Large Language Model–Assisted Clinical Decision Support for Regional Anesthesia Selection in Knee Arthroscopy Patients

Sebastian Hensley
Department of Biomedical Informatics, University of North Texas Health Science Center, Fort Worth, Texas, USA.
s.hensley@unthsc.edu

Evan Wainwright
School of Computing and Information Sciences, Florida International University, Miami, Florida, USA.
e.wainwrightr@fiu.edu

Daniel Brooks
Department of Health Systems Engineering, University of Missouri–Kansas City, Kansas City, Missouri, USA.
daniel.brooks@umkc.edu

Abstract

Regional anesthesia has become a fundamental component of perioperative pain management for knee arthroscopy procedures due to its ability to reduce opioid consumption, improve postoperative recovery, and enhance patient satisfaction. Nevertheless, selecting an appropriate regional anesthesia strategy remains a complex clinical decision-making process that requires consideration of patient-specific characteristics, surgical factors, institutional resources, practitioner expertise, and evolving evidence-based guidelines. Recent advances in large language models (LLMs) have created opportunities to augment clinical decision support systems through sophisticated reasoning capabilities, natural language interaction, and integration of heterogeneous medical information sources. This study explores a system-level framework for LLM-assisted clinical decision support in regional anesthesia selection for knee arthroscopy patients. Rather than focusing solely on predictive performance, the paper examines the architectural, organizational, and governance dimensions associated with integrating LLM technologies into perioperative care environments.

The proposed framework combines structured electronic health record data, clinical practice guidelines, historical anesthesia outcomes, and real-time clinical documentation within a multi-layer decision support architecture. Particular attention is given to explainability, reliability, fairness, workflow integration, and human oversight mechanisms necessary for safe deployment. The study analyzes how LLM-assisted systems can support individualized recommendations among femoral nerve blocks, sciatic nerve blocks, adductor canal blocks, combined approaches, and alternative analgesic pathways while maintaining clinician accountability. Furthermore, the paper evaluates implementation challenges including data quality, institutional interoperability, regulatory compliance, model drift, and infrastructure sustainability. Findings suggest that LLM-assisted decision support can substantially enhance consistency and knowledge accessibility in perioperative decision-making when embedded within robust governance structures. The research contributes a comprehensive socio-

technical perspective on the future role of generative artificial intelligence in regional anesthesia management and clinical decision support ecosystems.

Keywords

Large Language Models; Clinical Decision Support Systems; Regional Anesthesia; Knee Arthroscopy; Artificial Intelligence in Healthcare; Explainable AI; Perioperative Informatics; Health Systems Engineering.

1. Introduction

The rapid advancement of artificial intelligence technologies has fundamentally transformed contemporary healthcare information systems. Among emerging innovations, large language models have attracted significant attention due to their capacity to process complex clinical narratives, synthesize diverse information sources, and generate contextually meaningful recommendations. Unlike traditional rule-based decision support systems that rely on explicitly encoded knowledge structures, large language models provide flexible reasoning capabilities capable of incorporating both structured and unstructured clinical information. This capability is particularly relevant in perioperative medicine, where decision-making often depends on nuanced contextual interpretation rather than rigid protocol adherence.

Knee arthroscopy remains one of the most frequently performed orthopedic procedures worldwide. Although generally considered minimally invasive, postoperative pain management continues to represent a significant clinical challenge affecting patient outcomes, rehabilitation timelines, healthcare resource utilization, and overall patient satisfaction. Regional anesthesia techniques have become central components of multimodal analgesia strategies, offering targeted pain control while minimizing systemic opioid exposure. However, selecting the most appropriate regional anesthesia approach requires balancing multiple competing considerations, including anticipated pain intensity, patient comorbidities, neurological risk profiles, mobility requirements, surgical complexity, and institutional capabilities.

Historically, anesthesia selection decisions have depended heavily on individual practitioner experience, local practice patterns, and guideline interpretation. Such approaches often generate variability across institutions and providers. The increasing availability of electronic health records, perioperative databases, and clinical analytics platforms has encouraged the development of data-driven decision support systems intended to improve consistency and quality of care. Nevertheless, conventional systems frequently struggle to integrate narrative documentation, evolving evidence, and contextual clinical reasoning.

Large language models offer a potentially transformative solution by serving as intermediary reasoning engines capable of synthesizing diverse forms of medical knowledge. Rather than replacing clinicians, these systems may function as collaborative decision-support partners that enhance information retrieval, contextual analysis, and recommendation transparency. The objective of this paper is to examine the design, implementation, and governance of LLM-assisted clinical decision support systems for regional anesthesia selection in knee arthroscopy patients. The discussion emphasizes system architecture, operational deployment, ethical considerations, and long-term sustainability rather than solely focusing on algorithmic performance.

2. Background and Related Literature

Clinical decision support systems have evolved substantially during the past three decades. Early systems primarily utilized deterministic logic, predefined clinical pathways, and manually encoded expert knowledge. While these approaches demonstrated value in reducing medication errors and improving guideline adherence, their limitations became apparent in complex clinical environments characterized by uncertainty, incomplete information, and rapidly changing evidence bases [1].

Machine learning subsequently expanded the capabilities of decision support technologies through predictive analytics and risk stratification. Applications in perioperative medicine have included postoperative complication prediction, pain forecasting, patient monitoring, and resource allocation optimization [2,3]. However, many machine learning systems operate as narrow predictive tools that provide limited insight into underlying reasoning processes. Such limitations can reduce clinician trust and hinder adoption within high-stakes healthcare settings [4].

The emergence of transformer-based large language models has introduced new possibilities for healthcare applications. Recent investigations have demonstrated that advanced language models can perform medical question answering, clinical documentation assistance, diagnostic reasoning support, and evidence synthesis tasks with increasingly sophisticated performance characteristics [5–7]. These developments have generated significant interest in integrating generative artificial intelligence into healthcare delivery systems.

Within anesthesiology, artificial intelligence research has traditionally focused on physiological monitoring, perioperative risk prediction, and operating room optimization [8]. More recent studies have explored AI-supported pain management strategies, anesthesia planning systems, and personalized perioperative care pathways [9]. Nevertheless, relatively limited research has examined the specific role of large language models in regional anesthesia selection despite the complexity of this decision-making domain.

Regional anesthesia decisions require integration of numerous variables including age, body mass index, previous surgical history, chronic pain conditions, anticoagulation status, neurological disorders, functional recovery goals, and anticipated postoperative rehabilitation requirements. Existing evidence indicates that different nerve block techniques produce varying trade-offs regarding analgesic effectiveness, motor preservation, mobility outcomes, and complication risks [10]. Combined femoral-sciatic nerve block approaches have demonstrated effectiveness in knee arthroscopy pain management, particularly in cases involving more extensive procedures [11].

The growing complexity of perioperative care environments suggests a need for decision support systems capable of synthesizing guideline recommendations, patient-specific characteristics, institutional constraints, and emerging clinical evidence. Large language models may provide an important mechanism for achieving this integration while preserving clinician oversight and contextual flexibility.

3. System Architecture for LLM-Assisted Anesthesia Selection

The proposed LLM-assisted clinical decision support framework adopts a layered socio-technical architecture designed to balance analytical sophistication with operational reliability. At the foundational layer, the system aggregates information from multiple healthcare data sources, including electronic health records, anesthesia information management systems, perioperative registries, imaging reports, laboratory results, and institutional clinical guidelines.

A critical architectural principle involves maintaining separation between data acquisition, reasoning processes, and recommendation presentation. Such separation enhances transparency, facilitates auditing, and reduces the likelihood of uncontrolled model behavior. Structured clinical variables are processed through conventional analytics pipelines, while unstructured clinical narratives are interpreted using natural language processing components integrated with the large language model environment.

The reasoning layer represents the core innovation of the framework. Rather than generating recommendations from isolated prompts, the system employs retrieval-augmented generation mechanisms that ground model outputs within institutionally approved knowledge repositories. This approach reduces hallucination risk and improves alignment with evidence-based practice standards. Clinical guidelines, peer-reviewed literature, local protocols, and historical outcome databases collectively inform recommendation generation.

Within the anesthesia selection workflow, the model evaluates numerous contextual factors. For example, a younger patient undergoing a relatively minor arthroscopic intervention may benefit from an adductor canal block strategy prioritizing early mobility and functional recovery. Conversely, a patient undergoing extensive ligament reconstruction with anticipated substantial postoperative pain may warrant consideration of combined regional anesthesia approaches. The model's role is not to dictate decisions but rather to present evidence-informed options accompanied by explanatory reasoning.

The presentation layer emphasizes explainability and clinician interaction. Recommendations are delivered through structured summaries describing key influencing factors, relevant guideline considerations, potential benefits, and recognized limitations. By making reasoning pathways visible, the system seeks to enhance clinician trust while supporting informed shared decision-making.

Importantly, the architecture incorporates continuous feedback mechanisms. Clinicians can evaluate recommendation quality, document deviations from suggested pathways, and provide contextual explanations. These feedback loops contribute to system refinement and facilitate ongoing performance monitoring. The resulting architecture functions not merely as an artificial intelligence application but as an evolving organizational learning platform.

4. Explainability, Trust, and Human-AI Collaboration

Explainability represents one of the most significant challenges associated with deploying large language models in healthcare environments. While generative models may produce highly persuasive recommendations, healthcare professionals require transparent justification before integrating such outputs into patient care decisions. Trust in clinical decision support systems emerges not solely from predictive accuracy but from the ability to understand how recommendations are generated and how uncertainties are managed.

In the context of regional anesthesia selection, explainability assumes particular importance because clinical decisions frequently involve nuanced trade-offs. For instance, a recommendation favoring motor-sparing anesthesia techniques may prioritize rehabilitation objectives while potentially sacrificing certain analgesic advantages. Similarly, recommendations involving combined nerve block approaches may improve pain control while increasing procedural complexity and resource requirements.

The proposed framework addresses explainability through multi-layer reasoning disclosure. Rather than presenting single recommendations, the system communicates influencing patient characteristics, relevant evidence sources, confidence assessments, and alternative pathways. Such transparency allows clinicians to critically evaluate system outputs rather than accepting them uncritically.

Human-AI collaboration further requires clearly defined accountability structures. Clinical responsibility must remain with licensed healthcare professionals regardless of the sophistication of artificial intelligence technologies. Consequently, the system is designed to augment rather than replace clinician judgment. Recommendations function as advisory inputs that support decision-making while preserving professional autonomy and ethical responsibility.

Trust is also influenced by consistency. Clinicians are more likely to adopt decision support technologies when outputs demonstrate predictable behavior across comparable clinical scenarios. Therefore, governance frameworks should include regular auditing procedures, calibration assessments, and multidisciplinary oversight committees responsible for monitoring system performance over time.

5. Fairness, Bias, and Healthcare Equity

Healthcare artificial intelligence systems increasingly face scrutiny regarding fairness and equity implications. Large language models inherit characteristics from training data, institutional practices, and broader healthcare structures that may contain historical biases. Without appropriate safeguards, decision support systems risk perpetuating disparities across demographic, socioeconomic, and geographic populations.

Regional anesthesia provides a particularly relevant context for examining equity concerns because access to advanced anesthesia techniques varies considerably across healthcare settings.Regional anesthesia provides a particularly relevant context for examining equity concerns because access to advanced anesthesia techniques varies considerably across healthcare settings. Academic medical centers often possess specialized expertise, advanced ultrasound-guided capabilities, and dedicated regional anesthesia services, whereas smaller community hospitals may operate under significant resource constraints. If decision support systems are trained primarily on data originating from highly specialized institutions, recommendations may fail to account for operational realities in less resourced environments. Consequently, system design must incorporate contextual awareness regarding institutional capabilities and available expertise.

Bias may also emerge from historical treatment patterns embedded within electronic health record datasets. Certain demographic groups have historically experienced disparities in pain assessment, analgesic management, and access to specialized procedures [12]. Large language models trained on such datasets may inadvertently reproduce these inequities unless deliberate mitigation strategies are implemented. Fairness assessment therefore requires ongoing evaluation across age groups, sexes, racial and ethnic populations, insurance categories, and geographic regions.

A robust governance framework should incorporate fairness auditing as a continuous operational function rather than a one-time validation exercise. Outcome monitoring should examine whether recommendations disproportionately favor particular interventions for specific demographic populations without clinically justified rationale. Multidisciplinary review committees involving anesthesiologists, informaticians, ethicists, and health equity specialists can provide independent oversight regarding potential bias emergence.

Transparency also contributes to equity objectives. Clinicians should understand the factors influencing recommendations and retain the ability to challenge system outputs when contextual considerations warrant alternative approaches. Such mechanisms reduce the risk of automation bias while preserving individualized patient care. Ultimately, equitable deployment requires recognition that technical performance alone does not guarantee socially responsible healthcare innovation.

6. Deployment Infrastructure and Organizational Integration

Successful implementation of LLM-assisted decision support systems depends not only on model performance but also on organizational readiness and infrastructure maturity. Healthcare environments represent highly regulated socio-technical systems in which technology adoption is influenced by workflow integration, stakeholder acceptance, operational reliability, and institutional governance structures.

One of the primary deployment challenges involves interoperability. Modern healthcare organizations utilize heterogeneous information systems that frequently originate from different vendors and support varying data standards. Effective LLM-assisted decision support requires seamless integration across electronic health records, anesthesia management platforms, clinical documentation systems, scheduling applications, and quality improvement databases. Fragmented information ecosystems can substantially diminish recommendation quality by limiting access to relevant clinical context.

Infrastructure scalability represents another critical consideration. Large language models require substantial computational resources, particularly when supporting real-time clinical workflows. Organizations must evaluate whether inference operations should occur through cloud-based services, on-premises infrastructure, or hybrid deployment architectures. Each approach presents distinct trade-offs involving latency, cybersecurity, regulatory compliance, operational costs, and system resilience.

Cybersecurity concerns are particularly significant given the sensitivity of healthcare information. Clinical decision support systems process extensive quantities of protected health information and therefore constitute attractive targets for malicious actors. Robust security architectures should include encryption mechanisms, access controls, audit logging, anomaly detection capabilities, and incident response procedures. Furthermore, organizations must ensure compliance with evolving healthcare privacy regulations and information security standards.

Clinician engagement plays an equally important role in deployment success. Historical experience demonstrates that technically sophisticated decision support systems frequently fail when user-centered design principles are neglected. Excessive alerting, workflow disruption, and opaque recommendations can generate resistance among healthcare professionals. Consequently, implementation strategies should prioritize usability testing, clinician participation in system design, and iterative refinement based on real-world operational feedback.

Training and organizational learning also influence long-term sustainability. Clinicians must understand both the capabilities and limitations of large language model technologies. Educational programs should emphasize critical evaluation of AI-generated recommendations rather than passive acceptance. Developing AI literacy across healthcare organizations may ultimately prove as important as technical model development itself.

7. Regulatory Governance and Long-Term Sustainability

The regulatory landscape surrounding generative artificial intelligence in healthcare continues to evolve rapidly. Existing regulatory frameworks were largely developed for traditional medical devices and conventional software systems rather than adaptive language models capable of generating context-dependent outputs. As a result, healthcare organizations deploying LLM-assisted decision support systems face significant uncertainty regarding compliance requirements and liability allocation.

Regulatory authorities increasingly emphasize principles such as transparency, explainability, risk management, and post-deployment monitoring [13]. These principles align closely with the governance requirements identified throughout this study. Rather than treating regulatory compliance as an external obligation, organizations should integrate regulatory objectives directly into system architecture and operational processes.

Model drift represents another major sustainability challenge. Clinical knowledge, treatment guidelines, patient populations, and healthcare delivery practices evolve continuously over time. A model that performs effectively during initial deployment may gradually become misaligned with contemporary clinical realities if ongoing monitoring and updating mechanisms are absent. Sustainable governance therefore requires continuous validation, recalibration, and performance auditing processes.

Environmental sustainability has also emerged as an important consideration in large-scale AI deployment. Training and operating advanced language models require substantial computational resources and energy consumption. Healthcare organizations increasingly face pressure to balance technological innovation with environmental responsibility. Efficient model architectures, optimized inference strategies, and responsible computing practices may become important components of future healthcare AI governance frameworks.

Economic sustainability further influences long-term viability. While LLM-assisted decision support systems may improve care quality and operational efficiency, implementation costs can be substantial. Organizations must evaluate infrastructure investments, maintenance requirements, licensing expenses, workforce training costs, and governance overhead. Comprehensive value assessment should therefore consider both direct financial outcomes and broader organizational benefits such as improved clinical consistency, knowledge accessibility, and patient outcomes.

The most sustainable deployment models are likely to involve adaptive governance structures capable of evolving alongside technological advancements. Static regulatory approaches may prove inadequate in a domain characterized by rapid innovation and continuous capability expansion.

8. Discussion

The emergence of large language models represents a significant transition in the evolution of clinical decision support technologies. Previous generations of healthcare AI systems primarily focused on narrow predictive tasks or deterministic rule execution. In contrast, contemporary language models introduce a fundamentally different paradigm centered on knowledge synthesis, contextual reasoning, and natural language interaction. This shift has important implications for perioperative medicine and regional anesthesia management.

The present analysis suggests that the greatest value of LLM-assisted decision support may not reside in superior predictive accuracy alone. Rather, the technology's transformative potential lies in its ability to integrate fragmented knowledge sources, facilitate interdisciplinary communication, and enhance organizational learning processes. Regional anesthesia selection provides a useful case study because decisions inherently require balancing multiple clinical objectives, operational constraints, and patient preferences.

The proposed framework also highlights the importance of viewing AI implementation as a socio-technical transformation rather than a purely technical innovation. Healthcare outcomes are shaped by interactions among technologies, clinicians, organizations, regulatory institutions, and patients. Consequently, successful deployment depends on governance structures, workflow design, educational initiatives, and cultural adaptation as much as algorithmic sophistication.

Several broader implications emerge from this perspective. First, explainability should be understood as an organizational capability rather than merely a technical feature. Effective explanation requires communication mechanisms that align with clinical reasoning processes and support professional judgment. Second, fairness and equity considerations must be embedded throughout the system lifecycle rather than addressed only during model development. Third, continuous monitoring and adaptive governance are essential because healthcare environments and AI technologies evolve simultaneously.

Future research should explore prospective clinical evaluations of LLM-assisted anesthesia decision support systems across diverse institutional settings. Comparative studies examining clinician trust, workflow efficiency, recommendation quality, patient outcomes, and organizational learning effects would provide valuable evidence regarding real-world impact. Additional investigation is also needed regarding multimodal AI architectures capable of integrating imaging, physiological monitoring, genomic information, and longitudinal patient histories within unified decision-support frameworks.

The long-term future of clinical decision support will likely involve collaborative intelligence models in which human expertise and artificial intelligence capabilities complement one another. Under such arrangements, large language models may function as knowledge amplifiers that enhance clinician performance while preserving professional accountability and ethical oversight. Achieving this vision requires careful attention to governance, transparency, equity, and sustainability considerations from the earliest stages of implementation.

9. Conclusion

This study examined the role of large language model–assisted clinical decision support systems in regional anesthesia selection for knee arthroscopy patients from a system-level perspective. The analysis demonstrated that successful implementation extends beyond algorithmic performance and requires comprehensive consideration of architecture design, interoperability, explainability, fairness, governance, infrastructure, and regulatory compliance.

Large language models offer substantial potential to enhance perioperative decision-making by synthesizing structured and unstructured clinical information, facilitating evidence-informed recommendations, and supporting individualized care planning. However, realizing these benefits depends on the development of robust socio-technical frameworks that preserve clinician oversight, ensure transparency, mitigate bias, and support continuous organizational learning.

The future of AI-enabled perioperative care is likely to be characterized by collaborative intelligence ecosystems in which clinicians and intelligent systems work together to improve decision quality, operational efficiency, and patient outcomes. Within this evolving landscape, governance and sustainability considerations will be as important as technological innovation itself. Healthcare organizations that successfully integrate these dimensions may establish new standards for safe, equitable, and effective deployment of generative artificial intelligence in clinical practice.

References

1. Shortliffe, E. H., & Cimino, J. J. (Eds.). (2014). Biomedical informatics: Computer applications in health care and biomedicine (4th ed.). Springer.

2. Sendak, M. P., D'Arcy, J., Kashyap, S., Gao, M., Nichols, M., Corey, K., & Ratliff, W. (2020). A path for translation of machine learning products into healthcare delivery. EMJ Innovations, 4(1), 38–45.

3. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358.

4. Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2021). The false hope of current approaches to explainable artificial intelligence in health care. The Lancet Digital Health, 3(11), e745–e750.

5. Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H., ... Natarajan, V. (2023). Large language models encode clinical knowledge. Nature, 620(7972), 172–180.

6. Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., ... Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digital Health, 2(2), e0000198.

7. Nori, H., King, N., McKinney, S. M., Carignan, D., & Horvitz, E. (2023). Capabilities of GPT-4 on medical challenge problems. arXiv preprint arXiv:2303.13375.

8. Hashimoto, D. A., Witkowski, E., Gao, L., Meireles, O., & Rosman, G. (2018). Artificial intelligence in anesthesiology: Current techniques, clinical applications, and limitations. Anesthesiology, 129(2), 379–394.

9. Lee, C. K., Hofer, I., & Gabel, E. (2022). Machine learning and perioperative medicine. Anesthesiology Clinics, 40(4), 689–703.

10. Memtsoudis, S. G., Poeran, J., Cozowicz, C., Zubizarreta, N., Ozbek, U., Mazumdar, M., ... Mariano, E. R. (2018). The impact of peripheral nerve blocks on perioperative outcome in hip and knee arthroplasty. Anesthesia & Analgesia, 127(2), 406–415.

11. 金子, 吴川, 王秀丽, & 刘朋. (2014). 股神经联合坐骨神经阻滞用于膝关节镜诊治术. 实用医学杂志, 30(4), 666-667.

12. Hoffman, K. M., Trawalter, S., Axt, J. R., & Oliver, M. N. (2016). Racial bias in pain assessment and treatment recommendations. Proceedings of the National Academy of Sciences, 113(16), 4296–4301.

13. U.S. Food and Drug Administration. (2021). Artificial intelligence and machine learning software as a medical device action plan. FDA.

14. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56.

15. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... Dean, J. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24–29.

16. Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. JAMA, 319(13), 1317–1318.

17. Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., ... Goldenberg, A. (2019). Do no harm: A roadmap for responsible machine learning for health care. Nature Medicine, 25(9), 1337–1340.

18. Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. Health Affairs, 33(7), 1123–1131.

19. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453.

20. Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., ... Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. Stroke and Vascular Neurology, 2(4), 230–243.