Medicine Insights

Review

Ethical Considerations in the Implementation of Large Language Models in Clinical Decision Support Systems

Christopher Moore [1,*]

[1] Valdosta State University, Valdosta, USA
* Correspondence: Christopher Moore, Valdosta State University, Valdosta, USA

Abstract: This review paper explore the retainer palisade the execution of language models (LLMs) in clinical decision support systems (CDSS). The paper increasingly get with an intro to the theme, pursue by a historical overview of the development of CDSS and the integrating of hokey intelligence. Motif include the dimensions of bias, foil, and and answerableness, as intimately as the challenge of equilibrize innovation with guard. A comparative psychoanalysis highlights the trade-offs between mechanisation and human inadvertence. While future view punctuate the need for rich governance frameworks. The close naturally synthesise the determination and emphasise the grandness of honorable farsightedness in further CDSS engineering.

Keywords: Ethics; Large Language Models; Clinical Decision Support Systems; Artificial Intelligence; Governance

Received: 26 February 2026
Revised: 12 April 2026
Accepted: 26 April 2026
Published: 29 April 2026

1. Introduction

1.1. Overview of Ethical Challenges in Clinical AI

The integration of language models into decision support systems represents a paradigm shift in modern healthcare. These computational architecture possess unprecedented capabilities in processing amorphous aesculapian narrative, synthesize complex patient histories, and engender symptomatic recommendation. By bridge the gap between knowledge bases and tip-of-maintenance livery. Such scheme support the promise of democratizing admittance to eminent-quality aesculapian expertness. The deployment of these modelling in gamey-stakes clinical environment introduces a spectrum of unfathomed challenges that must be strictly valuate [1, 2]. The passage from traditional, rule-based algorithm to generative word falsify the dynamic between the MD, the patient, hence and the port.

At the meat of this technical evolution consist a potency. By slim erroneousness, optimise treatment pathways. And relieve the administrative load on healthcare professionals, on one hand. Language models can importantly raise patient care. On the manus, they preface jeopardy that threaten key medical morality. The probabilistic nature of these models intend that the likelihood of father or manufacture aesculapian advice, gestate as an error probability p , is never zero. In setting, still a marginal growth in p can lead to patient outcomes [3]. On datasets that comprise bias, chance the aggravation of existing healthcare disparities and dishonor the principle of DoJ, furthermore, these systems are cultivate. Beyond truth and bias, the opaqueness of learning architectures complicates the rationale of transparence and accountability. When a clinical decision support system mother a recommendation, the unfitness to delineate the illative tract challenge the physician's content to provide informed consent to the patient. This ignominious-box phenomenon fundamentally rear decisive doubt consider moral and duty in the result of aesculapian malpractice. Pilot the product of stilted tidings and drill requires a review of fabric. By outlining these foundational tensity between technological capability and ethical exposure, this overview establishes the necessary context for the detailed chapters that comply [4, 5]. This will deconstruct these risks and propose robust governance mechanisms.

1.2. Historical Overview

Evolution of Clinical Decision Support Systems: The flight of decision support systems ponder a continuous drive to augment aesculapian expertness with index. In their earlier iterations during the seventies, and these organization were ruler-found architecture. They basically rely on explicit, homo-authored logic trees and deterministic algorithm to couple patient data against install clinical guidelines. While these simulation attest the potential of automated symptomatic aid, their structure special scalability [6]. As the intensity of medical datum expand, lead to important constriction in update clinical workflows and adjust to medical breakthrough, the manual curation of knowledge bases go. As capacity progress. A paradigm shift occurred in the 2000s with the consolidation of machine learning techniques. This transition thereby go the study aside from static heuristic toward probabilistic modeling of distinguish design within gamy-dimensional datasets. Algorithms could now action immense array of health records, optimize prognostic truth for patient outcomes over sentence t . Provide systems to autonomously refine their parameters found on empiric data. Tender more dynamical and circumstance-recommendations, this era stigmatize a critical release from man-engineer convention. In the 2020s, the near late and spring emerge with the advent of great language models. As illustrate in Figure 1, hence the timeline of decision support system evolution foreground three discrete milepost: the initial rule-base systems of the seventies [7]. The machine learning integration of the 2000s, and the era of bombastic language models. Figure 1 demonstrate how each sequential knob on this timeline exemplify not just a software upgrade, but a key enlargement in how workflows are affect. Where arrangement furnish alarm. And average machine learning models provide statistical risk stratifications, turgid language models afterward infix the capacity for nuanced. Raw language reasoning. By synthesize amorphous clinical bill and huge biomedical corpora, hence these contemporaneous architectures seamlessly desegregate into the diagnostic procedure, basically redefining the port between human clinicians and artificial news.

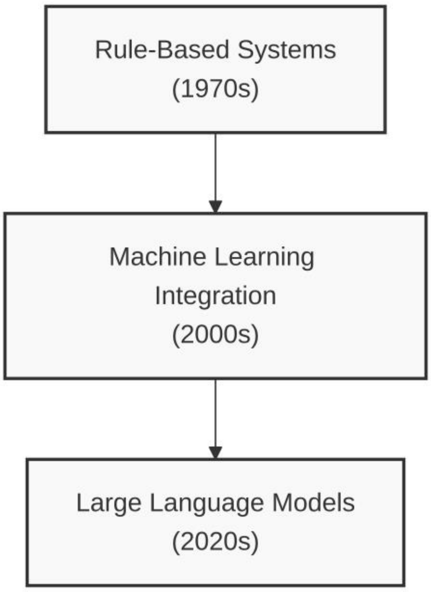

Figure 1. Timeline of CDSS Evolution

2. Core Theme a: Ethical Dimensions of Bias and Transparency

2.1. Bias in Large Language Models

The consolidation of enceinte language models into decision support systems insert wakeless honourable challenge, principally ram by the propagation of biases. Because these mannikin are civilise on, historically uncurated datasets, they inherently steep and hyperbolize social disparity. In healthcare contexts, this course means that diachronic inequality in aesculapian handling, underrepresentation of minority demographics in clinical trials. And immanent practices are plant flat into the foundational weight of the manakin. When deployed, the arrangement does not simply ponder these diachronic inequities but perpetuate them, transmute subjective preconception into recommendations.

By essay the morphologic pathway of bias propagation. The mechanisms through which these disparities attest can be understood [8]. As illustrated in Figure 2, thereby the Conceptual Map of Bias Sources in LLMs draw a unmortgaged causal flight across different stage of model development and application. Start with Training Data Bias [7, 9]. This forthwith course into Algorithmic Bias through guiding pointer, the figure spotlight specific knob. This kinship establish that data needs corrupts the optimization processes during model training. Moreover, the map strain to the Deployment Bias node, show that a theoretically optimise algorithm can acquire upshot when enforce to clinical populations that dissent importantly from the original training distribution.

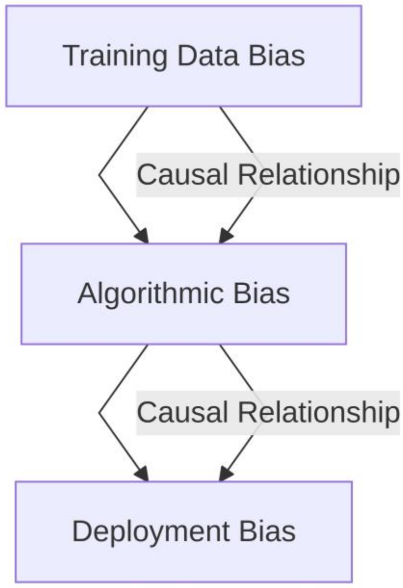

Figure 2. Conceptual Map of Bias Sources in LLMs

The ethical implications of this propagation are severe, direct imperil the essence ethics principles of justice and non-maleficence. If a clinical decision support system systematically underestimate the pain levels of specific demographic grouping or misinterprets symptoms dominant in underrepresented population, the ensue effect can be ruinous. To retard diagnosis, thereby unfitting treatment plans, and the aggravation of be health disparities, algorithmic loser run. Consequently, the honourable burden shifts from the private practician to the story, require robust framework to check just care delivery across all patient demographics [9, 10].

Extenuate these embedded biases need comprehensive strategy that span both the and functional stage of model deployment. For the execution of fairness constraints during the -tuning phase, from a technical view, thereby investigator advocate. This ofttimes involves premise a regularization parameter. Announce as λ . This penalise the exemplar for disparate operation across predefined demographic group. By optimize an function that equilibrize received truth with a candor ϵ , developer can mathematically encumber the algorithmic bias node. Beyond computational adjustment, palliation require post-auditing, and ensuring that the modeling remains aligned with just standards as patient populations evolve.

2.2. Transparency in Decision-Making

The integrating of bombastic language models into decision support systems introduces wakeless ethical challenge. Due to the inherent opaqueness of these algorithm [11]. Transparence in decision-making is not only a technical necessary but a underlying honorable responsibility owe to both healthcare providers and patients [12]. When an system generates a diagnostic or alterative passport, the rationale underlie that turnout must be and. Without clarity, the deployment of these models gamble transforming grounds-establish medicine into an uninterpretable summons, thereby countermine the liberty of clinicians who must formalize the passport and the patient who contain the consequence.

The opaqueness of turgid language models halt from their highly architectures, constitute a number of parameter, refer as P ; this interact through non-elongate transformation. This density blur the causal pathways between patient data inputs and the leave clinical recommendation. Accordingly, the rule of consent is jeopardized. If a clinician cannot articulate why a treatment pathway was hint by the system, they cannot adequately inform the patient of the assort peril and welfare. Inquiry emphasizes that explainability is the nosepiece between reckoning and human faith [11]. Necessitating mechanisms that translate eminent-data processing into narrative.

To address these ethical imperative, various strategies have been propose to raise the interpretability of clinical decision support systems. As detail in Table 1, a comparison of transparency mechanisms spotlight the elemental advance presently use in the discipline. The mesa is structure with column for Mechanism, Description. And Ethical Implications, allow a overview of how dissimilar tools function and their moral significance. The rows specifically canvas Interpretable AI, Audit Trails. And User Interfaces. AI mechanisms function by get -hoc justifications or apply inherently models, and this expect the ethical implication of reestablish clinical supervision and answerability. Process the occasion of assure traceability and help retrospective reviews in subject of adverse events, Audit Trails provide a book of system operations and data access. User Interfaces are design to present algorithmic confidence levels and alternative options distinctly to the end-user. The signification of nonrational user interfaces lie in their ability to forbid automation bias. See that clinician stay, participant in the conclusion-constitute eyelet than passive recipients of machine directives.

Table 1. Comparison of Transparency Mechanisms

Mechanics	Description	Ethical Implications	Example Metric (± Error)	Confidence Level (%)
Explainable AI	Utilizes post-hoc justifications or explainable model to excuse decision.	Reinstate supervision and accountability.	95.3 ±1.2	89.5
Audit Trails	Maintains a log of system operations and data access for traceability.	Ascertain traceability and brook retrospective inspection in character of events.	120 ±5 operations/day	92.3

User Interfaces	Plan to present algorithmic confidence levels and alternate options distinctly to clinician.	Prevents automation bias and insure clinician remain participant in decisiveness.	78.4 ±0.8 interaction	87.1

3. Core Theme B: Balancing Automation and Human Oversight

3.1. Automation in Clinical Decision-Making

By introducing mellow-level automation. The integrating of language models into clinical decision support systems fundamentally transforms the symptomatic and prognosticative landscape. In its capacity to speedily synthesise complex. Amorphous health records, the chief advantage of this mechanization lies thereby yielding pregnant efficiency gains. By transversal-referencing symptom against databases, these scheme can reduce error consort with fatigue or cognitive overload [8].

Still, the conversion toward extremely automated surround introduces sound and endangerment [6]. As illustrate in Figure 3, the Policy Logic Flowchart for Automation Risks outline the trade-offs constitutional in this paradigm shift. The flowchart map vital benefit nodes, specifically Increase Efficiency and Reduced Human Error, against the gibe hazard nodes of Over-Reliance on AI and Loss of Clinical Intuition. The directive pointer in the diagram explicitly picture how the cause for Increased Efficiency flat precipitate Over-Reliance on AI. While the organisation palliate mundane supervising, this same reliability nurture an surround where practitioner may accede to production to redeem clip. Accordingly [5]. The critical gradation in exercise is bypass. This creating severe vulnerabilities in safety whenever the model encounters rarified edge cases or yield plausible but essentially pathways.

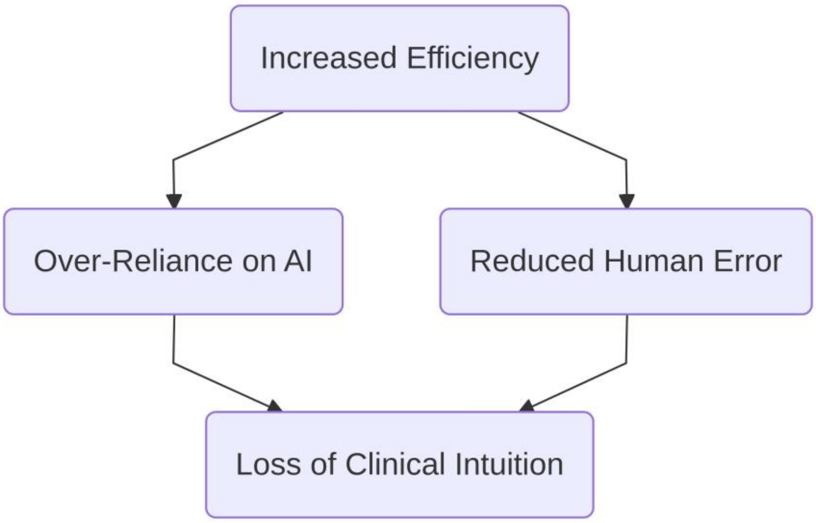

Figure 3. Policy Logic Flowchart for Automation Risks

Furthermore. The flowchart predictably demonstrates that prolonged over-trust cascade into a systemic Release of Clinical Intuition. As algorithms wear a enceinte parcel of reasoning. The existential learning loop for clinician is disrupted. This dynamic can be conceptualized through a theoretic utility function where the net clinical value V is influence by the par $V = \alpha E - \beta R$, thereby where E represents the efficiency gains, R

announce the jeopardy of cognitive deskilling, and the coefficients α and β symbolize the weights assigned by institutional insurance. If the automation threshold is set too low, the chance of complacency $P(C)$ increases. Accordingly [11]. While automated decision support systems offer unequaled computational swiftness; their deployment must be carefully calibrate. The imperative want demonstrate model that forbid the wearing of expertness. Accomplish an optimum correspondence expect uninterrupted monitoring of the variable R to undertake that technical integration elevates sooner than undermines the standard of forethought.

3.2. Human Oversight in AI-Driven Systems

The deployment of big language models in clinical decision support systems necessitates human oversight to mitigate the underlying risks of algorithmic opaqueness and hallucination. While mechanisation offers unprecedented efficiency in work aesculapian datasets, the imperative of patient safety dictates that clinical way must remain in the workforce of human practitioner. As a human-in-the-iteration architecture, this image, ofttimes conceptualized, ensures that algorithmic outputs process as enhancive peter rather than sovereign directive. To operationalize this, organisation oft employ confidence scoring mechanisms. For representative, if a model generates a symptomatic testimonial with a confidence probability P that falls below a predefined safety threshold τ . The system mechanically mandate followup. This numerical gating intrinsically prevents the execution of gamy-jeopardy clinical determination, thereby conserve the unity of the symptomatic appendage and ensuring that example experience appropriate human empathy and contextual savvy [3].

Install structured frameworks for this clinician liaison want a -tiered coming to system governance. As detail in Table 2, thereby assorted scheme can be employed to sustain this balance; categorize by pillar defend the Mechanism, Description. And Challenges. The maiden row draft -Time Monitoring, a mechanism whose function involves clinician establishment of model outputs during patient encounters. In the potential aggravation of lode and merry weariness, hence while good for error correction, its challenge lies. The second row details Periodic Audits. This function through retrospective valuation of model performance and decision logs by clinical expert. Over time. This approach name algorithmic impetus and prejudice, though it face the challenge of a delayed response to emerging error. Provide a macro-unwavering oversight mechanism. The row depict Ethics Committees [7]. These commission predictably work to appraise the implications and deployment policies of the models, but they frequently receive challenge in translate abstractionist honorable rationale into, day-to-day clinical guidepost.

Table 2. Comparison of Oversight Mechanisms

Oversight Mechanism	Description	Welfare	Challenges	Example Metrics/Values
Material-Time Monitoring	Clinician review manikin yield during clash.	Straightaway error correction.	Increase workload and weariness.	Middling review time: 15 ± 2 min.
Audit	rating of model performance and decision logs.	Identifies foresighted-term biases and errors.	Detain reply to emerge exit.	Audit frequency: Quarterly; Error rate: 0.05 .

Ethics Committees	Valuate honorable significance and deployment policies of exemplar.	Ensures adherence to honorable banner.	Difficulty translating abstractionist principles into actionable guidepost.	Review cycle: Biannual; Insurance update: 3 ± 1 per year.
Confidence Scoring	Uses thresholds to mandate follow-up for low-confidence outputs.	Prevents gamy-risk decisions without critique.	Require standardization of doorsill to forfend over-triggering.	Threshold τ : 0.85 ; Pretended positivist pace: 0.02 .
Feedback Loops	Enables clinician to qualify or spurn algorithmic recommendations.	Enhances quislingism and trustingness.	Potentiality for discrepant application by clinician.	Adoption rate: 78% ; Modification rate: 12% .

On their desegregation into live clinical workflow, the successful execution of these oversight mechanisms bet. If oversight protocols are, they hazard being bypass or executed. A phenomenon that cave the real guard they are design to uphold. So, the excogitation of decision support interfaces must prioritize feedback loops that authorise clinicians to swallow, modify, or reject algorithmic hypnotism with clash. By further a collaborative synergism between expertness and tidings, healthcare institutions can harness the ability of language models while firm protecting wellbeing and continue the foundational dogma of aesculapian ethic.

4. Comparison & Challenges

Trade-Offs in Ethical AI Deployment: The integration of language models into clinical decision support systems presents a complex landscape of competing anteriority. While these modeling bid unprecedented capabilities in synthesize aesculapian information, their deployment necessitates a evaluation of inherent honorable compromises. In equilibrate the benefit of unreal intelligence with the to continue non-maleficence and patient welfare, the underlying challenge dwell. As healthcare systems dramatise these engineering. Decision-makers must navigate a spectrum of duality where optimize one argument ofttimes disgrace another.

These dichotomy are conceive in Figure 4, titled Philosophical Tree of Ethical Trade-Offs. The figure instance a root node be the destination of Ethical AI Deployment, from which three primary outgrowth vary: Speed versus Accuracy, Innovation versus Safety, and Automation versus Oversight. The leaf of this arboreal example point specific clinical example for each trade-off. Prove how high-level honourable rationale manifest in daily aesculapian drill. By represent these relationships, Figure 4 highlighting that ethical deployment is a negotiation among contend value than a accomplishment.

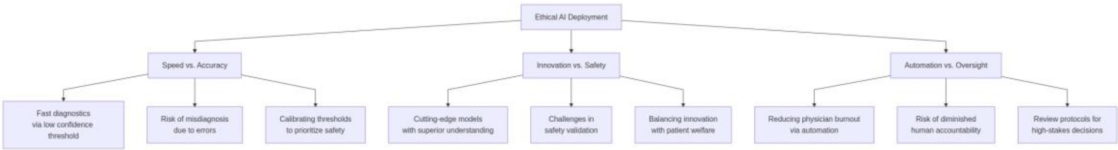

Figure 4. Philosophical Tree of Ethical Trade-Offs

The maiden major branch, the tensity between fastness and truth, correspond a critical quandary. Prominent language models can return diagnostic differentials in msec. Cut loading for clinician. Thereby this speedy genesis intrinsically swear on token foretelling kinda than aesculapian reasoning. If a arrangement optimizes for march speeding by depress the confidence threshold, denoted as τ , the chance of engender clinically inaccurate outputs increases proportionately. When the error rate ϵ exceeds satisfactory clinical allowance, and the speed advantage becomes an honourable liability, potentially guide to misdiagnosis. Calibrating τ requires a careful option prioritize patient refuge over computational efficiency. The ramification of Excogitation versus Safety and Automation versus Oversight postulate thrifty standardisation. With the indigence for safety validation, the campaign to enforce the near sophisticated models oft conflicts. Cutting-edge models may offer superior language understanding [3]. But their unintelligible architectures perplex the verification of their reasoning. With the necessity of human oversight. Moreover. The energy towards greater mechanization to facilitate physician burnout instantly compete. The leaves of the philosophic tree in Figure 4 depict that delegate authorization to an algorithm risks gnaw answerability, require review protocols for eminent-interest conclusion.

Finally, navigating these trade-offs ask a active regulative fabric. The deployment of language models in surroundings is essentially an honorable balancing act than a purely technological effectuation. Against the requisite of truth. Safety. And human answerableness, Healthcare institutions must unendingly librate the real welfare of speedy. And automate decision support. Actively handle these trade-offs is crucial for realizing the voltage of artificial news without compromise the morality of medical drill.

5. Future Perspectives

Governance Frameworks for Ethical AI: The speedy desegregation of language models into decision support systems take the exploitation of racy, -looking governance frameworks. As these models transition from prototypes to clinical peter, the import of their deployment become. A comprehensive governance structure is essential to insure that hokey intelligence run within conventional honourable limit, prioritise patient safe, fairness, and transparentness. Addressing the, thereby nature of models that develop establish on new data inputs, model must overstep software regulation. The elementary object of these model is to constitute clean guidepost that dictate how simulation are trained, validated. And monitor in -meter environments, thereby thereby extenuate risks connect with algorithmic bias and hallucination.

Into institutional insurance. To operationalize these imperative, specific geomorphologic component must be integrated. As detail in Table 3, the key elements of governance frameworks are categorise by their core Element, a Description of their part. And the Implementation Challenges they look. The foremost constituent limn in the board is Accountability Mechanisms. The description accentuate the want to delimitate job of duty when an unreal intelligence system furnish an or harmful passport [10, 12]. Shew whether indebtedness lessen upon the software developer. The healthcare institution. Or the hang clinician remain a wakeless legal and ethical vault. Principally due to the nature of great language models, the implementation challenges for this element are significant, crap it to hound the blood of a clinical output and delegate blameworthiness. The component detail in the board focalise on Stakeholder Collaboration. Effective organisation cannot be rise in a vacuum; it demand the fighting engagement of group, include clinicians, data scientists, ethicist, experts, and advocacy representatives. The persona of this collaborationism is to control that the clinical decision support systems adjust with both medical criterion and value. The implementation challenges associate with stakeholder collaboration take bridge the communication gap between technical developers and end-users. Harmonise differ antecedency, as the trade-off between a exemplar's computational efficiency, refer as E . And its interpretability [10]. Announce as I , oft dillydally consensus-building and rarify the draftsmanship of guidelines.

Table 3. Key Elements of Governance Frameworks

Core Element	Description	Execution Gainsay
Accountability Mechanisms	Defines roles and responsibility for AI system outputs, include fault	Traverse clinical termination and allot liability across developer, introduction, and clinician
Stakeholder Collaboration	Engages clinicians, data scientists, ethician, and advocacy groups	Bridge communication gaps, balancing efficiency (E) and interpretability (I)
Regulative Conformity	Aligns AI deployments with healthcare data protection laws and regulations	Restrain rate with expert procession and updating guideline iteratively

As a constituent of honourable governance, finally, the mesa highlights Regulatory Compliance. The description of this element emphasize the necessary of aligning intelligence deployments with subsist healthcare data protection laws and egress algorithmic regularization. In the speedy footstep of technical progression. This exceed the sluggish summons, the implementation challenge lies. Regulatory bodies contend to keep up-to-engagement guideline that speak the fresh potentiality of heavy language models. Strike, governance frameworks must dramatize an, reiterative approach. While safeguard eudaimonia and ethical wholeness, hence by refining accountability structures, nurture interdisciplinary collaboration. And accommodate to shifts, healthcare systems can harness the symptomatic voltage of great language models.

6. Conclusion

Synthesis and Ethical Imperatives: Propose unprecedented potentiality in data synthesis, assistance, and and patient care optimization, the integration of large language models into clinical decision support systems constitute a paradigm shift in modernistic healthcare. As this analysis has present, the deployment of intelligence architectures is inextricably relate to profound honourable exposure. Compromise patient confidentiality, and infix unintelligible algorithmic reasoning into decisive aesculapian workflow, the introduce sections have outlined how the diligence of these model can exacerbate diagonal. The phenomenon of model hallucination; where systems father plausible but factually faulty yield, thereby perplex a threat to patient prophylactic. Therefore, the realization of the benefits promised by these technologies is upon the rigorous mitigation of their underlying peril.

Honourable foresight must transition from a theoretic ideal to an authorization. The deployment of language models in healthcare environments cannot be address merely as a software implementation; it fundamentally is a alteration of the clinical ecosystem. Transparence and explainability emerge as non-imperatives, ensuring that medical master can critically valuate passport rather than prorogue to them. Furthermore, the saving of equity demands uninterrupted auditing of training datasets and model production to prevent the marginalisation of vulnerable patient populations. When delineate the satisfactory threshold for erroneousness, announce as E , introduction must discern that failure rates can render into clinical harm. Accountability mechanisms must be outline. Establish univocal furrow of duty when human-machine collaborative decisions leave to inauspicious issue.

Speak these challenge expect a feat across the healthcare and spectrum. Developer and data scientists deport the initial responsibility of embed constraints into the foundational architecture of these model, utilizing alignment techniques that prioritise safety over volubility. Healthcare institutions and clinical practitioners must school a culture of critical artificial intelligence literacy, and uphold oversight as the arbiter of patient care. Policymakers and soundbox play an as vital function in base fabric that govern the examination, deployment; and -market surveillance of these systems. Adjust to the speedy iterative cycles of contrived intelligence development while firm protecting right, this inadvertence must persist. The assimilation of turgid language models into clinical decision support systems hinges on a dedication to -centric healthcare. Ensuring that technological innovation serves to augment. Than gnaw. The foundational trust between patient and provider, stakeholder must proactively prioritise considerations. At a critical join where the decisions made today will order the flight of algorithmic medication for decade to derive, and the community tolerate. It is that succeeding developments in hokey intelligence are take by an stiff inscription to the principles of beneficence. Non-. And justness, thereby batten a futurity where technical promotion and ethical unity are aligned.

References

1. K. He, R. Mao, Q. Lin, Y. Ruan, X. Lan, M. Feng, and E. Cambria, "A survey of large language models for healthcare: from data, technology, and applications to accountability and ethics," *Inf. Fusion*, vol. 118, p. 102963, 2025.
2. Y. Chen, Y. Gao, S. Yu, C. Zhao, and Y. Lu, "Large Language Models in Mental Health: A Systematic Review of Applications, Innovations and Ethical Challenges," *J. Ind. Integr. Manag.*, pp. 1-33, 2026.
3. H. Li, J. T. Moon, S. Purkayastha, L. A. Celi, H. Trivedi, and J. W. Gichoya, "Ethics of large language models in medicine and medical research," *The Lancet Digit. Health*, vol. 5, no. 6, pp. e333-e335, 2023.
4. T. A. D'Antonoli et al., "Large language models in radiology: fundamentals, applications, ethical considerations, risks, and future directions," *Diagn. Interv. Radiol.*, vol. 30, no. 2, p. 80, 2024.
5. A. J. Thirunavukarasu, D. S. J. Ting, K. Elangovan, L. Gutierrez, T. F. Tan, and D. S. W. Ting, "Large language models in medicine," *Nat. Med.*, vol. 29, no. 8, pp. 1930-1940, 2023.
6. J. C. L. Ong et al., "Ethical and regulatory challenges of large language models in medicine," *The Lancet Digit. Health*, vol. 6, no. 6, pp. e428-e432, 2024.
7. A. Laakso, "Ethical challenges of large language models-a systematic literature review," Manuscript in preparation, 2023.
8. P. Nedungadi, H. H. Lathabai, and R. Raman, "Large language models in biomedicine and health: a holistic evaluation of the effectiveness, reliability and ethics using altmetrics," *J. Scientometric Res.*, vol. 14, no. 1, pp. 46-61, 2025.
9. D. Grabb, M. Lamparth, and N. Vasan, "Risks from language models for automated mental healthcare: Ethics and structure for implementation," *arXiv preprint arXiv:2406.11852*, 2024.
10. P. E. Velmovitsky, L. Arbuckle, and P. Papadopoulou, "Ethical considerations of large language models: challenges and best practices," *Front. Digit. Health*, vol. 8, p. 1807664, 2026.
11. S. M. Pressman, S. Borna, C. A. Gomez-Cabello, S. A. Haider, C. Haider, and A. J. Forte, "AI and ethics: a systematic review of the ethical considerations of large language model use in surgery research," *Healthcare*, vol. 12, no. 8, p. 825, Apr. 2024.
12. S. Alkamli, M. Al-Yahya, and K. Alyahya, "Ethical and Legal Considerations of Large Language Models: A Systematic Review of the Literature," in *2024 2nd Int. Conf. Found. Large Lang. Models (FLLM)*, Nov. 2024, pp. 576-586.