher-capacity cloud resources [23].

Effective design depends on aligning the healthcare application, device ecosystem, LLM role, deployment architecture, and privacy/security requirements, and on navigating the multi-objective trade-offs identified above. In this sense, success depends less on deploying the most powerful model and more on building context-aware, reliable, privacy-preserving, and clinically meaningful systems that reflect the integrated patterns revealed by this review.

6.9. Limitations of Our Approach

In this section we discuss the limitations of our approach. First, the literature search was primarily based on Google Scholar and Scholar Labs, which may not capture all relevant studies indexed in specialized databases such as Scopus, Web of Science, PubMed,

IEEE Xplore, or ACM Digital Library that may contain studies not readily retrieved through Google Scholar. Second, searches were limited to English-language studies, which may exclude relevant non-English publications. Third, we did not utilize independent screening and quality assessment by multiple coders independently, the screening process was conducted by a junior researcher and iteratively validated with a senior researcher. Consequently, no inter-reviewer agreement statistic was calculated. We also note that study-level quality assessment was not conducted; the quantitative synthesis reported reflects the frequency and architectural focus of current research rather than the clinical effectiveness or methodological rigor of the individual systems proposed in each literature.

Fourth, during preliminary testing with different natural-language prompts, Scholar Labs was observed to process a maximum of approximately 300 records for a single search. However, for the finalized prompt reported in this review, Scholar Labs screened approximately 250 indexed articles and returned 36 filtered results. To maintain methodological comparability, the traditional Google Scholar search was therefore limited to the first 300 results, representing the maximum screening range observed for the AI-assisted method. Consequently, potentially relevant studies beyond this threshold may not have been identified. Fifth, the AI-assisted search method has limited repeatability due to its non-deterministic nature. The retrieved studies, their ranking, and the underlying result index may vary among users submitting the same prompt and may also change when the same user repeats the search at different times. Thus, identical AI prompts may not consistently produce the same set or ordering of results. To reduce potential selection and coding bias, predefined inclusion and exclusion criteria and a structured data-extraction spreadsheet were applied consistently throughout the review.

The publication status of the studies included in the review is presented in Table 20.

Table 20. Peer-reviewed and preprint studies included in this review.

Peer-Reviewed	[20–30,32–35,38–43,45–48,51–55,57–65,67–73,76–82,85,86,88–90,92,93,100]
Preprint Studies	[19,31,36,37,44,49,50,56,66,74,75,83,87,91]

7. Conclusions

This systematic review examined the integration of Large Language Models with IoT-driven healthcare systems through the analysis of 75 research articles. This study provides both quantitative and conceptual insights into the current state of this emerging field. This review offers five primary contributions.

First, the comparative analysis of AI-powered prompt-based search and traditional keyword-based search shows that AI-augmented literature discovery can improve the precision and relevance of retrieved studies. In our comparative sample, prompt-based search yielded higher initial precision (86% vs. 42%), though supplementary traditional searches remain essential to ensure broader literature recall and compensate for AI non-determinism. Therefore, because the AI-assisted approach remains at an early stage, limitations related to article coverage and reproducibility should be addressed through hybrid search strategies that combine AI-assisted and traditional methods.

Second, the proposed multi-dimensional taxonomy demonstrates that LLM–IoT healthcare systems are not evenly distributed across application areas. The literature is concentrated in remote patient monitoring and personal health applications, suggesting that current research is shaped more by data availability and implementation feasibility than by clinical criticality. The taxonomy provides a structured basis for examining how healthcare domains, IoT devices, LLM roles, and deployment architectures interact.

Third, the cross-dimensional analysis using quantitative cross-tabulation reveals that system design is highly context-dependent though clear relationships emerge among healthcare application types, IoT device ecosystems, LLM functions, and architectural choices. These findings indicate that effective LLM–IoT integration requires coordinated design across multiple dimensions rather than isolated technological decisions.

Fourth, the assessment of privacy, security, and ethical challenges indicates that although encryption, federated learning, and edge processing are frequently discussed or adopted, security is often treated as a design feature rather than a rigorously evaluated system property. LLM-specific risks, including hallucination, prompt injection, adversarial manipulation, and unsafe outputs, remain insufficiently addressed, which reveals a substantial gap between system functionality and operational robustness.

Fifth, the identified research gaps and future directions show that the field remains at an early stage of maturity. Major limitations include the underrepresentation of emergency and critical care applications, limited autonomous orchestration, insufficient multimodal sensing integration, and limited real-world clinical validation. Future research should move beyond prototype development where security and privacy are merely added features, toward scalable, secure, privacy-aware, and clinically validated systems.

The findings show that LLM–IoT healthcare systems evolve along a capability–constraint spectrum in which design decisions are shaped by trade-offs among privacy, performance, latency, scalability, and reliability. Their effectiveness depends not only on advanced model capabilities but also on the alignment of application requirements, device ecosystems, architectural design, and security considerations. Integrating methodological comparison, taxonomy development, cross-dimensional analysis, and critical evaluation of security and research gaps, this study provides a foundation for developing intelligent, context-adaptive, and privacy-preserving LLM–IoT healthcare systems for real-world deployment.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/iot7030075/s1>, Table S1: PRISMA 2020 checklist; Table S2: Structured multi-label coding taxonomy and extracted data.

Author Contributions: Conceptualization, Y.K., P.R.M.; methodology, Y.K., P.R.M. and S.M.; writing—original draft preparation, Y.K., P.R.M. and S.M.; writing—review and editing, Y.K., P.R.M. and S.M.; visualization, P.R.M.; supervision, Y.K. and S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The structured multi-label coding taxonomy and extracted data are publicly available in the Supplementary Materials to improve transparency and support reproducibility.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Acronym	Description
DRL	Deep Reinforcement Learning
LoRA	Low-Rank Adaptation
IPFS	InterPlanetary File System
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Unit
DcDFS	Dependency-correlated Data Fusion Scheme

IRB	Institutional Review Board
HIPAA	Health Insurance Portability and Accountability Act
GDPR	General Data Protection Regulation
HITRUST	Health Information Trust Alliance
CSF	Common Security Framework
EU MDR	European Union Medical Device Regulation
TFDA	Taiwan Food and Drug Administration
DPIA	Data Protection Impact Assessment
AES	Advanced Encryption Standard
TLS	Transport Layer Security
RBAC	Role-Based Access Control
IAM	Identity and Access Management
RAG	Retrieval-Augmented Generation
SHAP	SHapley Additive exPlanations
LIME	Local Interpretable Model-Agnostic Explanations
PII/PHI	Personally Identifiable Information/Protected Health Information
DoS/DDoS	Denial-of-Service/Distributed Denial-of-Service
SGX	Software Guard Extensions
CGM	Continuous Glucose Monitoring
BLE	Bluetooth Low Energy
MoE	Mixture of Experts

References

1. Shaik, T.; Tao, X.; Higgins, N.; Li, L.; Gururajan, R.; Zhou, X.; Acharya, U.R. Remote Patient Monitoring Using Artificial Intelligence: Current State, Applications, and Challenges. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2023**, *13*, e1485. [\[CrossRef\]](#)
2. Li, C.; Wang, J.; Wang, S.; Zhang, Y. A Review of IoT Applications in Healthcare. *Neurocomputing* **2024**, *565*, 127017. [\[CrossRef\]](#)
3. Shah, S.; Madni, S.H.H.; Hashim, S.Z.M.; Samah, A.A.; Ayaz, M.; Ullah, S.; Faheem, M. Requirements Evaluation Framework for a Phase-Driven AIoT Architecture for Remote Patient Monitoring. *J. Eng.* **2025**, *2025*, e70149. [\[CrossRef\]](#)
4. Greco, L.; Percannella, G.; Ritrovato, P.; Tortorella, F.; Vento, M. Trends in IoT Based Solutions for Health Care: Moving AI to the Edge. *Pattern Recognit. Lett.* **2020**, *135*, 346–353. [\[CrossRef\]](#)
5. Baker, S.; Xiang, W. Artificial Intelligence of Things for Smarter Healthcare: A Survey of Advancements, Challenges, and Opportunities. *IEEE Commun. Surv. Tutor.* **2023**, *25*, 1261–1293. [\[CrossRef\]](#)
6. Alsabah, M.; Naser, M.A.; Albahri, A.S.; Albahri, O.S.; Alamoodi, A.H.; Abdulhussain, S.H.; Alzubaidi, L. A Comprehensive Review on Key Technologies toward Smart Healthcare Systems Based IoT: Technical Aspects, Challenges and Future Directions. *Artif. Intell. Rev.* **2025**, *58*, 343. [\[CrossRef\]](#)
7. Al Nazi, Z.; Peng, W. Large Language Models in Healthcare and Medical Domain: A Review. *Informatics* **2024**, *11*, 57. [\[CrossRef\]](#)
8. Bedi, S.; Liu, Y.; Orr-Ewing, L.; Dash, D.; Koyejo, S.; Callahan, A.; Fries, J.A.; Wornow, M.; Swaminathan, A.; Lehmann, L.S.; et al. Testing and Evaluation of Health Care Applications of Large Language Models: A Systematic Review. *JAMA* **2025**, *333*, 319–328. [\[CrossRef\]](#)
9. Ferreira, J.C.; Correia, R.; Rosa, I.M.D.; Ferreira, J.C.; Correia, R.; Rosa, I.M.D. Multi-Component Pipeline LLMs for Interoperable Healthcare Data: A Scoping Review from Clinical Summarization to Multimodal Integration. In *Machine Learning and Data Mining*; Aceves-Fernández, M.A., Ed.; IntechOpen: London, UK, 2026.
10. Ferrara, E. Large Language Models for Wearable Sensor-Based Human Activity Recognition, Health Monitoring, and Behavioral Modeling: A Survey of Early Trends, Datasets, and Challenges. *Sensors* **2024**, *24*, 5045. [\[CrossRef\]](#)
11. Kalejahi, B.K.; Khalimova, K.; Rajabi, M.J. Large Language and Generative Foundation Models for Cloud-Enabled Medical Imaging and Internet of Medical Things: A PRISMA Systematic Review of Architectures, Security, and Deployment. *Res. Sq.* **2026**, *PREPRINT (Version 1)*. [\[CrossRef\]](#)
12. Hasan Khan, M.; Nayfeh, A.; Masood, M.; Al-Shaikhi, A.A.; Mahboob, M.; Rahman, U.; Al-Naffouri, Y. ECG Foundation Models and Medical LLMs for Agentic Cardiovascular Intelligence at the Edge: A Review and Outlook. *arXiv* **2026**, arXiv:2604.02501.
13. Halevi, G.; Moed, H.; Bar-Ilan, J. Suitability of Google Scholar as a Source of Scientific Information and as a Source of Data for Scientific Evaluation—Review of the Literature. *J. Informetr.* **2017**, *11*, 823–834. [\[CrossRef\]](#)
14. Gusenbauer, M. Google Scholar to Overshadow Them All? Comparing the Sizes of 12 Academic Search Engines and Bibliographic Databases. *Scientometrics* **2018**, *118*, 177–214. [\[CrossRef\]](#)

15. Brereton, P.; Kitchenham, B.A.; Budgen, D.; Turner, M.; Khalil, M. Lessons from Applying the Systematic Literature Review Process within the Software Engineering Domain. *J. Syst. Softw.* **2007**, *80*, 571–583. [\[CrossRef\]](#)
16. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ* **2021**, *372*, n71. [\[CrossRef\]](#)
17. Shen, Z.; Ma, H.; Wang, K. A Web-Scale System for Scientific Knowledge Exploration. In *Proceedings of ACL 2018, System Demonstrations, Annual Meeting of the Association for Computational Linguistics (2018)*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 87–92. [\[CrossRef\]](#)
18. Song, Y.; Wang, G.; Li, S.; Lin, B.Y. The Good, The Bad, and The Greedy: Evaluation of LLMs Should Not Ignore Non-Determinism. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2025; Volume 1, pp. 4195–4206. [\[CrossRef\]](#)
19. Gao, Y.; Ye, Z.; Xiao, M.; Member, S.; Xiao, Y.; In Kim, D. Guiding IoT-Based Healthcare Alert Systems with Large Language Models. *arXiv* **2024**, arXiv:2408.13071.
20. Pawana, I.W.A.J.; Astillo, P.V.; You, I. Lightweight LLM-Based Anomaly Detection Framework for Securing IoTMD Enabled Diabetes Management Control Systems. *IEEE J. Biomed. Health Inform.* **2025**, *30*, 6356–6367. [\[CrossRef\]](#)
21. Kumari, A.; Shah, M.H.; Sheth, P.V. S-Health: LLM-Driven and Lightweight Blockchain-Based Secure Framework for Healthcare 5.0. In *Proceedings of the 2024 IEEE Globecom Workshops (GC Wkshps), Cape Town, South Africa, 12 August 2025*; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2025; pp. 1–6.
22. Yang, B.; Jiang, S.; Xu, L.; Liu, K.; Li, H.; Xing, G.; Chen, H.; Jiang, X.; Yan, Z. DrHouse: An LLM-Empowered Diagnostic Reasoning System through Harnessing Outcomes from Sensor Data and Expert Knowledge. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*; Association for Computing Machinery: New York, NY, USA, 2024; Volume 8. [\[CrossRef\]](#)
23. Zagar, P.; Ravi, V.; Aalami, L.; Krusche, S.; Aalami, O.; Schmiedmayer, P. Dynamic Fog Computing for Enhanced LLM Execution in Medical Applications. *Smart Health* **2025**, *36*, 100577. [\[CrossRef\]](#)
24. Fang, C.M.; Danry, V.; Whitmore, N.; Bao, A.; Hutchison, A.; Pierce, C.; Maes, P. PhysioLLM: Supporting Personalized Health Insights with Wearables and Large Language Models. In *Proceedings of the 2024 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), Houston, TX, USA, 10–13 November 2024*. [\[CrossRef\]](#)
25. Vinodh Kumar, M.; Ramesh, G.P. Smart IoT Based Health Care Environment for an Effective Information Sharing Using Resource Constraint LLM Models. *J. Smart Internet Things* **2024**, *2024*, 133–147. [\[CrossRef\]](#)
26. Karamanlioğlu, A.; Demirel, B.; Tural, O.; Doğan, O.T.; Alpaslan, F.N. Privacy-Preserving Clinical Decision Support for Emergency Triage Using LLMs: System Architecture and Real-World Evaluation. *Appl. Sci.* **2025**, *15*, 8412. [\[CrossRef\]](#)
27. Alsaif, K.; Albeshri, A.; Khemakhem, M.; Eassa, F. Healthcare 4.0: A Large Language Model-Based Blockchain Framework for Medical Device Fault Detection and Diagnostics. *Int. J. Adv. Comput. Sci. Appl.* **2025**, *16*, 980–992. [\[CrossRef\]](#)
28. Gosálvez-White, I.; Rodríguez-Martín, N.; Barajas-García, N.; Mena-Gallardo, C.; García-Teodoro, P. MEDIotWALL: Securing Smart Healthcare Environments Through IoT Firewalls. *Sensors* **2025**, *25*, 6235. [\[CrossRef\]](#)
29. Pap, I.A.; Oniga, S. EHealth Assistant AI Chatbot Using a Large Language Model to Provide Personalized Answers through Secure Decentralized Communication. *Sensors* **2024**, *24*, 6140. [\[CrossRef\]](#)
30. Alotaibi, O.; Tomy, S.; Pardede, E. Enhancing Decision-Making with Large Language Models: Context and Sensitivity Analysis. *Evol. Intell.* **2025**, *18*, 126. [\[CrossRef\]](#)
31. Lee, D.Y.; Weakley, A.; Wei, H.; Cardona, D.; Pan, S. Home Health System Deployment Experience for Geriatric Care Remote Monitoring. *arXiv* **2026**, arXiv:2601.17608.
32. Dall’Ora, N.; Felli, L.; Aldegheri, S.; Vicino, N.; Giuliano, R. LumiCare: A Context-Aware Mobile System for Alzheimer’s Patients Integrating AI Agents and 6G. *Electronics* **2025**, *14*, 3516. [\[CrossRef\]](#)
33. Mahmoudi, A.; Farahani, G.; Domanski, P.; Farahani, B.; Firouzi, F.; Chakrabarty, K. DiabLLM: An LLM-Based Framework for Blood Glucose Prediction in Type 1 Diabetes. *IEEE J. Biomed. Health Inform.* **2026**, *30*, 6446–6459. [\[CrossRef\]](#)
34. Liu, S.; Chen, Y.; Wang, J. A Lightweight Large Language Model for Personal Sleep Quality Estimation. In *Proceedings of the 2025 IEEE 20th Conference on Industrial Electronics and Applications (ICIEA), Yantai, China, 3–6 August 2025*; pp. 1–6. [\[CrossRef\]](#)
35. Liu, Z.; Chen, C.; Cao, J.; Pan, M.; Liu, J.; Li, N.; Miao, F.; Li, Y. Large Language Models for Cuffless Blood Pressure Measurement From Wearable Biosignals. In *Proceedings of the 15th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, Shenzhen, China, 22–25 November 2024*. [\[CrossRef\]](#)
36. Zhang, A.; Pillai, A.; Campbell, A.; Jacobson, N.C. MotionTeller: Multi-Modal Integration of Wearable Time-Series with LLMs for Health and Behavioral Understanding. *arXiv* **2025**, arXiv:2512.21506.
37. Pillai, A.; Spathis, D.; Nepal, S.; Collins, A.C.; Mackin, D.M.; Heinz, M.V.; Griffin, T.Z.; Jacobson, N.C.; Campbell, A. Time2Lang: Bridging Time-Series Foundation Models and Large Language Models for Health Sensing Beyond Prompting. *Proc. Mach. Learn. Res.* **2025**, *287*, 2025. [\[CrossRef\]](#)

38. Syed, A.A.; Kumar Sambamurthy, P.; Sajid Mohammed, A.; Mamodiya, U. Optimized Edge-AI Streaming for Smart Healthcare and IoT Using Kafka, Large Language Model Summarization, and On-Device Analytics. In *Proceedings of the International Conference on Computing, Intelligence, and Application, CIACON 2025*; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2025.
39. Kumar, R.M.; Adluru, S. LLM-Based Remote Patient Monitoring for Elderly Care Using IoT. In *Proceedings of the 2025 International Conference on Engineering Innovations and Technologies (ICoEIT), Bhopal, India, 4–5 July 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 1168–1174.
40. Raghavan, P.; Sura, R.; Taneja, A.; Tiwari, A. Federated Learning for Privacy-Preserving Multi-Agent Systems in IoT-Based Healthcare, Guided by Explainable LLMs. In *Proceedings of Fifth International Conference on Computing and Communication Networks. ICCCN 2025*; Nguyen, G.-N., Swaroop, A., Shukla, P., Eds.; Springer: Cham, Switzerland, 2026; pp. 566–584.
41. Gaitán-Guerrero, J.F.; Martínez-Cruz, C.; Espinilla, M.; Díaz-Jiménez, D.; López, J.L. A Novel Fine-Tuning and Evaluation Methodology for Large Language Models on IoT Raw Data Summaries (LLM-RawDMeth): A Joint Perspective in Diabetes Care. *Comput. Methods Programs Biomed.* **2025**, *269*, 108878. [\[CrossRef\]](#)
42. Zhang, T.; Xu, X.; Wang, Y.; Peng, R.; Kadoch, M. Optimizing Remote Medical Services with AIoT: Integration of Large Language Models and 6G Edge Computing. In *Proceedings of the Communications in Computer and Information Science*; Springer Science and Business Media Deutschland GmbH: Berlin/Heidelberg, Germany, 2025; Volume 2416 CCIS, pp. 278–294.
43. Hua, X.; Chu, Y.; Ma, Y.; Hao, E.; Cao, H.; Wang, B. Research on the Framework Design of an AI-IoT Integrated Evaluation Large Language Model Focused on Children’s Health and Disease Prevention at High Altitude. In *Proceedings of the 2025 5th International Conference on Artificial Intelligence, Big Data and Algorithms, CAIBDA 2025*; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2025; pp. 1172–1175.
44. Srivastava, A.K.; Lohia, A.; Raj, A.; Kumar, A.; Harsh, M.; Lakhotia, S.S. HealthHub: An IoT-Enabled Real-Time Health Monitoring System Using Retrieval-Augmented Generation for Comprehensive Healthcare Management. *Strad Res.* **2025**, *12*, 20–24. [\[CrossRef\]](#)
45. Islayem, R.; Musamih, A.; Salah, K.; Jayaraman, R.; Yaqoob, I. Enhancing Medical Digital Twins within Metaverse Using Blockchain, NFTs and LLMs. *Internet Things* **2025**, *32*, 101648. [\[CrossRef\]](#)
46. Liu, D.; Zheng, Y. Design of an Edge Computing Otological Disease Diagnosis System Based on Deep Learning and LLM. In *Proceedings of the International Conference on Machine Learning, Neural Networks, and Computer Software (MLNCS 2025), Hong Kong, China, 15–17 August 2025*; SPIE: Bellingham, WA, USA, 2025; Volume 13940, pp. 319–326.
47. Bououdina, S. Trustworthy Privacy-Preserving Federated Large Language Models for the Internet of Medical Things. In *Proceedings of the 2025 IEEE 36th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Istanbul, Turkiye, 12 December 2025*; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2025; pp. 1–7.
48. An, T.; Zhou, Y.; Zou, H.; Yang, J. IoT-LLM: A Framework for Enhancing Large Language Model Reasoning from Real-World Sensor Data. *Patterns* **2026**, *7*, 101429. [\[CrossRef\]](#)
49. Tahera, M.A.; Sidhu, K.S.; Dass, S.; Saha, S. SoK: Privacy-aware LLM in Healthcare: Threat Model, Privacy Techniques, Challenges and Recommendations. *arXiv* **2026**, arXiv:2601.10004.
50. Wang, M.H.; Zeng, P.; Li, X.; Chong, K.K.-L.; Yu, X. Enhancing Public Healthcare with IoT and AI: A Comprehensive Study on Improving Dry Eye Disease Diagnosis (Version 1). *Res. Sq.* **2024**, preprint. [\[CrossRef\]](#)
51. Mehdi, S.; Kouah, S.; Saighi, A.; Zertal, S. LLM-Cardio: A Large Language Model-Based Assistant for Cardiovascular Health Inquiry and Diagnostic Support Using Wearable Data. *Ing. Syst. Inf.* **2025**, *30*, 3225–3234. [\[CrossRef\]](#)
52. Patil, S.; Ramya, T.A. Chat-GPT Powered IoT Devices Using Regularizing the Data for an Efficient Management Systems. *J. Smart Internet Things* **2024**, *2024*, 180. [\[CrossRef\]](#)
53. Zhao, X.; Liu, S.; Yang, S.-Y.; Miao, C. A Smart Multimodal Healthcare Copilot with Powerful LLM Reasoning. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, Montreal, QC, Canada, 16–22 August 2025*; pp. 11132–11136. [\[CrossRef\]](#)
54. Rahman, M.A.; Al-Hazaa, S. Next-Generation Virtual Hospital: Integrating Discriminative and Large Multi-Modal Generative AI for Personalized Healthcare. In *Proceedings of the GLOBECOM 2024, 2024 IEEE Global Communications Conference, Cape Town, South Africa, 8–12 December 2024*; pp. 3509–3514. [\[CrossRef\]](#)
55. Englhhardt, Z.; Ma, C.; Morris, M.E.; Chang, C.C.; Xu, X.O.; Qin, L.; McDuff, D.; Liu, X.; Patel, S.; Iyer, V. From Classification to Clinical Insights: Towards Analyzing and Reasoning About Mobile and Behavioral Health Data with Large Language Models. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*; Association for Computing Machinery: New York, NY, USA, 2024; Volume 8. [\[CrossRef\]](#)
56. Ahsan, M.S.; Shaikh, A.S.; Zahid, M.; Irfan, U.; Mustafa, M.; Bhatti, N.A.; Alizai, M.H. LLM-Enhanced Wearables for Comprehensive Health Guidance in LMICs. *arXiv* **2026**, arXiv:2602.08701.

57. Chen, W.; Cheng, J.; Wang, L.; Zhao, W.; Matusik, W. Sensor2Text: Enabling Natural Language Interactions for Daily Activity Tracking Using Wearable Sensors. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*; Association for Computing Machinery: New York, NY, USA, 2024; Volume 8. [\[CrossRef\]](#)
58. Yu, X.; Hu, L.; Reichman, B.; Chu, D.; Chandrupatla, R.; Zhang, X.; Heck, L.; Rosing, T.S. SensorChat: Answering Qualitative and Quantitative Questions during Long-Term Multimodal Sensor Interactions. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*; Association for Computing Machinery: New York, NY, USA, 2025; Volume 9, pp. 1–35. [\[CrossRef\]](#)
59. Zhou, Z.; Yang, Y.; Ren, T. IHILLM-RAG: A Safe and Private Medical Large Language Model Based on Intelligent Hardware Interaction and Retrieval-Augmented Generation (RAG). In *Proceedings of the Fourth International Computational Imaging Conference (CITA 2024)*; SPIE: Bellingham, WA, USA, 2025; Volume 13542, pp. 1070–1079.
60. Vijayalakshmi, S.; Alwin, J.; Lekha, J. IoT-Enabled Intelligent Health Care Screen System for Long-Time Screen Users. *Eng. Proc.* **2024**, *82*, 96. [\[CrossRef\]](#)
61. Khasentino, J.; Belyaeva, A.; Liu, X.; Yang, Z.; Furlotte, N.A.; Lee, C.; Schenck, E.; Patel, Y.; Cui, J.; Schneider, L.D.; et al. A Personal Health Large Language Model for Sleep and Fitness Coaching. *Nat. Med.* **2025**, *31*, 3394–3403. [\[CrossRef\]](#)
62. Choudhari, R.D.; Khodwe, G. *Predictive Wellness Analytics Through Hybrid Sensor-LLM Integration*; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2025; pp. 1–7.
63. Wang, X.; Griffith, J.; Adler, D.A.; Castillo, J.; Choudhury, T.; Wang, F. Exploring Personalized Health Support through Data-Driven, Theory-Guided LLMs: A Case Study in Sleep Health. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*; Association for Computing Machinery: New York, NY, USA, 2025; pp. 1–15. [\[CrossRef\]](#)
64. Corda, E.; Massa, S.M.; Riboni, D. Context-Aware Behavioral Tips to Improve Sleep Quality via Machine Learning and Large Language Models. *Futur. Internet* **2024**, *16*, 46. [\[CrossRef\]](#)
65. Zheng, H.; Ai, X.; Liu, X.; Xing, X.; Xu, X. Profile-Aided Distillation Framework for Personalized Sleep Analysis with Compact Models Using LLM-Guided Synthetic Data. *Front. Physiol.* **2026**, *16*, 1678364. [\[CrossRef\]](#)
66. Khaokaew, Y.; Ji, K.; Nguyen, T.H.; Kegalle, H.; Alaofi, M.; Xue, H.; Salim, F.D. ZzzGPT: An Interactive GPT Approach to Enhance Sleep Quality. *arXiv* **2024**, arXiv:2310.16242.
67. Dharani, M.; Udhaya Kumar, M.; Latha, B.; Tanusree, S.R.; Sivashanmugha, V.; Vijay, M. Design and Development of Healthcare IoT Based Bots Using Different LLM Models: A Best Method Performance Evaluation. In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies-ICRDICCT'25*; SciTePress: Setúbal, Portugal, 2025; pp. 40–47.
68. Kurma, S.; Singh, K.; Paul, A.; Mumtaz, S.; Li, C.-P. Dual-LLM Integration with Reconfigurable Intelligent Surface for Healthcare Networks. *IEEE Trans. Commun.* **2025**, *73*, 11323–11339. [\[CrossRef\]](#)
69. Kurma, S.; Paul, A.; Singh, K.; Dev, K.; Li, C.P. LLMs for Resource Allocation in Next-Gen RIS-Aided Healthcare Wireless Networks. In *Proceedings of the IEEE Conference on Computer Communications Workshops, INFOCOM WKSHPS 2025*; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2025; pp. 1–6.
70. Nahid, N.; Ahad, M.A.R.; Inoue, S. Context-Aware Alerting in Elderly Care Facilities: A Hybrid Framework Integrating LLM Reasoning with Rule-Based Logic. *Sensors* **2025**, *25*, 6560. [\[CrossRef\]](#)
71. Marques, A.G.; Andrade, C.E.; Da Costa Nascimento, J.J.; Adelino Rodrigues, Y.O.; Severiano, G.F.B.; Gomes Abreu Filho, P.R.; Carlos Mauricio, C.M.J.; De Freitas Souza, L.F. Wearable Stroke Alert System-New Health of Things Approach Based on Generative AI and Datafusion for Real-Time Stroke Monitoring. In *Proceedings of the Brazilian Symposium on Computing System Engineering, SBESC*; IEEE Computer Society: Piscataway, NJ, USA, 2024; pp. 1–6.
72. Shukla, T.; Gupta, A. LLM-Augmented Multimodal Fusion for Accurate and Real-Time Wearable Event Prediction. In *Proceedings of the 9th SLAAI International Conference on Artificial Intelligence (SLAAI-ICAI), Colombo, Sri Lanka*; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2026; pp. 1–6.
73. Ghazal, T.M.; Islam, S.; Hasan, M.K.; Abu-Shareha, A.A.; Mokhtar, U.A.; Khan, M.A.; Baili, J.; Saeed, A.Q.; Bhatt, M.W.; Ahmad, M. Generative Federated Learning with Small and Large Models in Consumer Electronics for Privacy-Preserving Data Fusion in Healthcare Internet of Things. *IEEE Trans. Consum. Electron.* **2025**, *71*, 3413–3430. [\[CrossRef\]](#)
74. Borkowski, A.; Ben-Ari, A. Multi-Agent AI Systems in Healthcare: Technical and Clinical Analysis. 2024. Available online: https://www.preprints.org/frontend/manuscript/3a368827205a2bd53ef513f8b4c037ea/download_pub (accessed on 2 September 2026).
75. Chen, J.; Deng, L.; Qiu, Y.; Zhao, P.; Lei, H.; Song, J.; Wang, X. Llm-Based Automated Modeling in Symbolic Execution for Securing Medical Software. *SSRN* **2024**. [\[CrossRef\]](#)
76. Park, J.; Kim, Y.; Song, J.; Lee, H. Contextual Health State Inference from Lifelog Data Using LLM. In *Proceedings of the International Conference on ICT Convergence*; IEEE Computer Society: Piscataway, NJ, USA, 2024; pp. 816–821.

77. Yahia, Y.A.; Salem, O.; Mehaoua, A. Autonomous and Context-Adaptive Defense Mechanisms for Identity Access Management in Internet of Medical Things. In *Proceedings of the 2025 8th Conference on Cloud and Internet of Things (CIoT)*, London, UK, 29–31 October 2025; pp. 1–7. [\[CrossRef\]](#)
78. Tian, Y.; Ren, X.; Wang, Z.; Gungor, O.; Yu, X.; Rosing, T. DailyLLM: Context-Aware Activity Log Generation Using Multi-Modal Sensors and LLMs. In *Proceedings of the 2025 IEEE 22nd International Conference on Mobile Ad-Hoc and Smart Systems, MASS 2025, Chicago, IL, USA, 6 October 2025*; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2025; pp. 372–380.
79. López, C.Á.; Pérez, A.R.; Alonso-Rincón, R.; Carrera, A.; González, S.R. Digital Twins and Virtual Assistants for Proactive Long-Term Care- The SALUS Architecture. In *Proceedings of the International Conference on Practical Applications of Agents and Multi-Agent Systems*; Nongailard, A., Caron, A., González-Briones, A., Fernández, A., Durães, D., Sharaf, N., Eds.; Springer: Cham, Switzerland, 2025; pp. 42–53.
80. Lee, J.; Jung, G.; Chang, Y.; Kang, Y.; Kim, J. ChatGPT Powered Digital Healthcare System. In *Proceedings of the 14th International Conference on ICT Convergence, Jeju Island, Republic of Korea, 11–13 October 2023*; IEEE Computer Society: Piscataway, NJ, USA, 2023; pp. 1189–1193.
81. Sharma, R.; Mehta, M. Adaptive and Explainable AI Agents for Anomaly Detection in Critical IoT Infrastructure Using LLM-Enhanced Contextual Reasoning. In *Proceedings of the Fifth International Conference on Computing and Communication Networks*; Nguyen, G.-N., Swaroop, A., Shukla, P., Eds.; Springer: Cham, Switzerland, 2026; pp. 247–266.
82. Yu, C.-H.; Masum, M. HybridSense-LLM: A Structured Multimodal Framework for Large-Language-Model-Based Wellness Prediction from Wearable Sensors with Contextual Self-Reports. *Bioengineering* **2026**, *13*, 120. [\[CrossRef\]](#)
83. Fukazawa, Y.; Kodaka, M. LLM-Based Intermediate Interpretations for Predicting Nurse Stress and Its Causes from Step Count Data (Version 1). *Res. Sq.* **2025**, preprint. [\[CrossRef\]](#)
84. Friha, O.; Amine Ferrag, M.; Kantarci, B.; Cakmak, B.; Ozgun, A.; Ghoulmi-Zine, N. LLM-Based Edge Intelligence: A Comprehensive Survey on Architectures, Applications, Security and Trustworthiness. *IEEE Open J. Commun. Soc.* **2024**, *5*, 5799–5856. [\[CrossRef\]](#)
85. Su, C.; Wen, J.; Kang, J.; Wang, Y.; Su, Y.; Pan, H.; Zhong, Z.; Shamim Hossain, M.; Su, J. Hybrid RAG-Empowered Multi-Modal LLM for Secure Data Management in Internet of Medical Things: A Diffusion-Based Contract Approach. *IEEE Internet Things J.* **2025**, *12*, 13428–13440. [\[CrossRef\]](#)
86. Huang, G.-H.; Lin, Y.-P.; Shih, S.-F. Implementation of Multimodal LLM Agents and Biomechanical Analysis for Remote Elderly Healthcare in Taiwan. *Supply Chain Sustain. Res.* **2025**, *4*, 83–93.
87. Chandra, R.; Agarwal, S.; Singh, N. Real-Time Health Analytics Using Ontology-Driven Complex Event Processing and LLM Reasoning: A Tuberculosis Case Study. *arXiv* **2025**, arXiv:2510.09646.
88. Khan, F.; Tulon, T.A.; Masrur, N.; Hasan, M.J.; Badrul, T.; Islam, A. Development and Evaluation of ShasthoBondhu: MHealth App for Guiding Emergency Remote Healthcare with Wearable IoT and AI Fusion. In *Proceedings of the Advances in Science and Engineering Technology International Conferences, ASET, Abu Dhabi, United Arab Emirates, 3–5 June 2024*; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2024; pp. 1–10.
89. Almiani, M.; Al-Rahayfeh, A.; Atiewi, S.; Alweshah, M.; Abu-Salih, B.; Alrawashdeh, T. Scalable Healthcare Automation for Arabic Communities via IoT-Based Edge NLP Framework. In *Proceedings of the 2025 12th International Conference on Social Networks Analysis, Management and Security (SNAMS), Vienna, Austria, 25 November 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 405–409.
90. Ho, T.C.; Kharrat, F.; Abid, A.; Karray, F.; Koubaa, A. REMONI: An Autonomous System Integrating Wearables and Multimodal Large Language Models for Enhanced Remote Health Monitoring. In *Proceedings of the 2024 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Eindhoven, The Netherlands, 26 June 2024*; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2024; pp. 1–6.
91. Nazir, A.; Evangelista, E.; Bukhari, S.; Sharma, R.S. *Wearable IoMT, Edge AI and LLMs for Monitoring Health and Wellness: A Harmonization Framework for Responsible AI Interventions*; Cambridge University Press: Cambridge, UK, 2025; Preprint on Cambridge Open Engage. [\[CrossRef\]](#)
92. Subramanian, A.; Yang, Z.; Azimi, I.; Rahmani, A.M. Graph-Augmented LLMs for Personalized Health Insights: A Case Study in Sleep Analysis. In *Proceedings of the 2024 IEEE 20th International Conference on Body Sensor Networks, BSN 2024-Proceedings, Chicago, IL, USA, 17 October 2024*; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2024; pp. 1–4.
93. Kim, Y.; Xu, X.; McDuff, D.; Breazeal, C.; Park, H.W. Health-LLM: Large Language Models for Health Prediction via Wearable Sensor Data. In *Proceedings of the Fifth Conference on Health, Inference, and Learning*; Proceedings of Machine Learning Research: Cambridge, MA, USA, 2024; Volume 248, pp. 522–539.
94. Peyroteo, M.; Ferreira, I.A.; Elvas, L.B.; Ferreira, J.C.; Lapão, L.V. Remote Monitoring Systems for Patients with Chronic Diseases in Primary Health Care: Systematic Review. *JMIR mHealth uHealth* **2021**, *9*, e28285. [\[CrossRef\]](#)
95. Lin, C.; Kuo, C.F. Roles and Potential of Large Language Models in Healthcare: A Comprehensive Review. *Biomed. J.* **2025**, *48*, 100868. [\[CrossRef\]](#)

96. Taimoor, N.; Rehman, S. Reliable and Resilient AI and IoT-Based Personalised Healthcare Services: A Survey. *IEEE Access* **2022**, *10*, 535–563. [[CrossRef](#)]
97. Leiser, F.; Guse, R.; Sunyaev, A. Large Language Model Architectures in Health Care: Scoping Review of Research Perspectives. *J. Med. Internet Res.* **2025**, *27*, e70315. [[CrossRef](#)]
98. Alshehri, F.; Muhammad, G. A Comprehensive Survey of the Internet of Things (IoT) and AI-Based Smart Healthcare. *IEEE Access* **2021**, *9*, 3660–3678. [[CrossRef](#)]
99. Albshaier, L.; Almarri, S.; Albuali, A. Federated Learning for Cloud and Edge Security: A Systematic Review of Challenges and AI Opportunities. *Electronics* **2025**, *14*, 1019. [[CrossRef](#)]
100. Hu, H.; Song, Z.; Shi, W. Large AI Model-Enhanced Digital Twin-Driven 6G Healthcare IoE. *Electronics* **2026**, *15*, 619. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.