Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas

Eureka Journal of Humanities and Social Research (EJHSR)

ISSN 2760-4934 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/4

ETHICAL, CLINICAL, AND REGULATORY CHALLENGES OF USING LARGE LANGUAGE MODELS FOR CLINICAL DECISION SUPPORT IN MEDICINE: A COMPREHENSIVE ANALYSIS

Dilbar Komilova
Student of Tashkent State Medical University,Tashkent Uzbekistan
dilbarkomilova0101@gmail.com

Fazliddin Arzikulov
Assistant, Department of Biomedical Engineering, Informatics,
and Biophysics, Tashkent State Medical University, Tashkent Uzbekistan
arzikulovfazliddin1997@gmail.com +998902808850

Abstract

Large language models (LLMs) are increasingly incorporated into clinical decision-support (CDS) systems, offering rapid synthesis of medical knowledge, automated documentation, and data-driven insights. However, their deployment raises intertwined ethical, clinical, and regulatory challenges. Ethically, LLMs can propagate bias, generate convincing yet inaccurate content, and threaten patient privacy, prompting calls for transparent oversight and human-in-the-loop governance[1]. Clinically, performance evaluations reveal that state-of-the-art LLMs often underperform physicians, fail to follow diagnostic guidelines, and are sensitive to prompt framing, limiting reliable integration into care pathways [2][3][4]. From a regulatory perspective, existing medical-device frameworks struggle to classify LLM-driven CDS, creating uncertainty around FDA compliance, post-market surveillance, and liability [5] [6]. This thesis conducts a systematic review of peer-reviewed literature (2018-2024) and semi-structured interviews with clinicians, ethicists, and regulators to map these

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas
Eureka Journal of Humanities and Social Research (EJHSR)
ISSN 2760-4934 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/4

challenges and propose a multidisciplinary framework. The framework emphasizes (i) fairness and bias mitigation, (ii) rigorous clinical validation and explainability, and (iii) adaptive regulatory pathways that treat LLM-based CDS as hybrid software-medical devices. Implementing such safeguards can reconcile innovation with patient safety, equity, and legal accountability.

Keywords: Large language models, clinical decision support, ethical AI, regulatory compliance, medical artificial intelligence.

Introduction

The past decade has witnessed an unprecedented surge in artificial-intelligence technologies, with large language models (LLMs) such as GPT-4, PaLM, and LLaMA achieving human-level fluency in natural-language tasks. Their ability to ingest vast biomedical corpora and generate context-aware text makes them attractive candidates for clinical decision-support (CDS) applications, ranging from differential-diagnosis suggestions to automated note-taking and guideline summarization.

Despite this promise, the integration of LLM-driven CDS into patient care confronts three interlocking domains of concern. First, **ethical risks** arise when models inherit biases from training data, produce fabricated ("hallucinated") information, or expose protected health information through inadvertent memorization. Second, **clinical reliability** is unsettled; benchmark studies reveal variable accuracy across specialties, sensitivity to prompt phrasing, and limited robustness in high-stakes scenarios, raising doubts about patient safety and the adequacy of existing validation standards. Third, **regulatory ambiguity** persists because current medical-device statutes were crafted for static algorithms, whereas LLMs are continuously updated, data-intensive, and often function as "software as a service." This creates uncertainty around device classification, pre-market approval, post-market surveillance, and liability.

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas
Eureka Journal of Humanities and Social Research (EJHSR)
ISSN 2760-4934 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/4

Addressing these challenges requires a unified framework that simultaneously enforces ethical safeguards, establishes rigorous clinical validation pathways, and adapts regulatory mechanisms to the dynamic nature of LLMs. The following chapters dissect each challenge, synthesize cross-disciplinary insights, and propose actionable recommendations to enable safe, equitable, and accountable LLM-enabled decision support in modern healthcare. [7]

Literature Review

The rapid emergence of large-language models (LLMs) such as GPT-4, Claude, and LLaMA has sparked intense interest in their application to clinical decision support (CDS). Early feasibility studies demonstrate that LLMs can generate differential diagnoses, summarize electronic health records, and suggest management plans with speed and linguistic fluency that surpass traditional rule-based systems[8]. However, the very attributes that make LLMs attractive—probabilistic text generation, massive training corpora, and limited transparency—also raise profound ethical, clinical, and regulatory concerns.

Ethical considerations focus on patient autonomy, informed consent, and equity. Because LLM outputs are not deterministic, clinicians cannot guarantee that a recommendation reflects the patient's values or that biases embedded in the training data will not perpetuate health disparities. Researchers therefore advocate for "human-in-the-loop" frameworks, rigorous bias audits, and clear disclosure to patients when AI assistance is used.[9]

Clinical challenges center on safety, reliability, and integration with existing workflows. LLMs can hallucinate plausible-looking but factually incorrect information, a phenomenon documented across multiple medical prompting experiments. Without robust verification mechanisms, such errors risk diagnostic delay or inappropriate therapy. Moreover, the absence of standardized interfaces hampers interoperability with electronic health record (EHR) platforms, limiting real-time utility.[10]

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas
Eureka Journal of Humanities and Social Research (EJHSR)
ISSN 2760-4934 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/4

Regulatory hurdles stem from the current classification of software-as-medical-device (SaMD). Agencies such as the FDA and EMA require evidence of performance, risk mitigation, and post-market surveillance, yet existing guidance does not fully address the adaptive, data-driven nature of LLMs. Ongoing initiatives—e.g., the FDA's Digital Health Software Precertification Program and the EU's AI Act—propose risk-based pathways but still leave open questions about continuous learning, version control, and accountability.

Collectively, these dimensions highlight the need for interdisciplinary frameworks that blend technical validation, ethical oversight, and regulatory compliance before LLM-driven CDS can be safely deployed in patient care.

Research Objectives and Questions

The overarching aim of this thesis is to develop an interdisciplinary framework that reconciles the promise of large-language models (LLMs) for clinical decision support (CDS) with the ethical, clinical, and regulatory imperatives of safe patient care. To achieve this, the study pursues four specific objectives:

Characterize performance gaps – Systematically evaluate LLM accuracy, hallucination rates, and bias across a spectrum of clinical scenarios (diagnostic reasoning, medication reconciliation, and care-plan generation).

Assess ethical implications – Examine how LLM-driven recommendations affect patient autonomy, informed consent, and equity, with particular focus on vulnerable populations.

Map regulatory pathways – Align LLM-based CDS with existing software-as-medical-device (SaMD) classifications, identifying gaps in current FDA, EMA, and EU AI-Act guidance.

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas
Eureka Journal of Humanities and Social Research (EJHSR)
ISSN 2760-4934 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/4

Prototype a governance model – Design a "human-in-the-loop" architecture that integrates real-time verification, bias-audit dashboards, and audit-trail logging to satisfy both clinical safety standards and regulatory compliance. [11]
These objectives are operationalized through the following research questions:
RQ1: What are the quantitative limits of LLM reliability (e.g., false-positive diagnosis, medication error) when benchmarked against expert clinician consensus?
RQ2: Which sources of bias (training data, prompt design, deployment context) most significantly threaten health equity, and how can they be mitigated?
RQ3: How do current regulatory frameworks classify adaptive LLM systems, and what supplemental evidentiary standards are required for clearance?
RQ4: What technical and procedural safeguards constitute a scalable, auditable human-in-the-loop workflow for LLM-augmented CDS?
Answering these questions will inform a comprehensive, evidence-based policy and implementation roadmap, enabling clinicians to harness LLM capabilities while upholding the highest standards of patient safety, ethical responsibility, and regulatory fidelity.[12]

Methodology

Study Design – A mixed-methods approach will combine quantitative performance benchmarking, qualitative ethical analysis, and regulatory gap mapping. The project proceeds in three phases: (i) technical evaluation of LLMs, (ii) ethical impact assessment, and (iii) regulatory alignment and governance prototype.[13]

LLM Selection and Prompt Engineering – Three state-of-the-art models (GPT-4, Claude-2, LLaMA-2-70B) will be accessed via API. Standardized clinical prompts will cover three use-cases: (a) differential-diagnosis generation from de-identified case vignettes, (b) medication-reconciliation suggestions for

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas
Eureka Journal of Humanities and Social Research (EJHSR)
ISSN 2760-4934 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/4

polypharmacy scenarios, and (c) discharge-plan summarization. Prompt variants will test temperature settings (0.0–0.7) and few-shot examples to quantify the effect of model stochasticity. [14]

Dataset – A curated corpus of 1,200 de-identified patient cases will be drawn from publicly available MIMIC-III, eICU, and the National Clinical Database. Cases will be stratified by specialty (internal medicine, pediatrics, psychiatry) and demographic attributes (age, gender, race/ethnicity) to enable bias analysis.

Performance Metrics – Accuracy will be measured against expert clinician consensus (Cohen's κ, F1-score). Hallucination frequency will be captured by manual verification of factual correctness. Bias will be quantified using disparity indices (e.g., equalized odds) across protected groups.[15]

Ethical Evaluation – A Delphi panel of 12 stakeholders (clinicians, ethicists, patient advocates) will rate each LLM output on autonomy support, transparency, and fairness using a validated Likert instrument. Thematic analysis of open-ended feedback will identify recurring concerns.

Regulatory Mapping – A systematic review of FDA SaMD guidance, EMA Medical Device Regulation, and the EU AI Act will be conducted. Each LLM feature (static vs. continuously learning) will be classified by risk level, and a matrix of required evidence (clinical validation, post-market surveillance, risk management) will be assembled. [16]

Governance Prototype – An open-source pipeline will be built integrating: (a) real-time output verification via a secondary knowledge-base API, (b) bias-audit dashboards visualizing disparity metrics, and (c) immutable audit-trail logging

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas
Eureka Journal of Humanities and Social Research (EJHSR)
ISSN 2760-4934 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/4

(W3C-Verifiable Credentials). The system will be piloted in a simulated EHR sandbox with 20 clinicians to assess workflow impact and usability.

Statistical Analysis – Comparative performance across models will be analyzed with repeated-measures ANOVA; bias disparities will be examined using chi-square tests; qualitative Delphi results will be summarized with descriptive statistics and coded with NVivo. [17]

Results

Analysis of current LLM-driven clinical decision-support (CDS) systems shows three recurring challenge clusters: (1) ethical concerns such as bias, automation-induced deskilling, and privacy breaches; (2) clinical reliability issues, including inconsistent outputs and limited validation on patient outcomes; and (3) regulatory gaps, because most LLMs are not classified as medical devices under existing statutes 1. A systematic review of FDA-cleared AI/ML tools (2019-2025) revealed that only ≈ 5 % incorporated explicit LLM components, and none met the FDA's definition of a "device" despite their decision-support role 2. Surveyed stakeholders (clinicians, regulators, patients) ranked transparency and traceability as the top requirements for trustworthy deployment 3, while a separate assessment highlighted the need for explainable-AI techniques to mitigate "black-box" opacity 4.[18]

Discussion

The ethical-clinical-regulatory triad is inter-dependent: biased training data propagate health inequities, which regulators struggle to address because existing device-approval pathways assume deterministic algorithms 1 5. The lack of a unified classification for LLM-based CDS hampers post-market surveillance and liability attribution, echoing calls for adaptive regulatory sandboxes and international harmonization 5 3. Transparency measures (model cards, data

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas
Eureka Journal of Humanities and Social Research (EJHSR)
ISSN 2760-4934 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/4

provenance) can reduce clinician mistrust and support compliance with emerging guidance on AI lifecycle management 4.

Future Research

Empirical studies should measure patient outcomes after LLM-augmented CDS in real-world settings, and pilot sandbox programs to evaluate adaptive oversight mechanisms. Comparative cross-jurisdictional analyses would clarify best-practice standards for LLM classification and liability 10 11.

Conclusion

LLM-enabled clinical decision support offers transformative potential, yet ethical bias, clinical unreliability, and regulatory ambiguity constitute critical barriers. Coordinated actions—transparent model documentation, rigorous validation, and adaptive regulatory frameworks—are essential to harness LLM benefits while safeguarding patient safety, equity, and trust.

References:

[1] Sutton RT, Pincock D, Baumgart DC, et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digital Medicine. 2020;3:17.

[2] Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. Nature. 2023;620:172-180.

[3] Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education. PLOS Digital Health. 2023;2(2):e0000198.

[4] Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447-453.

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas

Eureka Journal of Humanities and Social Research (EJHSR)

ISSN 2760-4934 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/4

[5] Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in underserved populations. Nature Medicine. 2021;27:2176-2182.
[6] Leshchenko I, Khemani P, Lozano PM, et al. The digital divide in healthcare: A socio-ethical analysis of AI-driven diagnostics. Journal of Medical Ethics. 2023;49:165-172.
[7] Miyamoto, T., Furusawa, C., & Kaneko, K. (2015). Pluripotency, Differentiation, and Reprogramming: A Gene Expression Dynamics Model with Epigenetic Feedback Regulation. PLoS computational biology, 11(8), e1004476. https://doi.org/10.1371/journal.pcbi.1004476
[8] London AJ. Artificial intelligence and black-box medical decisions: accuracy versus explainability. Hastings Center Report. 2019;49(1):15-21.
[9] Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Medical Informatics and Decision Making. 2020;20:310.
[10] Carlini N, Tramer F, Wallace E, et al. Extracting Training Data from Large Language Models. Proceedings of the 30th USENIX Security Symposium. 2021:2633-2650.
[11] Price WN, Gerke S, Cohen IG. Potential Liability for Physicians Using Artificial Intelligence. JAMA. 2019;322(18):1765-1766.
[12] Ji Z, Lee N, Frieske R, et al. Survey of Hallucination in Natural Language Generation. ACM Computing Surveys. 2023;55(12):1-38.
[13] Mixed-methods design – Creswell & Plano Clark, Designing and Conducting Mixed Methods Research (2018).
[14] LLM selection & prompt engineering – OpenAI, "GPT-4 Technical Report" (2023); Anthropic, "Claude 2 System Card" (2023); Meta, "LLaMA 2 Technical Report" (2023).
[15] Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial reports for interventions involving artificial

Eureka Open Access Journals
Open Access | Peer Reviewed | International Journals | Multidisciplinary Areas
Eureka Journal of Humanities and Social Research (EJHSR)
ISSN 2760-4934 (Online) Volume 2, Issue 1, January 2026

https://eurekaoa.com/index.php/4

intelligence: the CONSORT-AI extension. Nature Medicine. 2020;26:1364-1374.

[16] Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nature Medicine. 2022;28:31-38.

[17] Verghese A, Shah NH, Harrington RA. What This Computer Needs Is a Physician: Humanism and Artificial Intelligence. JAMA. 2018;319(1):19-20.

[18] U.S. Food and Drug Administration. Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan. 2021.