In recent months, innovation in generative artificial intelligence using Large Language Models (LLMs) has accelerated dramatically. The creation of LLM strained by deep neural networks utilizing enormous text datasets and typically having billions of parameters, has significantly advanced natural language processing (NLP)[1]. Google first introduced the "Transformer" architecture for machine translation in 2017, and it subsequently achieved state-of-the-art performance in numerous NLP applications[2]. This led to the development of numerous LLMs with "Transformer" architecture, including BERT (Google) [3], GPT (OpenAI) [4], LLaMA (Meta AI)[5].

The rapid evolution of LLMs, such as OpenAI's GPT-3 and Google’s BERT, has demonstrated their potential to process and generate human-like text by leveraging vast amounts of training data [3], [6]. This development has significantly transformed various fields, including finance, education, transportation, and healthcare, leading to increased innovation and efficiency. In particular, the impact of LLMs on healthcare is remarkable. These models have the potential to greatly improve diagnostic accuracy, refine treatment plans, and enhance patient outcomes. A key advantage of LLMs is their capability to deliver clear and understandable explanations, which helps build trust and comprehension.

The integration of advanced computational models into healthcare has ushered in a new era of clinical decision support systems (CDSS), which are essential for enhancing diagnostic accuracy and improving patient outcomes. In the context of CDSS, LLMs can synthesize information from diverse sources, thereby assisting healthcare professionals in making informed decisions.

At the forefront of this technological revolution is the application of LLMs trained on diverse datasets, including academic literature, clinical guidelines, and patient records. These models, when coupled with Retrieval-Augmented Generation (RAG)[7] techniques and agent-based architectures[8] provide a robust framework for retrieving pertinent information from extensive knowledge bases.

However, the sheer volume of medical literature and the complexity of clinical information necessitate the use of RAG techniques to ensure that the most relevant and up-to-date knowledge is retrieved [7]. RAG is a hybrid approach that combines the generative capabilities of LLMs with retrieval systems to enhance the quality of generated responses. By retrieving relevant documents or data points from a knowledge base and integrating them into the response generation process, RAG significantly improves the accuracy and relevance of the information presented [9]. In a healthcare setting, this means that when a clinician queries the system regarding a specific disease or symptom, the model can retrieve the latest research findings, treatment guidelines, and case studies, thereby providing a comprehensive response that is grounded in evidence.

The incorporation of agent-based architectures further enhances the functionality of CDSS. Agents act autonomously to perform specific tasks, such as monitoring patient data, analyzing trends, and retrieving information from various sources. This autonomy allows for real-time decision-making and facilitates a more interactive experience for healthcare providers. For instance, an agent could continuously analyze a patient's vital signs and medical history, alerting the clinician to any anomalies that may require immediate attention. By combining LLMs with agent-based systems, we have created a dynamic and responsive CDSS that adapts to the needs of both clinicians and patients.

An essential aspect of deploying LLMs and RAG in healthcare is the need for an argumentative framework that allows the system to structure its reasoning and provide justifications for its diagnoses. Argumentative frameworks, which are rooted in formal logic and computational models of argumentation, enable systems to present evidence, counterarguments, and conclusions coherently. In a clinical context, this means that when the model suggests a diagnosis, it can articulate the rationale behind its recommendation by referencing specific studies, guidelines, or patient data. This not only aids in clinician understanding but also fosters trust in the system's recommendations.

Moreover, the use of XAI techniques is crucial for ensuring that the decision-making processes of LLMs are transparent and interpretable. XAI aims to make AI systems more understandable to users by providing insights into how decisions are made. In healthcare, where the stakes are high, clinicians must understand the reasoning behind a model's diagnosis or treatment recommendation. Techniques such as feature importance analysis, saliency maps, and model-agnostic explanations can help elucidate the factors that influenced the model's output [10].This enables the model to not only diagnose diseases but also to elucidate its reasoning through Explainable Artificial Intelligence (XAI)[11] techniques. By integrating XAI into CDSS, healthcare providers can make more informed decisions, ultimately leading to better patient care.

The convergence of LLMs, RAG, agent-based systems, argumentative frameworks, and XAI represents a significant advancement in the development of CDSS within the healthcare domain. This holistic approach not only enhances the accuracy and relevance of clinical recommendations but also addresses the critical need for transparency and interpretability in AI-driven healthcare solutions. As the healthcare landscape continues to evolve, the integration of these technologies will be pivotal in shaping the future of clinical decision-making.

In summary, the combination of LLMs trained on diverse datasets, RAG techniques for information retrieval, agent-based architectures for autonomous decision-making, argumentative frameworks for structured reasoning, and XAI for transparency presents a comprehensive solution for improving CDSS in healthcare. The potential benefits of this integrated approach are vast, ranging from enhanced diagnostic accuracy to improved clinician-patient communication. This introduction aims to explore the intricate interplay between these components and their implications for the healthcare domain.

References:

[1] R. Yang, T. F. Tan, W. Lu, A. J. Thirunavukarasu, D. S. W. Ting, and N. Liu, “Large language models in health care: Development, applications, and challenges,” *Health Care Sci.*, vol. 2, no. 4, pp. 255–263, Aug. 2023, doi: 10.1002/hcs2.61.

[2] A. Vaswani *et al.*, “Attention is All you Need”.

[3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” May 24, 2019, *arXiv*: arXiv:1810.04805. Accessed: Apr. 06, 2024. [Online]. Available: http://arxiv.org/abs/1810.04805

[4] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, “Improving Language Understanding by Generative Pre-Training”.

[5] H. Touvron *et al.*, “LLaMA: Open and Efficient Foundation Language Models,” Feb. 27, 2023, *arXiv*: arXiv:2302.13971. Accessed: Mar. 24, 2024. [Online]. Available: http://arxiv.org/abs/2302.13971

[6] T. Brown *et al.*, “Language Models are Few-Shot Learners,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2020, pp. 1877–1901. Accessed: May 24, 2025. [Online]. Available: https://proceedings.neurips.cc/paper\_files/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html

[7] P. Lewis *et al.*, “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2020, pp. 9459–9474. Accessed: May 24, 2025. [Online]. Available: https://proceedings.neurips.cc/paper\_files/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html

[8] C. Chen, B. Yao, Y. Ye, D. Wang, and T. J.-J. Li, “Evaluating the LLM Agents for Simulating Humanoid Behavior”.

[9] V. Karpukhin *et al.*, “Dense Passage Retrieval for Open-Domain Question Answering,” Sep. 30, 2020, *arXiv*: arXiv:2004.04906. doi: 10.48550/arXiv.2004.04906.

[10] F. Doshi-Velez and B. Kim, “Towards A Rigorous Science of Interpretable Machine Learning,” 2017, doi: 10.48550/ARXIV.1702.08608.

[11] V. Chamola, V. Hassija, A. R. Sulthana, D. Ghosh, D. Dhingra, and B. Sikdar, “A Review of Trustworthy and Explainable Artificial Intelligence (XAI),” *IEEE Access*, vol. 11, pp. 78994–79015, 2023, doi: 10.1109/ACCESS.2023.3294569.