ISSN: 2583-7354

International Journal of Emerging Knowledge Studies

Publisher's Home Page: https://www.ijeks.com/

Fully Open Access

Research Paper

The Impact of Large Language Models on Explainability and Transparency in Clinical Decision Support

Lakshmana Rao Padala[1*] Prof. Naresh Tangudu[2]

[1]Assistant Professor / Research Scholar, Aditya Institute of Technology and Management, Tekkali, Srikakulam, Andhra Pradesh, india.

[2]Department of Computer Science and Engineering (Data Science), Aditya Institute of Technology and Management, Tekkali, Srikakulam, Andhra Pradesh, india.

DOI: https://doi.org/10.70333/ijeks-04-05-025

*Corresponding Author: lakshmani003@gmail.com

Article Info:- **Received** : 17 March 2025 **Accepted** : 25 April 2025 **Published** : 30 May 2025

Abstract

The integration of Large Language Models (LLMs) into Clinical Decision Support Systems (CDSS) has emerged as a transformative advancement in healthcare, enabling the processing of unstructured clinical data, automated summarization, and improved diagnostic reasoning. While these systems show significant promise, their widespread adoption is constrained by persistent concerns regarding explainability and transparency. This study reviews recent literature to examine how LLMs influence interpretability, accountability, and clinician trust in CDSS. Findings indicate that LLMs can enhance efficiency and decision accuracy across domains such as radiology, oncology, and pathology. However, their "black-box" nature presents challenges, as clinicians often lack visibility into how outputs are generated. Post-hoc explainability methods such as SHAP, LIME, and Grad-CAM contribute to interpretability, but their scope is limited, and they cannot fully capture the reasoning processes of complex models. Transparency issues, including data provenance, process auditability, and compliance with regulatory frameworks such as the EU AI Act and HIPAA, further restrict clinical acceptance. The study concludes that the future of LLM-driven CDSS depends on explainability-first design, stronger governance frameworks, and empirical validation in real-world settings. By addressing these concerns, LLMs can evolve into trustworthy, transparent, and effective tools for advancing safe and accountable healthcare.

Keywords: *Large Language Models, Clinical Decision Support, Explainable AI, Transparency, Trustworthy AI.*

1. INTRODUCTION

Artificial Intelligence (AI) is transforming healthcare through innovations in diagnostics, treatment planning, and patient management. Among these advancements, Large Language Models (LLMs) such as GPT-4, Med-PaLM 2, and domain-specific adaptations have emerged as powerful tools for handling unstructured clinical data, supporting medical decision-making, and enhancing clinician–patient communication **(Jung, 2025; Nazi & Peng, 2024)**. Their capacity to generate human-like text and synthesize information across diverse sources has made them increasingly attractive for integration into Clinical Decision Support Systems (CDSS) **(Papageorgiou et al., 2025)**.

However, despite their potential, LLMs raise significant concerns regarding explainability and transparency. CDSS tools are high-stakes applications where clinicians must not only trust the accuracy of recommendations but also understand the rationale behind them. Opaque or "black-box" models risk undermining physician confidence, patient safety, and regulatory compliance **(Pierce et al., 2022; Quttainah et al., 2024)**. For example, while traditional rule-based CDSS systems provided logic that was relatively transparent, LLM-based systems often rely on complex, non-interpretable deep learning architectures, making their decision processes difficult to audit **(Rane et al., 2023)**.

Recent studies demonstrate both the promise and limitations of LLMs in healthcare. In radiology, LLMs have been used to summarize reports and provide differential diagnoses, with improvements in diagnostic-related group (DRG) assignment but risks of hallucinated findings that reduce reliability **(Papageorgiou et al., 2025)**. In oncology, a proof-of-concept small language model achieved 86% concordance with multidisciplinary tumor boards, showing the potential for interpretable, guideline-driven decision support **(Griewing et al., 2024)**. Similarly, explainable AI (XAI) frameworks such as SHAP and LIME have been applied to interpret model outputs in pathology and genomics, fostering clinician trust while exposing challenges of bias and computational cost **(Patidar et al., 2024; Umerenkov et al., 2023)**.

The tension between accuracy and explainability has become a focal point of scholarly debate. Some argue that accuracy alone provides sufficient justification for clinical decisions, while others stress that semantic transparency and causal interpretability are essential for both current and future patients **(Pierce et al., 2022)**. Regulatory frameworks, including the EU AI Act and existing healthcare compliance standards such as HIPAA and GDPR, now emphasize accountability and traceability, underscoring the need for LLMs that are not only powerful but also auditable and trustworthy **(Quttainah et al., 2024)**.

2. OBJECTIVES OF THE STUDY

- To analyze the role of Large Language Models (LLMs) in Clinical Decision Support Systems (CDSS).
- To evaluate their impact on explainability and transparency in clinical decision-making.
- To identify key challenges and risks such as bias, hallucinations, and privacy issues.
- To review existing frameworks and methods (e.g., SHAP, LIME, CUC-FATE) for enhancing trust.
- To propose future directions for explainability-first LLM-CDSS design.

3. SIGNIFICANCE OF THE STUDY

This study is significant because it addresses the trust gap in adopting Large Language Models (LLMs) for Clinical Decision Support Systems (CDSS). While LLMs promise improved diagnostic accuracy, documentation support, and patient–clinician communication, their black-box nature raises concerns about safety, accountability, and ethical use. By focusing on explainability and transparency, the research contributes to building trustworthy AI in healthcare, supporting clinicians in making informed decisions, and guiding policymakers in shaping regulatory frameworks.

4. SCOPE OF THE STUDY

4.1 The scope of this research includes:

- A review of current applications of LLMs in CDSS across domains such as radiology, pathology, oncology, and genomics.
- Examination of explainability techniques (e.g., SHAP, LIME, Grad-CAM, CUC-FATE) that enhance model interpretability.
- Analysis of transparency issues, including semantic clarity, accountability, and

compliance with regulatory frameworks (HIPAA, GDPR, EU AI Act).

- Identification of challenges and risks, such as hallucinations, bias, and workflow integration barriers.
- Recommendations for future directions, including explainability-first model design, multimodal integration, and federated learning.

5. LITERATURE REVIEW

Clinical Decision Support Systems (CDSS) have long been used in healthcare to assist physicians with diagnosis, treatment planning, and patient management. Traditional CDSS models relied heavily on rule-based algorithms and structured data from electronic health records (EHRs). While these systems provided a high degree of transparency, since clinicians could easily trace the rules behind each recommendation, their flexibility was limited and they were unable to process unstructured medical information such as free-text notes or imaging reports **(Pierce et al., 2022)**. With the development of deep learning and large language models (LLMs), CDSS have gained new capabilities to analyze multimodal data, but this advancement has come at the cost of reduced interpretability **(Nazi & Peng, 2024)**.

The emergence of LLMs in healthcare has introduced powerful new applications. Models such as GPT-4, Med-PaLM 2, and domain-specific small language models have been integrated into clinical workflows to support tasks ranging from radiology report summarization and coding to oncology decision-making and patient–doctor communication **(Jung, 2025; Papageorgiou et al., 2025)**. In oncology, for instance, a proof-of-concept study demonstrated that a small, guideline-driven LLM achieved 86% concordance with multidisciplinary tumor boards, suggesting the feasibility of interpretable and source-controlled clinical support **(Griewing et al., 2024)**. Similarly, in pathology, fine-tuning of embeddings revealed structured domain knowledge maps, enabling more interpretable decision outputs and improving the transparency of diagnostic processes **(Kraišniković et al., 2025)**. These findings highlight both the opportunities and risks inherent in adopting LLMs for critical medical decisions.

The concept of explainability in medical AI has become a central focus of research. Explainability refers to the extent to which an AI system's predictions can be understood and trusted by clinicians and patients. A wide range of Explainable AI (XAI) methods have been proposed, including SHAP, LIME, Grad-CAM, and attention visualization techniques. These approaches clarify feature importance, highlight relevant regions in images, and provide both local and global interpretability, thereby improving clinician trust in AI-generated outputs **(Rane et al., 2023; Patidar et al., 2024)**. Despite this, debates continue about whether accuracy alone provides sufficient justification for clinical decision-making or whether causal explanations are ethically necessary. **Pierce et al. (2022)** argue that while accuracy may suffice for immediate treatment decisions, long-term improvements in healthcare require semantic transparency and scientific understanding of AI reasoning. Complementing this, **Quttainah et al. (2024)** proposed the CUC-FATE framework (Cost, Usability, Credibility, Fairness, Accountability, Transparency, Explainability), which highlights explainability and transparency as central enablers of safe and effective LLM adoption in medicine.

Closely related to explainability is the principle of transparency, which extends beyond interpretability of model outputs to include auditability, accountability, and traceability. Transparency ensures that clinicians and regulators can evaluate not only what a model predicts but also how and why those predictions are generated **(Jo & Raj, 2024)**. This issue is especially critical in light of regulatory frameworks such as the EU AI Act, HIPAA, and GDPR, which emphasize the need for accountability in high-stakes decision-making systems **(Quttainah et al., 2024)**. A recent survey of XAI methods concluded that semantic transparency—the ability to interpret how an AI system processes data and generates outputs—is essential to building physician confidence in LLM-based CDSS **(Patidar et al., 2024)**. Conversely, when transparency is absent, even explainable results risk being rejected by clinicians due to mistrust, highlighting the need for robust accountability mechanisms **(Umerenkov et al., 2023)**.

Despite these advances, the adoption of LLMs in CDSS faces substantial challenges and risks. One of the most pressing issues is the problem of hallucinations, where LLMs produce convincing but incorrect outputs, which can compromise patient safety (Jung, 2025). Bias in training data also poses a significant threat, as it may lead to inequitable care and exacerbate existing health disparities (Rane et al., 2023). Practical barriers, such as integrating LLMs into hospital EHR systems, ensuring interoperability, and protecting data privacy, further complicate large-scale implementation (Papageorgiou et al., 2025). Ethical dilemmas remain unresolved, particularly regarding accountability: in cases of AI-driven medical errors, it is unclear whether responsibility lies with the clinician, the institution, or the developers of the model (Pierce et al., 2022).

In summary, existing literature highlights the transformative potential of LLMs in clinical decision-making while simultaneously pointing to critical issues of explainability, transparency, and trust. Although several frameworks and XAI techniques have been developed, there remains a gap in systematically analyzing how LLMs specifically affect both explainability and transparency together in CDSS. Most studies either emphasize the technical dimensions of interpretability or focus narrowly on clinical applications, leaving an urgent need for integrative research that bridges these perspectives. This study seeks to address that gap, contributing to the design of trustworthy, explainable, and transparent AI systems that can be responsibly adopted in healthcare.

6. METHODOLOGY

This study employs a narrative literature review design with elements of systematic analysis to examine the impact of large language models (LLMs) on explainability and transparency in Clinical Decision Support Systems (CDSS). A literature-based approach was chosen because the field of LLM-driven healthcare is still rapidly evolving, and existing evidence is widely dispersed across technical, clinical, and ethical domains. By synthesizing findings from prior studies, this review aims to build a coherent understanding of how LLMs influence interpretability, accountability, and trust in medical decision-making.

The research process began with the identification of relevant academic sources. Databases such as PubMed, IEEE Xplore, Scopus, SpringerLink, MDPI, and arXiv were systematically searched using keywords including "large language models," "explainable AI," "clinical decision support systems," "transparency," and "interpretability." Reference mining of influential works, such as those by Pierce et al. (2022), Quttainah et al. (2024), and Jung (2025), was also undertaken to ensure comprehensive coverage.

Clear inclusion and exclusion criteria were applied to refine the literature pool. Articles published between 2020 and 2025 were prioritized, reflecting the period of most active development in LLM applications. Studies were included if they directly addressed healthcare applications of LLMs with a focus on explainability, transparency, or interpretability. Eligible works encompassed peer-reviewed journal articles, conference proceedings, doctoral dissertations, and authoritative reviews. Exclusion criteria filtered out studies unrelated to healthcare, those with insufficient focus on LLMs, and non-academic opinion pieces.

The analysis of selected studies followed a thematic framework aligned with the research objectives. Four main dimensions were used for classification: (1) applications of LLMs in CDSS across domains such as radiology, oncology, pathology, and electronic health records; (2) explainability through post-hoc and ante-hoc interpretability methods, including SHAP, LIME, Grad-CAM, and attention visualization; (3) transparency in terms of auditability, accountability, and compliance with regulatory standards; and (4) challenges and risks such as hallucinations, bias amplification, privacy concerns, and barriers to workflow integration. Each study was examined for its context, methodology, findings, and limitations, and the extracted data were compared across cases to identify patterns and divergences.

To ensure clarity, findings from the reviewed studies were consolidated into summary tables and conceptual figures where appropriate. These visual representations were designed to highlight contrasts between traditional rule-based CDSS and LLM-augmented systems, as well as to compare different explainability and transparency approaches. Narrative synthesis was then applied to integrate these insights, providing a balanced

ISSN: 2583-7354

discussion of both the promises and pitfalls of LLM-driven decision support.

Finally, the methodology acknowledges certain limitations. Since this study relies entirely on secondary data, there is a risk of publication bias and variation in methodological quality across the reviewed works. Additionally, as healthcare applications of LLMs are advancing at a rapid pace, some findings may quickly become outdated. Nevertheless, this review offers a valuable foundation for understanding how explainability and transparency are shaped by LLMs and provides directions for future research and practice.

7. FINDINGS AND RESULTS

The review of literature demonstrates that Large Language Models (LLMs) are making a significant impact on the design and adoption of Clinical Decision Support Systems (CDSS). Findings indicate that while LLMs offer unprecedented capacity to process unstructured medical data, summarize clinical narratives, and assist in diagnostic reasoning, they also introduce challenges that directly influence explainability and transparency. The results are organized thematically around applications, explainability, transparency, and challenges.

7.1 Applications of LLMs in Clinical Decision Support

Across diverse clinical domains, LLMs are already being tested or integrated into CDSS. In oncology, Griewing et al. (2024) demonstrated that a transparent, small language model achieved 86% concordance with multidisciplinary tumor boards, suggesting that interpretable AI can complement clinical expertise. In radiology and nuclear medicine, Papageorgiou et al. (2025) reported improved diagnostic-related group (DRG) assignment, which enhanced coding efficiency but raised questions about hallucinated findings. Similarly, in pathology, fine-tuned embeddings successfully revealed domain-specific knowledge, making diagnostic predictions more interpretable (Kraišniković et al., 2025). These applications highlight that LLMs are not only improving efficiency but also reshaping how clinicians engage with AI-generated recommendations.

Table-1: Applications of LLMs in CDSS Across Clinical Domains

Clinical Domain	Example Study	LLM Contribution	Impact on Explainability/Transparency
Oncology	Griewing et al. (2024)	Small LLM for tumor board support	Transparent and interpretable decisions
Radiology	Papageorgiou et al. (2025)	DRG assignment and coding	Enhanced efficiency but hallucination risk
Pathology	Kraišniković et al. (2025)	Fine-tuned embeddings	Domain-specific interpretability
Medical Education	Quttainah et al. (2024)	CUC-FATE framework	Emphasis on fairness, accountability, transparency

7.2 Impact on Explainability

LLMs have introduced new possibilities for explainability in CDSS. Unlike traditional rule-based systems that offered ante-hoc interpretability, LLMs often require post-hoc methods to justify predictions. Techniques such as SHAP and LIME provide feature attribution, while Grad-CAM highlights relevant regions in medical imaging. Studies show that these methods can improve clinician confidence by making AI reasoning more visible (Rane et al., 2023; Patidar et al., 2024). However, results also reveal limitations: explanations may be computationally expensive, inconsistent across cases, or overly simplified for complex medical decisions.

For example, Umerenkov et al. (2023) found that providing LLM-generated explanations for diagnoses improved clinician trust in some cases but also risked misleading interpretations when explanations were vague or inconsistent. Pierce et al. (2022) further argued that explainability must go beyond “surface-level rationales” and instead enable semantic understanding of how models reach decisions.

ISSN: 2583-7354

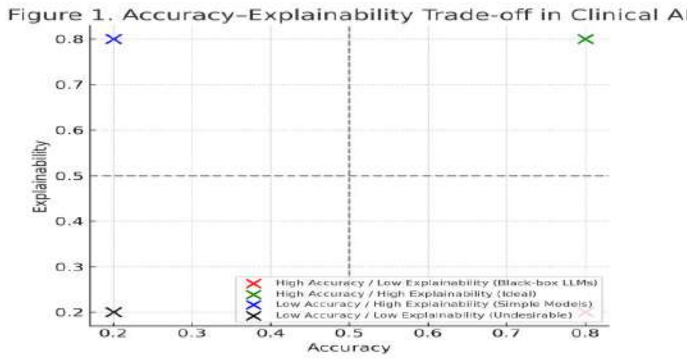

Fig-1: Accuracy–Explainability Trade-off in Clinical AI

- **A quadrant diagram illustrating:**
  - High accuracy/low explainability (black-box deep models).
  - High accuracy/high explainability (ideal but rare).
  - Low accuracy/high explainability (simple but limited models).
  - Low accuracy/low explainability (undesirable).

This figure illustrates the central dilemma in healthcare AI adoption.

7.3 Impact on Transparency

Transparency emerged as a broader concern, extending beyond interpretability of outputs to include data provenance, accountability, and system auditability. Studies highlight that transparency is critical in high-stakes decision-making, especially for compliance with the EU AI Act, GDPR, and HIPAA **(Quttainah et al., 2024)**. **Jo and Raj (2024)** emphasized that LLMs require traceable pipelines where clinicians can understand how input data is processed and how outputs are generated.

Transparency is also linked to trust and adoption. In a proof-of-concept study, clinicians expressed higher confidence in systems where model outputs included citations, confidence levels, and traceable decision paths **(Griewing et al., 2024)**. Conversely, models lacking transparency, even when accurate, were often distrusted.

Table-2: Dimensions of Transparency in LLM-Driven CDSS

Dimension	Description	Example
Data Transparency	Clear source of training and clinical data	Documented datasets in radiology
Process Transparency	Visibility into model pipeline	Algorithm flowcharts for CDSS
Decision Transparency	Interpretability of output	Confidence scores with citations
Regulatory Transparency	Compliance with legal frameworks	GDPR-compliant AI audit trails

7.4 Challenges and Risks Identified

Despite progress, findings highlight persistent risks. Hallucinations remain a major concern, as LLMs sometimes generate plausible but incorrect outputs, potentially endangering patients **(Jung, 2025)**. Bias amplification was frequently reported, with LLMs reflecting disparities present in their training data **(Rane et al., 2023)**. Another challenge is workflow integration: embedding LLMs into EHRs requires interoperability, validation, and clinician training, which remain underdeveloped **(Papageorgiou et al., 2025)**. Finally, accountability gaps persist: when AI-driven errors occur, responsibility is often unclear between clinicians, institutions, and developers **(Pierce et al., 2022)**.

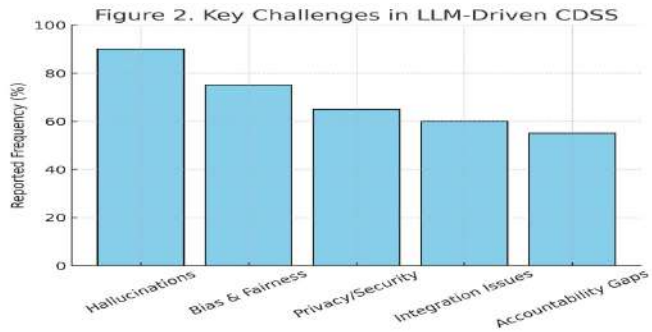

Fig-2: Key Challenges in LLM-Driven CDSS

- **A bar chart showing frequency of reported issues in literature:**
  - Hallucinations (most common).
  - Bias and fairness concerns.
  - Data privacy/security issues.
  - Integration difficulties.
  - Accountability gaps.

7.5 Synthesis of Results

Overall, the findings suggest that LLMs offer substantial benefits for efficiency, domain coverage, and interpretability when paired with XAI methods, but their integration into CDSS remains limited by trust, transparency, and regulatory concerns. The results converge on three insights:

- LLMs can enhance diagnostic accuracy and clinician support but only if supported by robust explainability frameworks.
- Transparency is essential not only for clinicians but also for policy compliance and ethical responsibility.
- Without addressing risks of hallucinations, bias, and accountability, LLM-driven CDSS may face resistance despite technical advances.

8. DISCUSSION

The findings of this study reveal that Large Language Models (LLMs) are reshaping the future of Clinical Decision Support Systems (CDSS) by enabling more sophisticated data interpretation and diagnostic reasoning, yet their integration into healthcare remains hindered by persistent concerns over explainability and transparency. The discussion reflects on these findings in the broader context of clinical practice, ethical considerations, and technological development.

One of the central insights is the accuracy–explainability trade-off that characterizes most AI systems. While LLMs have demonstrated high performance across domains such as oncology, radiology, and pathology, their predictions often lack sufficient interpretability for clinicians to trust them without reservation. This aligns with **Pierce et al. (2022)**, who argued that accuracy alone cannot justify clinical reliance without transparent rationales. The introduction of post-hoc interpretability methods such as SHAP, LIME, and Grad-CAM has mitigated this challenge to some extent, but the results indicate that these methods may offer partial or context-dependent explanations rather than fully faithful interpretations of model reasoning. This tension underscores the need for explainability-first design principles, where transparency is embedded during model development rather than retrofitted afterward.

The results also highlight that transparency is not merely a technical issue but a multidimensional construct involving data provenance, accountability, and regulatory compliance. Transparency ensures that AI-driven decisions can be traced, audited, and justified in alignment with legal frameworks such as the EU AI Act, HIPAA, and GDPR **(Quttainah et al., 2024)**. In this regard, LLMs pose unique challenges: their training data often remain opaque, their decision pathways non-linear, and their outputs probabilistic. This contrasts with traditional CDSS, which were more limited in scope but offered explicit logic that could be directly inspected. The literature therefore suggests that transparency must be pursued not only at the level of model outputs but also at the level of data pipelines and governance structures.

Another significant theme is the trust gap between clinicians and AI systems. Studies such as **Griewing et al. (2024)** demonstrated that interpretable small language models were more readily accepted by physicians, suggesting that trust is linked to transparency of operation rather than sheer model complexity. Clinicians are more inclined to adopt AI recommendations when they are accompanied by citations, confidence scores, and decision traces, features that make the system accountable and align with established medical reasoning practices. Without these, even accurate systems may be met with skepticism or outright rejection, slowing adoption in clinical environments.

The study further emphasizes the risks associated with LLM adoption in healthcare. Hallucinations—the generation of plausible but incorrect information—were identified as one of the most serious threats, raising concerns about patient safety **(Jung, 2025)**. Similarly, bias and fairness issues persist, as LLMs trained on historical medical data may inadvertently reinforce existing inequalities in care delivery **(Rane et al., 2023)**. These risks highlight that explainability and transparency are not merely desirable attributes but essential safeguards against harm. The challenge lies in balancing the innovative potential of LLMs with rigorous ethical oversight.

When compared with traditional CDSS, LLM-augmented systems offer greater adaptability and multimodal capabilities but at the expense of interpretability and clarity. This comparison suggests that the future of CDSS may depend on hybrid systems that integrate the flexibility of LLMs with the traceability of rule-based logic. Such systems would allow clinicians to benefit from advanced reasoning while maintaining the ability to audit and verify outputs.

Finally, the findings carry important policy and governance implications. Regulators and healthcare institutions must move toward frameworks that mandate transparency reporting, standardized interpretability metrics, and robust auditing processes for AI systems. Without these safeguards, the adoption of LLMs in CDSS risks being fragmented and untrustworthy. The growing emphasis on trustworthy AI in healthcare therefore aligns with both clinician expectations and public interest in safe and accountable AI systems.

9. FUTURE DIRECTIONS

The integration of Large Language Models (LLMs) into Clinical Decision Support Systems (CDSS) presents both opportunities and challenges, and the future of this field will depend on the ability to balance innovation with trustworthiness. One of the most promising directions is the development of explainability-first LLMs, where interpretability is embedded into the architecture rather than added as a post-hoc solution. This shift would require interdisciplinary collaboration between computer scientists, clinicians, and ethicists to ensure that models are designed with clarity and accountability from the outset.

Another critical direction is the evolution of multimodal LLMs that can process text, imaging, genomic, and physiological data simultaneously. By integrating diverse modalities, these models have the potential to provide more holistic insights into patient care. However, such complexity also demands stronger frameworks for interpretability. Future research should explore how multimodal reasoning can be made transparent to clinicians through intuitive visualizations, structured explanations, and domain-specific fine-tuning.

Federated learning and privacy-preserving techniques will also play a pivotal role. As LLMs require large-scale datasets for training, ensuring data confidentiality while maintaining transparency in model behavior will be vital. Techniques such as federated learning, differential privacy, and secure model auditing can help reconcile the tension between patient data protection and the demand for transparent decision-making pathways.

Policy and governance frameworks will shape the trajectory of LLM adoption in healthcare. The EU AI Act, HIPAA, and GDPR already provide foundations for accountability, but future regulations must go further by mandating standardized interpretability metrics, independent audits, and explainability benchmarks. Aligning LLM deployment with these regulatory expectations will be crucial for scaling their adoption in clinical environments.

Clinician training and human–AI collaboration must also evolve in parallel. Future CDSS should not only generate transparent outputs but also incorporate features such as confidence scores, uncertainty quantification, and evidence-based citations, empowering clinicians to evaluate AI recommendations critically. Educational programs that equip healthcare professionals with AI literacy will be necessary to ensure safe and effective adoption.

Lastly, more empirical research is needed to test LLM-CDSS in real-world settings. Current findings are often limited to proof-of-concept studies or laboratory evaluations. Future work should involve longitudinal clinical trials to assess the reliability, explainability, and trustworthiness of LLMs across diverse patient populations and clinical contexts. Such studies will provide the evidence base required to transition from experimental systems to fully deployed, transparent, and accountable CDSS.

In sum, the future of LLM-driven CDSS lies in designing models that are not only powerful but also interpretable, transparent, and trustworthy. Achieving this vision will require advancements in technical frameworks, regulatory alignment, clinician engagement, and real-world validation. Only through such multidimensional progress can LLMs fulfill their potential to transform healthcare decision-making responsibly.

10. CONCLUSION

The rise of Large Language Models (LLMs) has opened new possibilities for enhancing Clinical Decision Support Systems (CDSS), offering advanced capabilities in processing unstructured medical data, generating diagnostic insights, and improving patient–clinician interactions. However, this review demonstrates that their adoption is inseparable from the dual imperatives of explainability and transparency. While LLMs can improve efficiency and accuracy in domains such as radiology, oncology, and pathology, their "black-box" nature often limits clinician trust and raises concerns over accountability and safety.

Findings from the reviewed literature show that post-hoc explainability methods like SHAP, LIME, and Grad-CAM contribute to making LLM outputs interpretable, yet they are not sufficient to fully address the complexity of medical reasoning. Similarly, transparency at multiple levels—including data provenance, process auditability, and compliance with legal frameworks—remains an essential but underdeveloped aspect of LLM deployment. The risks of hallucinations, bias, and unclear accountability further highlight the urgency of explainability-first design approaches.

Ultimately, the study underscores that the future of LLM-driven CDSS depends on creating systems that are not only accurate but also auditable, accountable, and trustworthy. Achieving this will require interdisciplinary collaboration, integration of ethical frameworks, robust regulatory oversight, and continuous clinical validation. If these challenges are addressed, LLMs hold the potential to transform CDSS into tools that are not only innovative but also safe, transparent, and aligned with the ethical obligations of healthcare.

REFERENCES

Chittimalla, S. K., & Potluri, L. K. M. (2025, March). Explainable AI frameworks for large language models in high-stakes decision-making. In 2025 International Conference on Advanced Computing Technologies (ICoACT) (pp. 1–6). IEEE.

Griewing, S., Lechner, F., Gremke, N., Lukac, S., Janni, W., Wallwiener, M., ... & Kuhn, S. (2024). Proof-of-concept study of a small language model chatbot for breast cancer decision support: A transparent, source-controlled, explainable and data-secure approach. Journal of Cancer Research and Clinical Oncology, 150(10), 451–460. https://doi.org/10.1007/s00432-024-05964-3

Hsu, A. (2025). Toward trustworthy language models: Interpretation methods and clinical decision support applications (Doctoral dissertation, University of California, Berkeley).

Jo, A. A., & Raj, E. D. (2024, November). Exploring explainable AI in large language models: Enhancing transparency and trust. In 2024 11th International Conference on Advances in Computing and Communications (ICACC) (pp. 1–7). IEEE.

Jung, K. H. (2025). Large language models in medicine: Clinical applications, technical challenges, and ethical considerations. Healthcare Informatics Research, 31(2), 114–124. https://doi.org/10.4258/hir.2025.31.2.114

Kraišniković, C., Harb, R., Plass, M., Al Zoughbi, W., Holzinger, A., & Müller, H. (2025). Fine-tuning language model embeddings to reveal domain knowledge: An explainable artificial intelligence perspective on medical decision making. Engineering Applications of Artificial Intelligence, 139, 109561. https://doi.org/10.1016/j.engappai.2025.109561

Nazi, Z. A., & Peng, W. (2024, August). Large language models in healthcare and medical domain: A review. Informatics, 11(3), 57. https://doi.org/10.3390/informatics11030057

Papageorgiou, P. S., Christodoulou, R. C., Pitsillos, R., Petrou, V., Vamvouras, G., Kormentza, E. V., ... & Georgiou, M. F. (2025). 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, 15(16), 9005. https://doi.org/10.3390/app15169005

Patidar, N., Mishra, S., Jain, R., Prajapati, D., Solanki, A., Suthar, R., ... & Patel, H. (2024). Transparency in AI decision making: A survey of explainable AI methods and applications. Advances of Robotic Technology, 2(1), 000110. https://doi.org/10.23880/art-16000110

Pierce, R. L., Van Biesen, W., Van Cauwenberge, D., Decruyenaere, J., & Sterckx, S. (2022). Explainability in medicine in an era of AI-based clinical decision support systems. Frontiers in Genetics, 13, 903600. https://doi.org/10.3389/fgene.2022.903600

Quttainah, M., Mishra, V., Madakam, S., Lurie, Y., & Mark, S. (2024). Cost, usability, credibility, fairness, accountability, transparency, and explainability framework for safe and effective large language models in medical education: Narrative review and qualitative study. JMIR AI, 3(1), e51834. https://doi.org/10.2196/51834

Rajashekar, N. C., Shin, Y. E., Pu, Y., Chung, S., You, K., Giuffre, M., ... & Shung, D. (2024, May). Human-algorithmic interaction using a large language model-augmented artificial intelligence clinical decision support system. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (pp. 1–20). ACM. https://doi.org/10.1145/3613904.3642024

Rane, N., Choudhary, S., & Rane, J. (2023). Explainable artificial intelligence (XAI) in healthcare: Interpretable models for clinical decision support. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4637897

Scientific, L. L. (2025). Revolutionizing healthcare with large language models: Advancements, challenges, and future prospects in AI-driven diagnostics and decision support. Journal of Theoretical and Applied Information Technology, 103(9), 1450–1468.

Shrivastava, G., Veeraiah, V., Praveenkumar, S., Pathak, P., Sharma, T., & Gupta, A. (2024, June). Explainable AI in NLP: Interpretable models for transparent decision making. In 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1–6). IEEE.

Talukder, N. (2024). Clinical decision support system: An explainable AI approach (Master's thesis).

Umerenkov, D., Zubkova, G., & Nesterov, A. (2023). Deciphering diagnoses: How large language models explanations influence clinical decision making. arXiv preprint arXiv:2310.01708. https://doi.org/10.48550/arXiv.2310.01708

Veeramani, M., Karthick, P., Venkateswaran, S., Sriman, B., Bhanu, S. T., & Devi, V. S. (2025). Transparency in text: Unraveling explainability in healthcare natural language processing. In Explainable Artificial Intelligence in the Healthcare Industry (pp. 131–160). Springer.

Wankhade, K., Jain, D., & Suryavanshi, S. (2024). Developing interpretable large language models for high-stakes decision-making in healthcare: Insights from an XAI review perspective. [Conference/Journal name if available].

Xuyan, H., Meng, S., Chengxing, S., Haoxuan, L., & Jianlin, Z. (2025). Visual-language reasoning large language models for primary care: Advancing clinical decision support through multimodal AI. The Visual Computer. Advance online publication.

Cite this article as: Lakshmana Rao Padala and Prof. Naresh Tangudu., (2025). The Impact of Large Language Models on Explainability and Transparency in Clinical Decision Support. International Journal of Emerging Knowledge Studies. 4(5), pp. 731 - 740. https://doi.org/10.70333/ijeks-04-05-025