Academic Journal of Management and Social Sciences
ISSN: 2958-4396 | Vol. 14, No. 3, 2026

Large Language Models in Healthcare Management: Implementation Evidence, Governance Architecture, and Strategic Pathways

Ran Zhang [1, 2], Saibing Wang [1]

[1] Henan University of Science and Technology, Luoyang, Henan, 471000, China
[2] The First Affiliated Hospital of Henan University of Science and Technology, Luoyang, Henan, 471000, China

Abstract: Large language models (LLMs) are rapidly shifting from exploratory tools to production infrastructure in healthcare management. This review focuses on five implementation domains: model foundations and adaptation, knowledge retrieval and question answering, document intelligence, operations and decision support, and governance and safety. Across these domains, the literature converges on one key result: organizational value is created less by raw model capability and more by workflow design, policy alignment, and institutional accountability. Reported benefits include faster information turnaround, lower administrative friction, improved triage support, and broader access to decision resources. Conversely, reported risks encompass hallucination, uneven reliability across contexts, privacy leakage, automation bias, and governance lag. To address these, we identify recurrent sociotechnical patterns that differentiate robust deployments from fragile pilots: retrieval grounding, role-bounded human oversight, auditable interaction logs, staged rollout, and continuous quality monitoring. We further propose an implementation maturity pathway linking technical controls to management outcomes, emphasizing total cost of ownership, workforce adaptation, and cross-functional governance capability. The review concludes that the next phase of healthcare LLM adoption should prioritize longitudinal multi-site evaluation, management-centered benchmar

Keywords: Large Language Models (LLMs); Artificial Intelligence Governance; Sociotechnical Systems; Implementation Science.

1. Introduction

Healthcare management currently operates under simultaneous pressures: rising service demand, persistent staffing constraints, regulatory complexity, and intensified expectations for timeliness and transparency in organizational decision-making. These pressures have expanded interest in LLMs from conversational novelty to operational tooling capable of supporting communication-heavy managerial work [1, 2]. Early discourse often framed LLMs as generalized productivity aids, but health-system leadership increasingly evaluates them through institution-level outcomes—such as cycle-time reduction, policy adherence, and risk containment—rather than isolated model scores [3]. The central issue is no longer whether an LLM can produce plausible language, but whether it can produce decision support that remains reliable under heterogeneous workflows, constrained data access, and legal accountability boundaries [4,5]. In management contexts, seemingly small failures in summarization fidelity, escalation routing, or exception handling can propagate into staffing, billing, quality, and patient communication workflows. Consequently, implementation quality depends on control design as much as algorithmic performance [6].

Recent healthcare market analyses show that organizations are moving toward productized LLM integrations promising immediate throughput gains, especially in documentation, messaging, and internal knowledge navigation [7]. However, institutional outcomes vary widely across sites, even for similar technologies, because success is heavily mediated by baseline governance maturity, process standardization, and data stewardship practices. This variance has strategic significance: executive teams treating model deployment purely as a technical purchase may systematically underperform those treating it as organizational redesign [8]. The current evidence base is therefore both promising and structurally incomplete. Reviews published in 2024-2025 report expanding use cases, from retrieval-augmented knowledge assistance to multi-agent coordination prototypes and specialized decision support [9,10]. Yet many studies remain short-horizon, single-site, and metric-fragmented, making it difficult to infer durable effects on quality, cost, safety, and staff burden. The field has generated abundant proof-of-concept literature, but fewer studies demonstrate strong external validity for management decisions at scale [6].

This review addresses that gap by synthesizing evidence through a management implementation lens, focusing on organizationally relevant mechanisms: how LLM outputs are constrained, validated, escalated, and audited inside real workflows. We analyze not only where performance improves, but also where risk concentrates and why those risks persist even when benchmark performance appears strong. Our scope includes hospital operations, health-system administration, quality and safety governance, and cross-functional coordination tasks that link clinical and non-clinical units [1, 7]. Five domain questions structure the review: (1) which foundation-model adaptation patterns are most credible; (2) under what conditions retrieval and question-answering pipelines produce trustworthy operational knowledge; (3) when document intelligence generates net productivity gains; (4) how robust current evidence is for LLM-supported operations; and (5) what governance architecture is required to maintain safety, legitimacy, and public trust during scaling [8, 10]. Across these domains, we argue that high-performing programs are sociotechnical by design. They combine technical adaptation with institutional capability—policy engineering, role

accountability, workforce training, and lifecycle monitoring—ensuring LLMs act as bounded components in managed service architectures that explicitly preserve human responsibility for consequential judgments [6, 9].

The review contributes by consolidating implementation evidence, offering a governance-oriented interpretation of LLM value durability, and proposing strategic pathways for healthcare leaders balancing innovation speed with operational assurance [2, 4]. From a management theory perspective, LLM deployment creates value when organizations can repeatedly sense information needs and reconfigure governance routines faster than external demand changes, explaining why capability heterogeneity often outweighs model heterogeneity [10, 11]. Furthermore, management decisions require net performance accounting—evaluating model-assisted speed against correction burden, escalation delay, monitoring labor, and compliance overhead—revealing that governance design determines whether productivity remains positive after scale-up [12, 13]. Finally, because healthcare organizations operate under public-trust obligations exceeding ordinary enterprise software adoption criteria, ethical and public-health scholarship suggests that legitimacy must be engineered through transparency, communication quality, and clear responsibility boundaries from the very first implementation stage [14, 15].

2. LLM Foundations and Adaptation

The first strategic task in healthcare management deployment is selecting and adapting model foundations to organizational constraints. Educational overviews consistently emphasize that model capability varies by task type, domain specificity, context window management, and prompting regime [16, 17]. This implies that generic model selection based on broad benchmark rankings is insufficient; institutions need explicit task-model matching that accounts for the informational and accountability structure of each workflow. Work on medical grounding and domain adaptation indicates that management-facing tasks require tighter coupling between generation and verifiable knowledge artifacts than open-domain use cases typically provide [18]. Targeted adaptation can materially improve relevance and consistency for bounded tasks, but gains can degrade when prompts or context windows drift beyond training assumptions [19]. Consequently, prompting itself has become an organizational competence, as institutions with codified prompt-policy patterns achieve better reproducibility and lower variance across users [20, 21].

Foundation decisions are increasingly linked to infrastructure architecture. Secure retrieval and privacy-preserving extraction studies show that adaptation quality depends on where data are stored, how context is injected, and which controls govern access to sensitive material [22, 23]. Dual-model and multi-component integration approaches illustrate another design trend: separating reasoning and validation roles rather than forcing a single model to perform all functions [24, 25]. Furthermore, adaptation quality must be evaluated at the service level—tracking exception rates, correction latency, policy-violation frequency, and burden transfer to human supervisors [12]. Robust adaptation strategy therefore relies on domain grounding, policy-bounded prompting, architecture-level data controls, and service-level performance monitoring; programs operationalizing all four are more likely to convert technical capability into measurable managerial value [9]. Procurement governance must also weigh contractual constraints regarding data portability and update transparency, as these influence strategic flexibility as much as immediate technical performance [26]. Finally, architectural planning should prioritize interface stability to overcome integration friction with legacy systems [27, 28], include local linguistic validation [8, 19], and sequence adaptation along an explicit maturity roadmap into higher-consequence operations [29, 30].

3. Q&A and Knowledge Retrieval

Knowledge retrieval is where healthcare management teams often see the earliest practical gains from LLM adoption, as retrieval-augmented pipelines can reduce search friction for policies and procedures when staff need rapid answers under time pressure [22, 31]. Recent healthcare retrieval frameworks show that combining lexical retrieval, semantic matching, and reasoning-aware augmentation improves answer completeness compared with naive prompting [18, 32]. However, retrieval design is fundamentally a data-governance decision [22, 23]. Hallucination analyses document that retrieval presence alone does not eliminate factual error; quality depends on source authority, chunking strategy, and post-retrieval verification logic [33]. Operational Q&A studies illustrate mixed outcomes: while systems can support routine information access, reliability drops in ambiguous scenarios requiring policy interpretation [34-36]. Thus, institutions with mature knowledge governance realize disproportionate gains because their retrieval systems have dependable source corpora to work with [7, 8].

Moreover, management Q&A should prioritize explanation and provenance as first-class outputs, reducing automation bias and improving audit readiness [33]. Strategically, retrieval should be implemented as a monitored service with explicit governance hooks, such as source quality checks and mandatory escalation for low-confidence responses [31, 32]. Effective systems classify requests into operational categories to determine appropriate retrieval depth, confidence thresholds, and review requirements [35, 37]. Interaction design must also foreground uncertainty cues and mandatory acknowledgement of high-risk outputs to mitigate behavioral risks like over-trust [38, 39]. In multilingual and cross-jurisdictional contexts, localized corpora and domain-specific glossaries are necessary to preserve fidelity [40, 41]. Ultimately, organizations must connect retrieval telemetry to management KPIs to ensure that the infrastructure costs of retrieval investments are justified by systemic reductions in downstream rework and decision latency [42, 43].

4. Document Intelligence

Document intelligence is one of the most immediate LLM use domains in healthcare management because drafting documents consumes substantial professional time, and LLM assistance can significantly reduce first-draft latency [28, 37]. Evidence suggests that LLM-supported drafting can lower manual effort for routine documentation and data cleaning tasks [44, 45]. However, field-experiment evidence indicates that productivity effects are heterogeneous across user roles and task complexity. In operational coding and financial management, LLMs have the potential to assist throughput only when outputs remain tightly constrained by coding rules

and specialist oversight [27, 46]. Furthermore, quality and readability assessments in public-facing communication reinforce the need for explicit post-editing protocols to prevent subtle inaccuracies from undermining trust [47]. Integrated systems, rather than standalone tools, are essential because they support traceability and align with existing records and governance workflows. Institutions that monitor cycle-time reduction, rework rates, and reviewer burden can iteratively redesign templates to maintain quality at scale.

The most credible path forward is assistive generation with explicit accountability, rather than the full automation of consequential documentation [37, 46]. To achieve this, organizations must map document workflows prior to rollout to target automation where marginal benefit is highest and governance burden is manageable [45]. Establishing institution-wide language standards and template governance can substantially improve both automation quality and review efficiency [28, 44]. Furthermore, document intelligence programs should utilize multi-dimensional quality dashboards, as readability, accuracy, and decision relevance can diverge [48]. Capturing structured feedback on edits and overrides generates strategic data assets that improve institutional memory, thereby enhancing consistency and reducing avoidable rework over time [46, 49].

5. Decision Support and Operations

Decision support and operations represent the highest-value but highest-risk frontier, with applications emerging in triage workflows, resource coordination, and agent-based process orchestration [50, 51]. While decomposing workflows into specialized agent roles can improve coordination quality [52], empirical analyses demonstrate that technical performance improvements do not automatically translate into organizational trust or adoption [19, 42]. Operations-oriented studies indicate that LLMs can reduce latency in information synthesis and routine decision preparation [46, 49]. However, frameworks emphasizing protocol-driven coordination and state management are necessary to provide stronger operational assurance and auditability [24, 30]. For administrative operations, use cases such as procurement planning and staffing coordination carry lower immediate clinical risk and can serve as intermediate deployment stages [27].

A significant risk in this domain is "authority drift," which can occur even when teams formally maintain human-in-the-loop policies [38]. Mitigating this requires systems designed for dependable collaboration that preserve explicit human accountability for final decisions [43, 50]. Operational rollout is most effective when paired with simulation-based stress testing to evaluate edge cases and tune escalation thresholds and fallback designs [51]. Clear decision-rights matrices should specify which decisions can be model-assisted versus those excluded entirely [42, 50]. Furthermore, value durability depends on embedding economic evaluations into operations governance, ensuring that overhead remains proportional to risk and use-case complexity. Finally, establishing formal cross-functional governance councils with defined escalation pathways can improve responsiveness and reduce policy drift during expansion.

6. Governance and Safety

Governance and safety ultimately determine whether deployments scale to institutional legitimacy, as generative systems can unintentionally amplify misinformation when controls for source quality and audience adaptation are weak 40]. The literature converges on a core principle: safety is a property of a managed lifecycle encompassing design, deployment, monitoring, and incident response [17, 33]. Because high fluency can mask low epistemic reliability in low-frequency scenarios [4, 6], architecture must mandate confidence signaling, evidence citation, and explicit fallback behaviors. Furthermore, equity, transparency, and accountability must be operationalized as concrete requirements rather than abstract principles [39]. With capability velocity often outpacing policy adaptation [53, 54], organizations need change-management controls that connect model updates to validation checkpoints. Frontline acceptance heavily depends on perceived fairness, workload impact, and clarity of accountability when errors occur [10, 26].

Governance must also address tone and contextual fit in patient-facing communication, which remain recurrent failure points [35, 48]. A robust architecture includes policy-bounded interaction design, continuous monitoring with actionable thresholds, and transparent documentation for internal and external accountability [55]. Maturity is operationalized through assurance cases that explicitly link claims, evidence, and controls, updated after major model changes or significant incidents [17]. Dedicated incident management infrastructure featuring standardized taxonomies enables trend analysis and targeted corrective action across departments [38]. Transparent records of model scope and human oversight responsibilities reduce legal ambiguity and strengthen public trust during adverse events [26, 39]. Ultimately, programs that treat every near miss as a learning input thrive in rapidly evolving ecosystems where fixed policies can become obsolete quickly .

7. Discussion

The cross-domain evidence supports a clear synthesis: healthcare management value from LLMs is real but conditional, proving strongest in information-dense workflows bounded by robust policy and review architectures [1, 3]. Three strategic patterns consistently emerge: retrieval grounding yields better operational utility for policy and guideline queries [31, 32]; role-bounded human oversight is essential as a normal operational feature rather than an emergency fallback [42, 47]; and lifecycle monitoring is a stronger predictor of durable performance than one-time validation [6, 53]. However, methodological limitations persist, as few evaluations fully account for the total cost of ownership, including governance labor and retraining costs [12, 43]. Additionally, many successes emerge from sites with high digital maturity or unusually cooperative implementation contexts [7, 10].

For management practice, staged adoption remains the most defensible strategy. Expansion into higher-stakes decision support should occur only after demonstrating stable control performance in earlier stages [27, 46]. Treating governance capability as a strategic asset allows organizations to create reusable control infrastructure for subsequent AI initiatives [20, 30]. The field urgently needs integrated, management-centered benchmark frameworks that capture safety-adjusted productivity, rework, and equity-sensitive performance across realistic workflow scenarios [9, 18]. Future research should utilize multi-site comparative designs to evaluate governance interventions like mandatory

review policies and escalation routing design [42, 47], while combining operational data with workforce outcomes to assess full institutional impact [10, 41]. Finally, integrating implementation science with health policy analysis will improve decision quality for executives operating in complex, multi-jurisdictional systems .

In the near term, LLM systems will likely evolve into policy-aware service components embedded in enterprise workflow platforms, improving traceability provided that model orchestration aligns with explicit responsibility models [29, 52]. A convergence between retrieval engineering and governance engineering is imminent, making dynamic source validation and provenance-rich formats standard features rather than optional add-ons [22, 23]. Workforce strategy will become a decisive factor, demanding training that emphasizes decision hygiene and organizational learning loops over mere tool familiarization [16, 41]. Over the medium term, progress toward interoperable governance standards—such as shared audit schemas—will accelerate cross-site learning and policy maturation across systems [55]. Consequently, evaluation pipelines will become continuously refreshed to track drift and trigger targeted revalidation, maintaining connection to current operational reality rather than historical test sets [6, 18].

8. Conclusion

LLMs create durable value only when technical capability translates into governed organizational practice, heavily relying on retrieval grounding, workflow integration, explicit human accountability, and continuous monitoring [8, 50]. The strategic imperative is to evaluate safety-adjusted productivity, transferability, and total cost of ownership before scaling high-consequence use cases [5, 13]. Disciplined prioritization of use cases with explicit decision pathways and tractable data provenance improves the return on governance investment and reduces implementation volatility during scale-up [3, 43]. For hospital and health-system leaders, the credible path forward is staged adoption characterized by "governance-by-design"—building capability in lower-risk workflows and expanding scope only when assurance performance remains stable under operational stress [34, 49]. Under this model, LLMs can accelerate reliable management decisions while preserving professional responsibility and public legitimacy.

References

[1] Yu P, Xu H, Hu X, Deng C. Leveraging Generative AI and Large Language Models:A Comprehensive Roadmap for Healthcare Integration. Healthcare. 2023, Vol. 11 (No. 20), p. 2776.

[2] AmorimLopes M, Martins H, Correia T. Artificial intelligence and the future in health policy, planning and management. The International Journal of Health Planning and Management. 2023, Vol. 39 (No. 1), p. 3-8.

[3] Carroll W M. Generative AI in clinical practice and operations. Nursing Management. 2023, Vol. 54 (No. 10), p. 56.

[4] Gao Y, BaptistaHon D T, Zhang K. The inevitable transformation of medicine and research by large language models: The possibilities and pitfalls. MedComm Future Medicine. 2023, Vol. 2 (No. 2).

[5] Xue V W, Lei P, Cho W C. The potential impact of ChatGPT in clinical and translational medicine. Clinical and Translational Medicine. 2023, Vol. 13 (No. 3), p. e1216.

[6] Aditya G. Understanding and Addressing AI Hallucinations in Healthcare and Life Sciences. International Journal of Health Sciences. 2024, Vol. 7 (No. 3), p. 1-11.

[7] Sharifi S, Namvar M, Intezari A, Akhlaghpour S. Healthcare LLMs Go to Market:A Realist Review of Product Launch News. Studies in health technology and informatics. 2024, Vol. 316, p. 712-716.

[8] Sun X, Tang W, Huang Z, Long E, Wan P. From general to specific: Tailoring large language models for realworld medical communications. Clinical and Translational Medicine. 2024, Vol. 15 (No. 1), p. e70157.

[9] Fadul N, Alaskar M F, Jillahi K B, El-Khaled D B. Generative AI in Healthcare:An Analytical Review of Models, Clinical Applications, and Decision-Support Implications. Journal of Future Artificial Intelligence and Technologies. 2025, Vol. 2 (No. 4), p. 587-615.

[10] Niemelä M, Kemppainen T, Vuollet J. Artificial intelligence in wellbeing services counties: management and experts perspectives on opportunities, challenges and AI readiness. Finnish Journal of eHealth and eWelfare. 2025, Vol. 17 (No. 1).

[11] Song Y, Yan T, Jia F, Chen L, Li H. Developing Generative AI for Value CoCreation:An InterventionBased Randomized Field Experiment in a Healthcare Context. Journal of Operations Management. 2025.

[12] Ong J C L, Michael C, Ng N, et al. Generative AI and Large Language Models in Reducing Medication Related Harm and Adverse Drug Events A Scoping Review. 2024.

[13] Fahy S, Niemann M, Böhm P, et al. Assessment of the Quality and Readability of Information Provided by ChatGPT in Relation to the Use of Platelet-Rich Plasma Therapy for Osteoarthritis. Journal of Personalized Medicine. 2024, Vol. 14 (No. 5), p. 495.

[14] Joseph J, Jose B, Jose J. The generative illusion: how ChatGPT-like AI tools could reinforce misinformation and mistrust in public health communication. Frontiers in Public Health. 2025, Vol. 13, p. 1683498.

[15] Jeyaraman M, Balaji S, Jeyaraman N, Yadav S. Unraveling the Ethical Enigma:Artificial Intelligence in Healthcare. Cureus. 2023, Vol. 15 (No. 8), p. e43262.

[16] Idan D, Einav S. Primer on large language models: an educational overview for intensivists. Critical Care. 2025, Vol. 29 (No. 1), p. 238.

[17] Teles A S, AbdAlrazaq A, Heston T F, et al. Editorial: Large Language Models for medical applications. Frontiers in Medicine. 2025, Vol. 12, p. 1625293.

[18] Gao Y, Li R, Croxford E, et al. Large Language Models and Medical Knowledge Grounding for Diagnosis Prediction. 2023.

[19] Miao J, Thongprayoon C, Fülöp T, Cheungpasitporn W. Enhancing clinical decision-making: Optimizing ChatGPT's performance in hypertension care. Journal of Clinical Hypertension. 2024, Vol. 26 (No. 5), p. 588-593.

[20] Korzyski P, Mazurek G, Krzypkowska P, Kurasiski A. Artificial intelligence prompt engineering as a new digital competence: Analysis of generative AI technologies such as ChatGPT. Entrepreneurial Business and Economics Review. 2023, Vol. 11 (No. 3), p. 25-37.

[21] Sai S, Gaur A, Sai R V, et al. Generative AI for Transformative Healthcare: A Comprehensive Study of Emerging Models, Applications, Case Studies, and Limitations. IEEE Access. 2024, Vol. 12, p. 31078-31106.

[22] Jiang E, Chen A P, Tenison I, Kagal L. MediRAG: Secure Question Answering for Healthcare Data. 2024, p. 6476-6485.

[23] Wiest I C, Ferber D, Zhu J, et al. From Text to Tables: A Local Privacy Preserving Large Language Model for Structured Information Retrieval from Medical Documents. 2023.

[24] Kurma S, Singh K, Paul A, et al. Dual-LLM Integration With Reconfigurable Intelligent Surface for Healthcare Networks. IEEE Transactions on Communications. 2025, Vol. 73 (No. 11), p. 11323-11339.

[25] Borkowski A, Ben-Ari A. Muli-Agent AI Systems in Healthcare: Technical and Clinical Analysis. Preprints.org. 2024.

[26] Low D A, Halvorsen P H, Hedrick S. Will large language model AI (ChatGPT) be a benefit or a risk to quality for submission of medical physics manuscripts?. Medical Physics. 2025, Vol. 52 (No. 4), p. 1974-1977.

[27] Aguero D, Nelson S D. The Potential Application of Large Language Models in Pharmaceutical Supply Chain Management. The Journal of Pediatric Pharmacology and Therapeutics. 2024, Vol. 29 (No. 2), p. 200-205.

[28] Li X, Guo H, Li D, Zheng Y. Engine of Innovation in Hospital Pharmacy: Applications and Reflections of ChatGPT. Journal of Medical Internet Research. 2024, Vol. 26, p. e51635.

[29] Yao Z, Yu H. A Survey on LLM-based Multi-Agent AI Hospital. 2025.

[30] ElSayed Z, Erickson C A, Pedapati E V. MCP-AI: Protocol-Driven Intelligence Framework for Autonomous Reasoning in Healthcare. ArXiv.org. 2025.

[31] Miao J, Thongprayoon C, Suppadungsuk S, et al. Integrating Retrieval-Augmented Generation with Large Language Models in Nephrology: Advancing Practical Applications. Medicina. 2024, Vol. 60 (No. 3), p. 445.

[32] Zhao X, Liu S, Yang S-Y, Miao C. MedRAG: Enhancing Retrieval-augmented Generation with Knowledge Graph-Elicited Reasoning for Healthcare Copilot. 2025, p. 4442-4457.

[33] Ilicki J. A Framework for Critically Assessing ChatGPT and Other Large Language Artificial Intelligence Model Applications in Health Care. Mayo Clinic Proceedings Digital Health. 2023, Vol. 1 (No. 2), p. 185-188.

[34] Berry P, Dhanakshirur R R, Khanna S. Utilizing large language models for gastroenterology research: a conceptual framework. Therapeutic Advances in Gastroenterology. 2025, Vol. 18, p. 17562848251328577.

[35] Chow J C L, Li K. Large Language Models in Medical Chatbots: Opportunities, Challenges, and the Need to Address AI Risks. Information. 2025, Vol. 16 (No. 7), p. 549.

[36] Lee T, Staller K, Botoman V, et al. ChatGPT Answers Common Patient Questions About Colonoscopy. Gastroenterology. 2023, Vol. 165 (No. 2), p. 509-511.e7.

[37] Patel S B, Lam K. ChatGPT: the future of discharge summaries?. The Lancet Digital Health. 2023, Vol. 5 (No. 3), p. e107-e108.

[38] BélislePipon J. Why we need to be careful with LLMs in medicine. Frontiers in Medicine. 2024, Vol. 11, p. 1495582.

[39] Islam M R, Urmi T J, Mosharrafa R A, et al. Role of ChatGPT in health science and research: A correspondence addressing potential application. Health Science Reports. 2023, Vol. 6 (No. 10), p. e1625.

[40] Lam K. ChatGPT for low- and middle-income countries: a Greek gift?. The Lancet Regional Health - Western Pacific. 2023, Vol. 41, p. 100906.

[41] Ghadban Y A, Lu H, Adavi U, et al. Transforming Healthcare Education: Harnessing Large Language Models for Frontline Health Worker Capacity Building using Retrieval-Augmented Generation. 2023.

[42] Delourme S, Redjdal A, Bouaud J, Séroussi B. Measured Performance and Healthcare Professional Perception of Large Language Models Used as Clinical Decision Support Systems: A Scoping Review. Studies in health technology and informatics. 2024, Vol. 316, p. 841-845.

[43] Sangal R B, Sharifi M, Rhodes D J, Melnick E R. Clinical Decision Support: Moving Beyond Interruptive Pop-up Alerts. Mayo Clinic Proceedings. 2023, Vol. 98 (No. 9), p. 1275-1279.

[44] Li X, Guo H, Li D, Zheng Y. Engine of Innovation in Hospital Pharmacy: Applications and Reflections of ChatGPT. 2023.

[45] Amgad M, Shamoon B, Mekhemar H, et al. Digitization of Medical Data: LLM-Based Data Entry, Cleansing, and Visualization Solutions. 2025, p. 339-342.

[46] Papageorgiou P S, Christodoulou R C, Pitsillos R, et al. The Role of Large Language Models in Improving Diagnostic-Related Groups Assignment and Clinical Decision Support in Healthcare Systems: An Example from Radiology and Nuclear Medicine. Applied Sciences. 2025, Vol. 15 (No. 16), p. 9005.

[47] Soddu M, Vito A D, Madeddu G, et al. Assessing the Accuracy, Completeness and Safety of ChatGPT-4o Responses on Pressure Injuries in Infants: Clinical Applications and Future Implications. Nursing Reports. 2025, Vol. 15 (No. 4), p. 130.

[48] Vicnesh J, Salvi M, Hagiwara Y, et al. Application of Infrared Thermography and Artificial Intelligence in Healthcare: A Systematic Review of Over a Decade (2013-2024). IEEE Access. 2025, Vol. 13, p. 5949-5973.

[49] Anthony P. Artificial Intelligence and Patient Support Models for Enhanced Chronic Disease Management and Adherence. International Journal of Future Engineering Innovations. 2025, Vol. 2 (No. 5), p. 65-82.

[50] Ogdu C U, Gurbuz S, Karaköse M, Hanolu E. Medical Implications of LLM Based Clinical Decision Support Systems in Healthcare. 2025, p. 1-4.

[51] Han S, Choi W. Development of a Large Language Model-based Multi-Agent Clinical Decision Support System for Korean Triage and Acuity Scale (KTAS)-Based Triage and Treatment Planning in Emergency Departments. Advances in Artificial Intelligence and Machine Learning. 2025, Vol. 05 (No. 01), p. 3261-3275.

[52] Joy M. Agentic Workflows in Healthcare: Advancing Clinical Efficiency through AI Integration. International Journal of Scientific Research in Computer Science Engineering and Information Technology. 2025, Vol. 11 (No. 2), p. 567-575.

[53] Temsah A, Alhasan K, Altamimi I, et al. DeepSeek in Healthcare: Revealing Opportunities and Steering Challenges of a New Open-Source Artificial Intelligence Frontier. Cureus. 2025, Vol. 17 (No. 2), p. e79221.

[54] Yuan H. Agentic Large Language Models for Healthcare: Current Progress and Future Opportunities. Medicine Advances. 2025, Vol. 3 (No. 1), p. 37-41.

[55] Agliata A, Pilato A D, Mariacarmen S, et al. Generative AI and Emotional Health: Innovations with Haystack. 2024, p. 1-4.