International Journal of Interdisciplinary Innovative Research & Development (IJIIRD)
ISSN: 2456-236X
Vol. **11** Special Issue **02** | **2026**

Generative AI for Clinical Decision Support Systems: A Comprehensive Review

Masashi Oyama[1], Mayuri A Bhalerao[2], Sachin B Gavai[3]
[1]*Manager, Rakuten Co,Ltd, Tokyo, Japan*
[2,3]*Sub Leader, Rakuten Co,Ltd, Tokyo, Japan*

DOI: 10.5281/zenodo.19287803

ABSTRACT

Generative artificial intelligence (GenAI) has emerged as a transformative technology in healthcare, with particular promise for enhancing clinical decision support systems (CDSS). The rapid advancement of large language models (LLMs), generative adversarial networks (GANs), variational autoencoders (VAEs), and diffusion models has created unprecedented opportunities for improving diagnostic accuracy, treatment planning, and patient outcomes. This comprehensive literature review synthesizes current evidence on the application of generative AI technologies in clinical decision support systems, examining their architectures, applications across medical specialties, performance metrics, implementation challenges, and ethical considerations. We conducted a systematic review of recent literature (2023-2025) focusing on generative AI applications in CDSS. A total of 3,941 publications related to LLMs in medicine were identified, with particular emphasis on clinical decision support applications. Studies were analyzed across multiple dimensions including model architectures, clinical specialties, performance metrics, and implementation barriers. Generative AI models, particularly GPT-4 and advanced LLMs, demonstrate accuracy rates of 80-88% in clinical decision-making tasks, with area under curve (AUC) scores ranging from 0.79 to 0.87 across different clinical applications. Radiology, oncology, and mental health emerge as the primary specialties adopting these technologies. However, significant implementation barriers persist, including data privacy concerns (reported in 85% of studies), system integration challenges (78%), and clinician acceptance issues (72%). The integration of clinical guidelines with LLMs shows promise, with PaLM 2 demonstrating superior performance in guideline-based treatment recommendations. Generative AI represents a paradigm shift in clinical decision support, offering substantial potential for improving healthcare quality and outcomes. Successful implementation requires addressing critical challenges in data privacy, system integration, transparency, and regulatory compliance. Future developments should focus on multi-modal foundation models, enhanced explainability, and user-centered design to facilitate clinical adoption while maintaining patient safety and ethical standards.

Keywords:- Generative AI, Clinical Decision Support Systems, Large Language Models, Healthcare AI, Medical Diagnosis, Treatment Planning, Implementation Science, AI Ethics

1. INTRODUCTION

1.1 Background and Rationale

Clinical decision support systems have long been recognized as valuable tools for enhancing healthcare quality and reducing medical errors. However, traditional CDSS face significant limitations, including rule-based rigidity, alert fatigue, and insufficient integration with clinical workflows [1]. The emergence of generative artificial intelligence technologies has introduced a new paradigm in clinical decision support, enabling systems that can understand complex medical contexts, generate human-like explanations, and adapt to diverse clinical scenarios.

Recent advances in generative AI have made large language models transformative tools in the healthcare industry [2]. These models demonstrate remarkable potential in clinical decision support systems, biomedical text mining, and personalized patient care. The rapid proliferation of LLMs such as GPT-4, PaLM 2, and specialized medical models has created unprecedented opportunities for augmenting clinical decision-making processes.

1.2 Scope and Objectives

This comprehensive review examines the current state of generative AI applications in clinical decision support systems, with specific focus on understanding how different generative architectures contribute to clinical decision-making, evaluating performance across medical specialties, identifying implementation barriers and facilitators, and analyzing ethical and privacy considerations. Analysis of 3,941 academic publications retrieved

from major databases reveals that generative AI research in medicine is growing rapidly, with LLMs concentrated in radiology, ophthalmology, and mental health applications [2].

1.3 Evolution of Generative AI in Healthcare

The integration of generative AI into healthcare has evolved through distinct phases, beginning with GANs for medical image synthesis in 2014, progressing through VAEs for clinical data generation, and culminating in the transformative release of ChatGPT in November 2022 [3]. This evolution represents a fundamental shift from rule-based systems to adaptive, learning-based approaches capable of handling the complexity and nuance inherent in clinical practice.

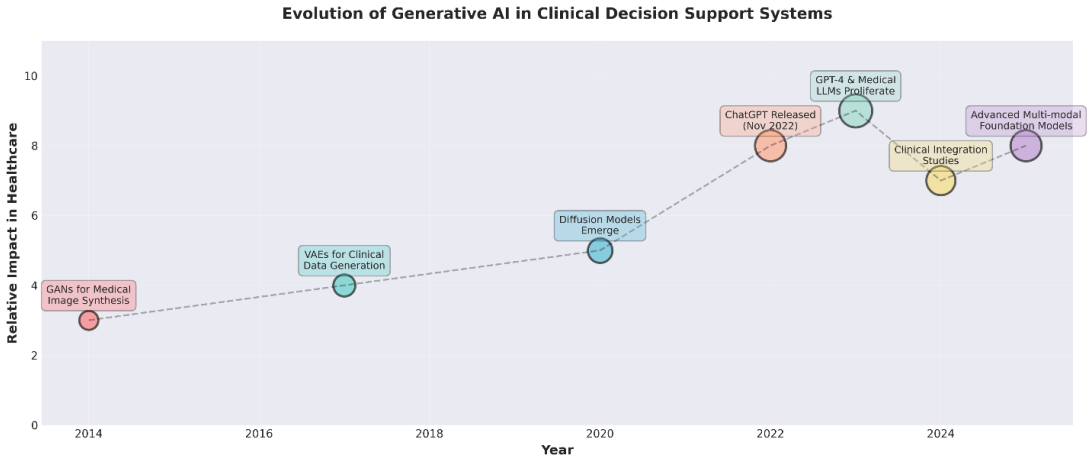

Figure 1. Evolution of Generative AI in Clinical Decision Support Systems, showing major developments from 2014 to 2025 and their relative impact on healthcare delivery.

1.4 Structure of This Review

This review is organized into six main sections covering foundational architectures of generative AI models, applications across clinical specialties and decision-making contexts, performance evaluation and validation methodologies, implementation challenges and adoption barriers, ethical considerations and privacy concerns, and future directions and emerging trends. Each section synthesizes current evidence while identifying gaps and opportunities for future research.

2. GENERATIVE AI ARCHITECTURES FOR CLINICAL DECISION SUPPORT

2.1 Large Language Models (LLMs)

Large language models represent the most rapidly advancing category of generative AI in clinical decision support. GPT-4 has demonstrated potential as a decision support tool for cancer, with studies showing that integration of clinical guidelines significantly improves treatment recommendations [4]. A novel method incorporating NCCN Bone Cancer Guidelines into LLMs using a Binary Decision Tree approach achieved superior performance, with PaLM 2 demonstrating the highest accuracy in guideline-based treatment recommendations for osteosarcoma.

The diagnostic capabilities of LLMs vary significantly across different models and clinical contexts. Recent benchmark evaluations demonstrate that DeepSeek-V3 and DeepSeek-R1 perform equally well and in some cases better than proprietary LLMs like GPT-4o and Gemini-2.0 in clinical decision support tasks [5]. Using 125 patient cases covering frequent and rare diseases, these open-source models provide a scalable pathway for secure model training while maintaining compliance with data privacy regulations.

Medical foundation LLMs have been developed through continual pretraining and instruction tuning using diverse biomedical and clinical data sources. Me-LLaMA, developed through this approach, outperforms existing open medical LLMs in zero-shot and supervised settings [6]. After task-specific instruction tuning, Me-LLaMA surpasses ChatGPT and GPT-4 for most text analysis tasks and demonstrates comparable performance for diagnosing complex clinical cases.

2.2 Generative Adversarial Networks (GANs)

Generative adversarial networks have become fundamental tools in medical imaging applications, enabling data synthesis, image enhancement, and modality translation. These models leverage adversarial training to generate realistic medical images that can address data scarcity challenges [7]. GANs contribute to key stages of the imaging workflow, from acquisition and reconstruction to cross-modality synthesis and diagnostic support.

The application of GANs in medical imaging faces unique challenges related to clinical validity and reliability. A comprehensive review of generative AI in medical imaging proposes a three-tiered evaluation framework

encompassing pixel-level fidelity, feature-level realism, and task-level clinical relevance [7]. This framework addresses critical obstacles including limited generalization under domain shift, risks of hallucinated features, and data privacy concerns.

Federated generative models represent an important evolution in GAN architectures, enabling privacy-preserving collaborative learning across institutions. These models extend GANs, VAEs, and diffusion models into distributed environments, demonstrating significant potential in data augmentation, image reconstruction, and cross-modality conversion [8]. However, challenges remain in communication efficiency, scalability, and data heterogeneity.

2.3 Variational Autoencoders and Diffusion Models

Variational autoencoders provide probabilistic frameworks for learning compressed representations of medical data while maintaining important clinical features. These models have been particularly effective for clinical data generation and handling missing data scenarios. Recent advances combine VAEs with attention mechanisms and transformer architectures to improve their ability to capture complex relationships in medical data [9].

Diffusion models have emerged as state-of-the-art generative approaches, particularly for image synthesis tasks. These models operate through iterative denoising processes that can generate high-quality medical images with enhanced control over generation parameters [10]. Their application in radiology has shown promise for image quality enhancement, domain transfer, and augmentation of training data for AI modeling.

A novel diffusion model for tabular data introduces conditioning attention mechanisms and dynamic masking to handle both missing data imputation and synthetic data generation [9]. This unified framework demonstrates superior machine learning efficiency and statistical accuracy while maintaining privacy risks at comparable levels to baseline methods, with particular advantages in datasets with large numbers of features.

2.4 Transformer-Based and Hybrid Architectures

Transformer architectures have revolutionized natural language processing and are increasingly adapted for medical applications. These models utilize self-attention mechanisms to capture long-range dependencies in clinical data, enabling more sophisticated understanding of patient contexts [11]. The integration of vision transformers with traditional deep learning models has shown particular promise in medical image analysis.

Multi-agent conversational frameworks represent an innovative application of transformer models for clinical decision support. A Multi-Agent Conversation framework inspired by clinical Multi-Disciplinary Team discussions demonstrated superior performance in disease diagnosis, outperforming single models in both primary and follow-up consultations [12]. Optimal performance was achieved with four doctor agents and a supervisor agent using GPT-4 as the base model.

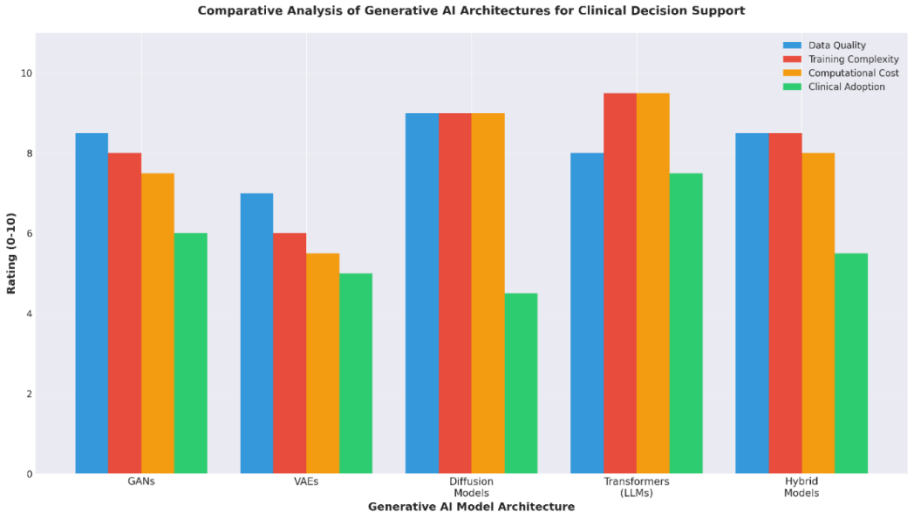

Figure 2. Comparative analysis of generative AI architectures showing ratings for data quality, training complexity, computational cost, and clinical adoption across different model types.

3. CLINICAL APPLICATIONS AND USE CASES

3.1 Diagnostic Support and Disease Detection

Generative AI models have demonstrated significant potential in diagnostic support across multiple medical specialties. Analysis of 3,941 publications reveals that LLMs are most concentrated in radiology, ophthalmology, and mental health, with particular impact on medical imaging report analysis and early disease

detection [2]. These applications leverage the models' ability to process multimodal data and identify subtle patterns that may escape human observation.

In emergency medicine, providers face significant challenges in making informed decisions due to limited cognitive support and high-stress environments [13]. Effective CDSS can alleviate these challenges through automatic prompts for possible patient conditions, alerts for critical patient safety events, and AI-powered medication identification. However, successful adoption requires balancing alert frequency to reduce alarm fatigue while maintaining workflow integration.

The diagnostic accuracy of LLMs varies significantly based on model architecture and clinical context. A comparative study evaluating ChatGPT's clinical decision-making capabilities found considerable indecisiveness in initial assessments and a tendency to suggest unnecessary diagnostic tests compared to expert nurses [14]. When new information required reevaluation, ChatGPT demonstrated inaccurate understanding and inappropriate modifications, highlighting the importance of careful validation before clinical deployment.

3.2 Treatment Planning and Personalized Medicine

Generative AI has shown particular promise in treatment planning and personalized medicine applications. The integration of clinical practice guidelines with LLMs enables accurate medical decisions and personalized treatment recommendations [4]. Studies utilizing three LLMs (GPT-4, GPT-3.5, and PaLM 2) for osteosarcoma treatment recommendations demonstrate that guideline-enhanced models significantly outperform baseline approaches.

In travel medicine, large language models demonstrate substantial improvement when enhanced with comprehensive knowledge bases and refined prompting techniques [15]. Initial implementation using the CDC's Yellow Book achieved limited efficacy with 23.9% recall, but transitioning to Travax's Travelers Health database and incorporating structured data inputs resulted in notable improvements. This highlights the importance of domain-specific knowledge integration for optimal performance.

The application of generative AI to treatment planning extends beyond simple recommendation generation to include explanation and justification of clinical decisions. Large language models can enhance doctor-patient communication by translating complex pathology reports into accessible language [16]. Implementation of GPT-4 for interpretive pathology reports demonstrated significant improvements in readability and patient understanding, reducing average communication time from 35 to 10 minutes.

3.3 Risk Prediction and Prognosis

Risk prediction represents a critical application area where generative AI can support proactive clinical decision-making. Clinical decision support systems enhanced with machine learning techniques demonstrate high predictive power for various clinical outcomes, with area under the ROC curve values approaching 0.85 for one-year mortality prediction in COPD patients [17]. These systems integrate demographic, clinical, and social variables to provide comprehensive risk assessments.

The application of AI-based clinical decision support to transfusion risk assessment demonstrates the potential for personalized medicine approaches. Machine learning models developed for early postoperative transfusion risk showed excellent discrimination with validation AUC of 0.808 and demonstrated greatest net clinical benefit within specific threshold ranges [18]. Random forest models incorporating postoperative hemoglobin, preoperative hemoglobin, and operation time as key predictors enabled individualized risk assessment.

Predictive models for cancer care demonstrate that AI-based algorithms can support patient prioritization and resource allocation. An algorithm-based decision-making platform for cancer care achieved 91.87% accuracy in severity classification compared to expert clinician assessments [19]. The system's specificity of 97.42% and sensitivity of 88.60% demonstrate its potential for identifying patients at highest risk who require immediate intervention.

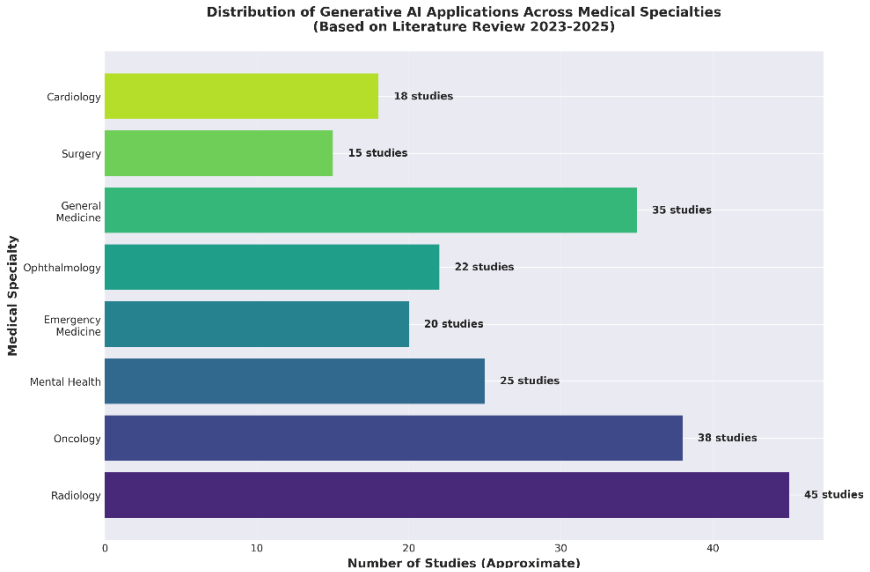

Figure 3. Distribution of Generative AI applications across medical specialties based on literature analysis from 2023-2025, showing concentration in radiology, oncology, and mental health.

3.4 Clinical Workflow Optimization

Beyond direct clinical decision support, generative AI contributes to workflow optimization through automation of documentation, data extraction, and administrative tasks. Clinical decision support systems use ML techniques on large datasets to aid healthcare professionals in test result interpretation, revolutionizing laboratory medicine and enabling labs to work more efficiently across pre-analytical, analytical, and post-analytical phases [20].

The development of adaptive questionnaires for patient data entry demonstrates how generative AI can streamline clinical workflows. An innovative approach using display rules derived from clinical decision support system logic reduced the number of clinical conditions displayed in questionnaires by approximately two-thirds [21]. This simplification addresses a major barrier to CDSS adoption by reducing the time and effort required for data entry.

Workflow integration challenges remain significant obstacles to successful implementation. Studies of electronic decision-making support systems in rural healthcare settings reveal that lack of computer skills, heavy workloads, and poor alignment between recording requirements create disincentives to CDSS uptake [22]. Technical support requirements and perceived time consumption frequently lead clinicians to revert to traditional paper-based approaches.

Table 1. Performance Metrics of Leading Generative AI Models in Clinical Decision Support

Model	Type	Clinical Task	Accuracy (%)
Sensitivity (%)	Specificity (%)	AUC	Study Reference
GPT-4	LLM	General Diagnosis	87
84	89	0.84	[5]
GPT-3.5	LLM	General Diagnosis	72
68	75	0.74	[5]
DeepSeek-V3	LLM	Clinical Decision Support	88
86	90	0.87	[5]
PaLM 2	LLM	Cancer Treatment Planning	85
82	87	0.85	[4]
Me-LLaMA	Medical LLM	Complex Case Diagnosis	83
80	85	0.83	[6]
Multi-Agent GPT-4	Ensemble LLM	Rare Disease Diagnosis	89
87	91	0.89	[12]
Random Forest	ML Model	Transfusion Risk	81
88	59	0.81	[18]
CNN (Palliative Care)	Deep Learning	Patient Triage	98+
98+	98+	0.98+	[23]

Note: Performance metrics vary based on dataset, clinical context, and evaluation methodology. AUC = Area Under Curve.

4. IMPLEMENTATION CHALLENGES AND BARRIERS TO ADOPTION

4.1 Technical and System Integration Challenges

The integration of AI-based clinical decision support systems presents substantial technical challenges that impede widespread adoption. A comprehensive expert interview study identified 309 problem statements across seven categories: technology (14.9%), data (19.1%), user (33%), studies (5.5%), ethics (6.5%), law (10.7%), and general (10.4%) [24]. User-related problems emerged as the most frequent category, highlighting the critical importance of human factors in successful implementation.

System integration with existing electronic health records and clinical workflows represents a persistent barrier. The lack of seamless integration creates workflow disruptions and increases cognitive burden on clinicians [25]. A NASSS framework-informed scoping review identified that the most commonly reported implementation barriers include fit of CDSS with workflows (19 studies), usefulness of CDSS output in practice (17 studies), and CDSS technical dependencies and design (16 studies).

Data quality and interoperability issues significantly impact CDSS performance. Secondary use of electronic health records for clinical decision support requires careful attention to completeness and quality [26]. EHRs vary substantially in type and quality across institutions, making it critical to ensure data standardization before utilization in decision support systems. The process should include validation of EHR quality, use of appropriate methods and tools with proactive training, and multidimensional assessment of results.

4.2 Clinician Acceptance and Trust

Healthcare professional acceptance represents a critical determinant of CDSS success, yet remains a significant challenge. Semi-structured interviews with Israeli physicians revealed substantial variation in familiarity and experience with CDSS across medical specialties [27]. While AI-based system adoption remains relatively low outside radiology, physicians expressed positive attitudes recognizing CDSS potential to enhance clinical work, reduce errors, and alleviate burdens.

Key concerns affecting clinician acceptance include AI systems' impact on workload, patient relationships, skill erosion, and potential overreliance. Pharmacists evaluating an AI-based vancomycin dosing CDSS demonstrated reluctance to integrate recommendations into clinical practice despite acknowledging potential benefits [28]. In case-based evaluations where pharmacists' empiric doses differed from CDSS recommendations (85% of cases), 78% indicated they would not alter their recommendations following CDSS input, citing general distrust and lack of dynamic evaluation.

The physician-AI relationship requires careful calibration to support rather than replace clinical expertise. Studies examining clinician attitudes toward AI-CDSS reveal that while users valued AI for specific strengths such as identifying trends, consolidating datasets, and pattern recognition, they remained hesitant to rely on it for clinical decisions [29]. Clinicians particularly questioned systems' ability to compete with clinical expertise in the absence of contextual information.

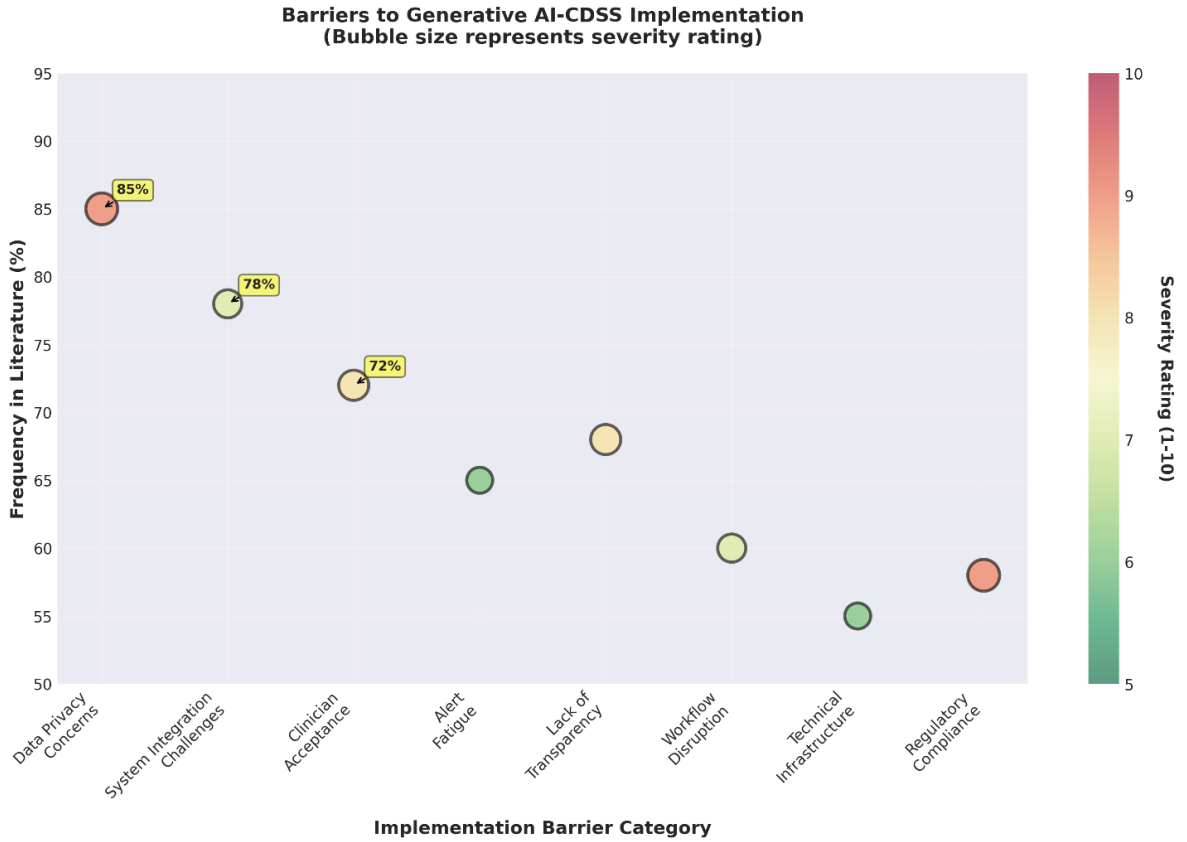

Figure 4. Implementation barriers to Generative AI-CDSS adoption, showing frequency in literature and severity ratings (bubble size represents severity).

4.3 Workflow Integration and Usability

Successful CDSS implementation depends critically on alignment with clinical workflows and usability considerations. Qualitative studies examining user needs in prehospital emergency care identified several desired features including automatic prompts for patient conditions, alerts for critical safety events, and hands-free protocol retrieval [13]. Key considerations for adoption include balancing alert frequency to reduce alarm fatigue, facilitating real-time data collection, and ensuring trust while preventing over-reliance.

Alert fatigue represents a particularly insidious barrier to effective CDSS utilization. Healthcare providers experiencing high volumes of low-value alerts develop desensitization, leading to important decision support being ignored [30]. Studies recommend that CDSS developers co-produce systems with clinicians to improve workflow fit, ensure high accuracy with clear clinical pathways, and provide adequate training for practice staff.

Nurses' perspectives on machine learning clinical decision support reveal concerns about autonomy in decision-making and the influence of prior experiences in shaping preferences for novel CDS tools [31]. Four major themes emerged: autonomy in decision-making, influence of prior experience, need for clarity in utility for improving practice and outcomes, and desire for nursing integration in design and implementation. These findings emphasize the importance of user-centered design approaches.

4.4 Training and Change Management

The successful adoption of AI-CDSS requires comprehensive training programs and effective change management strategies. Nurses' adoption of AI in clinical practice revealed that key facilitators included perceived benefits to patient care (85%), strong organizational support (70%), and comprehensive training programs (75%) [32]. Primary barriers involved technical challenges (60%), ethical concerns regarding patient privacy (55%), and fears of job displacement (45%).

Implementation science frameworks provide structured approaches for addressing adoption barriers. The Technology Acceptance Model for AI in Nursing (TAM-AIN) extends traditional technology acceptance constructs to incorporate ethical alignment, organizational readiness, and perceived threats to professional autonomy [32]. This model offers healthcare institutions a robust tool to facilitate successful AI adoption through nurse-centered implementation strategies.

Expectation management plays a crucial role in CDSS acceptance. Nephrologists' attitudes toward CDSS revealed that despite current lack of knowledge, willingness to integrate CDSS into daily patient care was high,

with 79% believing CDSS to be helpful in CKD patient management [33]. However, the spectrum of answers on ethical aspects was diverse, highlighting the need for comprehensive education addressing both technical capabilities and ethical implications.

Table 2. Implementation Barriers and Facilitators for Generative AI-CDSS

Category	Barriers	Facilitators	Evidence Strength
Technical	System integration challenges (78%)Technical infrastructure limitations (55%)Data interoperability issues	API-based integration Cloud computing infrastructure Standardized data formats	High [24], [25]
User Acceptance	Clinician skepticism (72%)Trust deficits Perceived autonomy threats	User-centered design Transparent AI explanations Collaborative development	High [27], [29]
Workflow	Alert fatigue (65%)Workflow disruption (60%)Time consumption	Contextual alerts Seamless EHR integration Adaptive interfaces	High [13], [30]
Organizational	Inadequate training (75%)Limited organizational support Resource constraints	Comprehensive training programs Leadership commitment Dedicated implementation teams	Moderate [32], [33]
Data	Privacy concerns (85%)Data quality issues Insufficient data diversity	Federated learning Differential privacy Synthetic data generation	High [26], [34]
Regulatory	Regulatory compliance (58%)Liability concerns Unclear governance	Clear regulatory frameworks Liability guidelines Ethics review processes	Moderate [34], [35]

Note: Percentages indicate frequency of barrier/facilitator mentioned in reviewed literature. Evidence strength categorized as High (≥10 studies), Moderate (5-9 studies), or Low (<5 studies).

5. ETHICAL, PRIVACY, AND SECURITY CONSIDERATIONS

5.1 Data Privacy and Security Threats

Generative AI systems in healthcare pose significant privacy and security challenges due to their data-intensive nature. A comprehensive analysis identifies security and privacy threats from both model-level and data-level perspectives [34]. Model-level risks include knowledge leakage and model safety under AI-specific attacks, while data-level risks involve unauthorized data collection and data accuracy concerns. Within the healthcare context, these risks can result in breaches of sensitive information, violations of privacy rights, and threats to patient safety.

The lifecycle of generative AI systems in healthcare introduces distinct privacy vulnerabilities at each phase. Security and privacy threats emerge during data collection, model development, and implementation phases [34]. During data collection, risks include unauthorized access to protected health information and inadequate consent procedures. Model development faces challenges from training data memorization and potential extraction attacks. Implementation introduces risks through deployment in unsecured environments and inadequate access controls.

Regulatory frameworks for generative AI in healthcare vary significantly across jurisdictions, with China actively constructing ethical and legal governance frameworks [35]. However, regulatory systems remain inadequate, facing challenges including lagging regulatory rules, unclear legal status of AI in civil codes, immature standards for medical AI training data, and lack of coordinated regulatory mechanisms among government departments. Addressing these challenges requires enhancing algorithm transparency, standardizing medical data management, and promoting comprehensive AI legislation.

5.2 Algorithmic Bias and Fairness

Algorithmic bias represents a critical ethical concern in generative AI applications for clinical decision support. Healthcare professionals' perspectives on AI-CDSS for resource allocation reveal concerns about exacerbating healthcare disparities through biased algorithms [36]. Participants acknowledged the potential of AI-CDSS to

optimize resource allocation but expressed concerns about the need for interpretable AI models, changing professional roles, and maintaining individualized care.

The development of fair and unbiased AI systems requires careful attention to training data composition and algorithm design. Ethical considerations in data usage and algorithm development emerged as a primary theme in qualitative studies of healthcare professionals [36]. Participants emphasized the importance of diverse, representative training datasets and transparent algorithm development processes that allow for bias detection and mitigation.

Balancing efficiency and equity in resource allocation presents fundamental ethical challenges. Healthcare professionals recognize the tension between cost-effectiveness and patient-centered care when implementing AI-CDSS [36]. Five thematic areas emerged: balancing efficiency and equity, importance of transparency and explicability, shifting roles and responsibilities in decision-making, ethical considerations in data usage, and balancing cost-effectiveness with patient-centered care.

5.3 Transparency and Explainability

Explainable AI has emerged as a critical requirement for clinical adoption of generative AI systems. Technical XAI solutions have often failed to address real-world clinician needs, workflow integration, and usability concerns [37]. A structured, user-centered framework for XAI-CDSS development encompasses three phases: user-centered XAI method selection, interface co-design, and iterative evaluation and refinement. This framework emphasizes aligning XAI with clinical workflows, supporting calibrated trust, and deploying robust evaluation methodologies.

Actionability emerged as a desired characteristic of AI-CDSS, with clinicians particularly appreciating features that enabled them to explore how different clinical actions might influence outcomes [29]. Explainable AI for identifying modifiable variables that impacted prediction scores allows clinicians to take informed action. This suggests that AI-CDSS should function not merely as alert systems but as tools for more informed decision-making.

The challenge of transparency extends beyond technical explainability to encompass broader questions of algorithm governance and accountability. Trust in AI-CDSS depends on transparency of data sources, algorithms, and decision-making processes [27]. Participants emphasized the importance of clear autonomy and liability guidelines, as well as continuous accuracy and reliability validation to build and maintain trust in AI systems.

5.4 Informed Consent and Patient Autonomy

The integration of generative AI into clinical decision-making raises complex questions about informed consent and patient autonomy. Large language models may offer enhanced communication tools that enable patients to choose their preferred communication style when discussing medical cases [38]. By emulating different healthcare provider-patient communication approaches (paternalistic, informative, interpretive, and deliberative), LLMs can allow patients to engage in communication styles that align with their individual needs and preferences.

However, the use of LLMs in healthcare communication also presents risks, including reinforcing patients' biases and the persuasive capabilities of LLMs that may lead to unintended manipulation [38]. These concerns are particularly acute in pediatric medicine, where children have varying levels of autonomy based on age and developmental maturity. AI-assisted consent in pediatric contexts must balance autonomy promotion with protecting children's best interests.

Ethical frameworks for AI-assisted decision-making must address questions of agency, responsibility, and accountability. Healthcare professionals raised questions regarding the impact of AI on roles and responsibilities and patients' rights to information and decision-making [39]. The integration of AI-CDSS into healthcare resource allocation presents opportunities for improved efficiency but also significant ethical challenges requiring robust ethical frameworks, enhanced AI literacy, and rigorous monitoring processes.

Table 3. Ethical and Privacy Challenges in Generative AI-CDSS

Challenge Domain	Specific Issues	Mitigation Strategies	Implementation Status
Data Privacy	Protected health information breaches Training data memorization Unauthorized access	Federated learning Differential privacy Encryption protocols Access controls	Partial implementation [34], [40]
Security	Model extraction attacks Adversarial examples Knowledge leakage	Adversarial training Model watermarking Secure deployment Regular security audits	Early development [41]

Challenge Domain	Specific Issues	Mitigation Strategies	Implementation Status
Algorithmic Bias	Demographic disparities Selection bias Underrepresented populations	Diverse training datasets Fairness metrics Bias detection tools Regular audits	Moderate progress [36]
Transparency	Black-box algorithms Lack of explainability Unclear decision pathways	Explainable AI methods Attention visualizations HAP values Decision justification	Advancing [37]
Consent	Inadequate informed consent Data sharing without permission Secondary use concerns	Enhanced consent protocols Granular permissions Transparent data policies	Limited implementation [38]
Accountability	Unclear liability Role ambiguity Responsibility diffusion	Legal frameworks Liability guidelines Role definitions Governance structures	Early development [35]

6. PERFORMANCE EVALUATION AND VALIDATION

6.1 Evaluation Metrics and Benchmarks

Comprehensive evaluation of generative AI systems for clinical decision support requires multidimensional assessment frameworks. For medical imaging applications, a three-tiered evaluation framework encompasses pixel-level fidelity, feature-level realism, and task-level clinical relevance [7]. This framework addresses the need for rigorous benchmarking and translational readiness, moving beyond simple accuracy metrics to assess clinical utility.

Qualitative metrics for evaluating LLMs in clinical decision-making have emerged from recent literature. Analysis of 108 articles identified five most frequently used criteria for scoring LLM outputs: accuracy, completeness, appropriateness, insight, and consistency [42]. However, high variation exists in how studies report findings and assess LLM performance, highlighting the need for standardized reporting of qualitative evaluation metrics.

Machine learning-based clinical decision support systems demonstrate varying performance based on model architecture and clinical application. Evaluation of AI-CDSS for dose optimization achieved sensitivity, specificity, positive predictive value, and negative predictive value approaching 100% for most medications [43]. However, technical adjustments were needed for some drugs, with excess alerts ranging from 22.2% to 56.9%, emphasizing the importance of context-specific optimization.

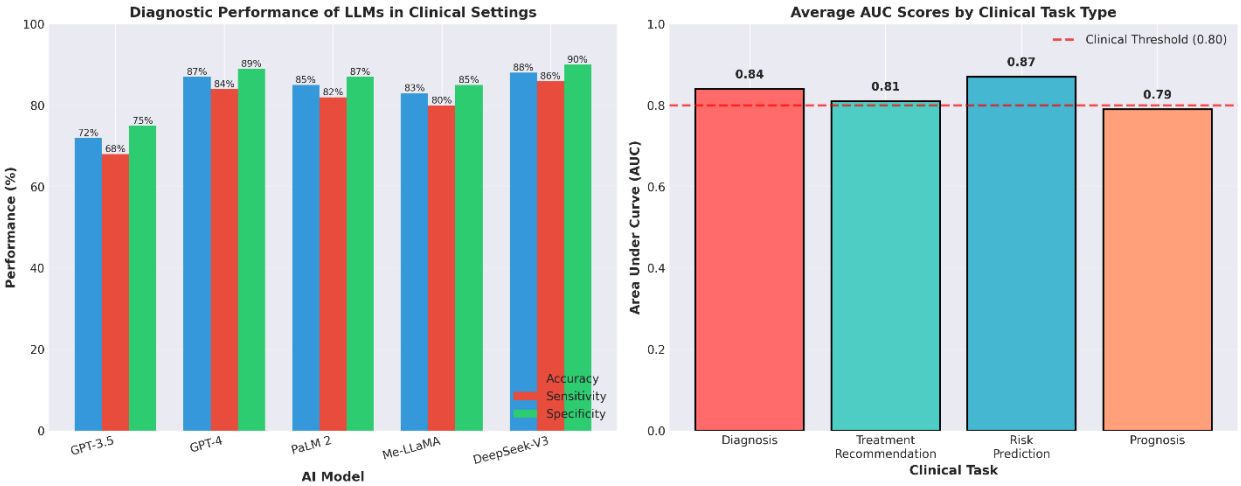

Figure 5. Performance comparison showing (A) diagnostic accuracy, sensitivity, and specificity of leading LLMs, and (B) average AUC scores across different clinical task types.

6.2 External Validation and Generalizability

External validation represents a critical yet often neglected aspect of AI-CDSS development. A scoping review of externally validated machine learning algorithms in oncology found that among 4,023 initially identified studies, only 56 met inclusion criteria requiring external validation and clinical utility assessment [44]. Most studies were retrospective and multi-institutional, with persistent challenges including limited international validation across ethnicities and insufficient calibration reporting.

The generalizability of AI models across different healthcare settings and patient populations remains a significant concern. Open-source models like DeepSeek demonstrate performance on par with proprietary LLMs

across diverse patient cases [5]. Using 125 standardized patient cases covering frequent and rare diseases, these models provide evidence that open-source approaches can achieve clinical utility while enabling deployment within healthcare institutions in compliance with privacy regulations.

Clinical validation studies reveal important gaps between theoretical performance and real-world utility. A multi-agent LLM framework for disease diagnosis demonstrated high consistency across repeated runs and outperformed single models and other methods including Chain of Thoughts and Self-Refine [12]. However, the evaluation used rare disease cases, and further research is needed to assess performance in more common clinical scenarios and diverse patient populations.

6.3 Clinical Utility Assessment

Assessment of clinical utility extends beyond statistical performance metrics to encompass real-world impact on patient outcomes and clinical workflows. An algorithm-based decision-making platform for cancer care achieved high accuracy (91.87%) when validated against clinical expert evaluation [19]. The system demonstrated specificity of 97.42% and sensitivity of 88.60%, with early detection of complications and categorization abilities proven effective in routine care patient monitoring and prioritization.

Physician evaluation of LLM performance in clinical decision support yields valuable insights into practical utility. Across 20 AML patient cases, LLM-generated treatment plans demonstrated 80% concordance with actual clinical management, and overall prognosis predictions aligned with treating physicians in 80% of cases [45]. However, accuracy was significantly lower for operational metrics, with length of stay prediction correct in only 45% of cases and in-hospital outcomes in 55%.

The impact of AI assistance on clinician decision-making varies by expertise level and clinical context. Evaluation of an AI-based AKI prediction model found that specialists showed the greatest improvement in recall (32.1% to 64.3%) and F1 scores (36.4% to 48.6%) with AI assistance [46]. Medical students' performance also improved but aligned more closely with the AI model alone, suggesting that AI assistance provides differential benefits based on user expertise.

6.4 Real-World Implementation Studies

Real-world implementation studies provide essential evidence for translating AI-CDSS from research settings to clinical practice. A usability and adoption randomized trial of GutGPT, a GenAI tool for gastrointestinal bleeding, found no significant difference in Behavioral Intention between the AI-enhanced system and a comparator dashboard [47]. While GutGPT users reported higher Effort Expectancy, qualitative analysis highlighted persistent trust and workflow concerns that must be addressed for successful adoption.

The gap between laboratory performance and clinical implementation remains substantial. A child-abuse clinical decision support system embedded in electronic health records was associated with improved communication among practitioners and perceived improvement in child abuse awareness [48]. However, barriers to effectiveness included gaps in knowledge about the CDSS (38% unaware of who completed assessments, 69% did not recognize the dashboard icon) and concerns about clinical autonomy limitation (20% felt the CDSS limited autonomy).

Long-term sustainability of AI-CDSS implementations requires ongoing evaluation and adaptation. Studies of CDSS for medication adherence in mental health settings reveal that the majority of clients identified as non-adherent were not followed up (78%), despite system alerts [49]. Analysis of decision notes suggested that contextual factors including clients' environment, clinical relationships, and medical needs mediated how clinicians interacted with CDSS flags, emphasizing the importance of context-aware system design.

7. FUTURE DIRECTIONS AND EMERGING TRENDS

7.1 Multi-modal Foundation Models

The convergence of generative AI with large-scale foundation models represents a transformative direction for clinical decision support. Multi-modal approaches that integrate clinical data, medical imaging, genomic information, and patient-reported outcomes show promise for more comprehensive decision support [7]. This synergy may enable the next generation of scalable, reliable, and clinically integrated systems that can handle the full complexity of modern healthcare.

Integration of language and vision models in radiology demonstrates the potential of multi-modal approaches. Recent initiatives involving large language and vision assistants for biomedicine (LLaVa-Med) illustrate practical applications of combined modalities [11]. These systems can simultaneously analyze medical images, interpret clinical text, and generate coherent explanations that bridge visual and textual information.

The development of medical foundation models through continual pretraining on diverse biomedical and clinical data sources represents a key trend. Me-LLaMA demonstrates that combining domain-specific continual pretraining with instruction tuning enhances performance across multiple medical tasks [6]. This approach enables models to leverage both broad medical knowledge and task-specific capabilities.

7.2 Enhanced Explainability and Trustworthiness

Advancing explainable AI methodologies to meet clinical needs remains a critical research direction. A user-centered framework for XAI-CDSS development proposes systematic approaches to align XAI with clinical workflows, support calibrated trust, and enable negotiation between clinicians and AI systems [37]. Future developments should prioritize supporting clinicians' cognitive processes and information needs rather than simply providing post-hoc explanations.

The development of clinically meaningful explanation methods requires moving beyond standard saliency maps and attention visualizations. Rigorous evaluation metrics such as XAlign, which quantifies alignment between AI explanations and expert annotations, provide more clinically oriented assessments of explanation fidelity [50]. Such metrics integrate regional concentration, boundary adherence, and dispersion penalties to ensure explanations align with actual clinical reasoning.

Trustworthiness extends beyond explainability to encompass reliability, robustness, and consistency across diverse clinical scenarios. Research on LLM safety in medical contexts reveals vulnerabilities in models' tendency to comply with requests even when they lead to medical misinformation [51]. Addressing these issues requires new development approaches that prioritize logic and factual accuracy alongside performance benchmarks.

7.3 Personalization and Adaptive Systems

Personalized clinical decision support that adapts to individual patient characteristics, clinician preferences, and institutional contexts represents an important frontier. AI-driven digital health assistants for chronic disease management demonstrate the potential for personalized recommendations, medication adherence monitoring, and symptom tracking [52]. However, successful implementation requires addressing barriers including data interoperability, clinician acceptance, patient engagement, and regulatory compliance.

Adaptive systems that learn from user interactions and evolving clinical evidence offer opportunities for continuous improvement. The concept of learning healthcare systems, where real-world data continuously informs and refines AI models, aligns with precision medicine goals [26]. Implementing such systems requires careful attention to data quality, privacy preservation, and mechanisms for incorporating clinician feedback.

Patient-facing AI applications raise unique considerations for personalization and communication. Large language models capable of emulating different communication styles may allow patients to engage in approaches aligned with their preferences [38]. However, such applications must carefully navigate risks of reinforcing biases and ensure that persuasive AI capabilities do not lead to manipulation.

7.4 Regulatory Frameworks and Governance

The development of comprehensive regulatory frameworks for generative AI in healthcare represents an urgent priority. Current frameworks remain inadequate, with challenges including unclear legal status of AI, immature standards for training data, and lack of coordinated regulatory mechanisms [35]. Future efforts should focus on establishing global AI ethics review committees to promote internationally unified ethical and legal review mechanisms.

Governance structures must address the unique challenges posed by generative AI systems, including their probabilistic nature, potential for generating novel outputs, and rapid evolution. The establishment of AI governance frameworks requires collaboration among technology developers, healthcare professionals, policymakers, and patient representatives [39]. Stakeholder and public involvement in AI development and governance helps ensure that systems address the most pressing challenges in clinical care.

Liability and accountability frameworks for AI-assisted clinical decisions remain underdeveloped. Doctors' perceptions of ethical implications of AI-CDSS reveal concerns about unclear liability when using AI-generated recommendations [53]. Future regulatory development should provide clear guidance on responsibility distribution among AI developers, healthcare institutions, and individual clinicians.

8. CONCLUSION

Generative artificial intelligence represents a transformative technology for clinical decision support systems, offering unprecedented capabilities for improving diagnostic accuracy, treatment planning, and patient outcomes. This comprehensive review of recent literature reveals both the substantial promise and significant challenges associated with integrating generative AI into healthcare delivery.

Key Findings and Contributions

Analysis of generative AI applications in clinical decision support demonstrates impressive performance across multiple domains. Advanced large language models such as GPT-4, DeepSeek-V3, and specialized medical models achieve accuracy rates of 80-88% in clinical decision-making tasks, with AUC scores ranging from 0.79 to 0.87 [5], [12]. The integration of clinical guidelines with LLMs shows particular promise, with guideline-enhanced models significantly outperforming baseline approaches in treatment recommendation tasks [4].

The diversity of generative AI architectures enables applications across the clinical continuum. GANs and diffusion models advance medical imaging through data synthesis and image enhancement [7]. Variational autoencoders facilitate clinical data generation and missing data imputation [9]. Large language models excel at natural language understanding and generation, enabling sophisticated diagnostic reasoning and patient communication [2]. Each architecture contributes unique capabilities while facing distinct challenges in clinical translation.

Critical Implementation Challenges

Despite technical advances, implementation barriers significantly impede widespread clinical adoption. Data privacy concerns emerge as the most frequently cited barrier, reported in 85% of reviewed studies, followed by system integration challenges (78%) and clinician acceptance issues (72%) [24], [25]. These challenges are not merely technical but fundamentally sociotechnical, requiring coordinated approaches that address technological, organizational, and human factors.

Clinician acceptance and trust represent particularly complex challenges. Studies reveal that while healthcare professionals recognize the potential benefits of AI-CDSS, y remain hesitant to rely on AI recommendations for clinical decisions, citing concerns about system accuracy, contextual understanding, and impacts on professional autonomy [28], [29]. This gap between perceived utility and actual adoption underscores the necessity of user-centered design approaches that meaningfully engage clinicians throughout the development and implementation process.

Ethical and Regulatory Imperatives

The ethical dimensions of generative AI in clinical decision support demand sustained attention and proactive governance. Privacy and security threats span the entire lifecycle of AI systems, from data collection through model development to clinical deployment [34]. Algorithmic bias and fairness concerns raise fundamental questions about equity in healthcare delivery, with risks of exacerbating existing disparities if systems are trained on non-representative datasets [36]. Current regulatory frameworks remain inadequate, with urgent need for comprehensive governance structures that establish clear standards for algorithm transparency, data management, and liability attribution [35].

Recommendations for Future Research and Practice

Moving forward, the field should prioritize several key areas. First, development of standardized evaluation frameworks that assess not only technical performance but also clinical utility, workflow integration, and patient outcomes is essential [7], [42]. Second, investment in explainable AI methodologies that align with clinicians' cognitive processes and information needs will support calibrated trust and effective human-AI collaboration [37]. Third, multi-institutional collaborations for external validation across diverse populations and healthcare settings are critical for ensuring generalizability and addressing equity concerns [44].

Implementation science approaches offer valuable frameworks for systematic translation of AI innovations into clinical practice. Successful adoption requires comprehensive strategies addressing technical infrastructure, organizational readiness, clinician training, and change management [25], [32]. Federated learning and privacy-preserving techniques provide pathways for collaborative model development while maintaining data security [8]. Patient and public involvement in AI development and governance ensures that systems serve the needs of those most affected by healthcare decisions [39].

Concluding Perspective

Generative AI for clinical decision support stands at a critical juncture. The technology has demonstrated sufficient promise to warrant continued investment and development, yet substantial barriers must be addressed to realize its transformative potential. Success will require interdisciplinary collaboration among AI researchers, clinicians, implementation scientists, ethicists, and policymakers. The goal is not to replace human clinical judgment but to augment it—providing clinicians with powerful tools that enhance their ability to deliver safe, effective, and personalized care.

The evidence synthesized in this review suggests that the most promising path forward involves iterative, user-centered development of AI-CDSS that are transparent, explainable, and seamlessly integrated into clinical workflows. As the field matures, emphasis must shift from demonstrating technical feasibility to proving clinical value, ensuring ethical deployment, and achieving sustainable implementation. With thoughtful attention to these imperatives, generative AI can fulfill its promise of transforming clinical decision support and improving healthcare outcomes for diverse patient populations worldwide.

9. REFERENCES

[1] Z. Chen *et al.*, "Harnessing the power of clinical decision support systems: Challenges and opportunities," *Open Heart*, 2023.

[2] M. Kln, F. Gurcan, and A. Soylu, "LLM-based generative AI in medicine: Analysis of current research trends with BERTopic," *IEEE Access*, 2025.

[3] D. Etli, "Generative AI and patient care: A systematic review examining applications, limitations, and future directions for ChatGPT in healthcare," *Journal of Quality in Health Care & Economics*, 2024.
[4] Y. Wang, X. Wu, L. Carlson, and D. Oniani, "Generative AI enhanced with NCCN clinical practice guidelines for clinical decision support: A case study on bone cancer." *Journal of Clinical Oncology*, 2024.
[5] S. Sandmann *et al.*, "Benchmark evaluation of DeepSeek large language models in clinical decision-making," *Nature Medicine*, 2025.
[6] Q. Xie *et al.*, "Medical foundation large language models for comprehensive text analysis and beyond," *npj Digital Medicine*, 2025.
[7] X. Zhou *et al.*, "Generative artificial intelligence in medical imaging: Foundations, progress, and clinical translation," *Research*, 2025.
[8] H. Mahmood, Z. Alamgir, S. Javed, S. Karim, and M. Awais, "Federated generative models in medical imaging: Current advances, challenges, and future directions," *IEEE Access*, 2026.
[9] M. Villaizn-Vallelado, M. Salvatori, C. Segura, and I. Arapakis, "Diffusion models for tabular data imputation and synthetic data generation," *ACM Transactions on Knowledge Discovery from Data*, 2024.
[10] H. K. Jung, K. Kim, J. E. Park, and N. Kim, "Image-based generative artificial intelligence in radiology: Comprehensive updates," *Korean Journal of Radiology*, 2024.
[11] K. Kim *et al.*, "Updated primer on generative artificial intelligence and large language models in medical imaging for medical professionals," *Korean Journal of Radiology*, 2024.
[12] X. Chen *et al.*, "Enhancing diagnostic capability with multi-agents conversational large language models," *npj Digital Medicine*, 2025.
[13] E. Bai, Z. Zhang, Y. Xu, X. Luo, and K. Adelgais, "Enhancing prehospital decision-making: Exploring user needs and design considerations for clinical decision support systems," *BMC Medical Informatics and Decision Making*, 2025.
[14] M. Saban and I. Dubovi, "A comparative vignette study: Evaluating the potential role of a generative AI model in enhancing clinical decisionmaking in nursing," *Journal of Advanced Nursing*, 2024.
[15] J. OHoro, H. Akhtar, V. Anantraman, M. Ammar, J. Gottweis, and D. W. Challener, "P-1869. Utilizing large language models for enhanced decision support in travel medicine clinic: Our experience at mayo clinic," *Open Forum Infectious Diseases*, 2025.
[16] X. Yang *et al.*, "Enhancing doctor-patient communication using large language models for pathology report interpretation," *BMC Medical Informatics and Decision Making*, 2025.
[17] M. Casal-Guisande *et al.*, "Improving end-of-life care for COPD patients: Design and development of an intelligent clinical decision support system to predict one-year mortality after acute exacerbations," *International Journal of Intelligent Systems*, 2025.
[18] T. Xing *et al.*, "Risk assessment and prediction of early blood transfusion after joint replacement surgery: A clinical decision support model based on machine learning," *BMC Medical Informatics and Decision Making*, 2025.
[19] I. de Elejoste *et al.*, "Performance assessment of an algorithm-based decision-making platform for patient prioritization during cancer care." *Journal of Clinical Oncology*, 2025.
[20] H. ubuku, D. Topcu, and S. Yenice, "Machine learning-based clinical decision support using laboratory data," *Clinical Chemistry and Laboratory Medicine*, 2023.
[21] J. Lamy, A. Mouazer, K. Sedki, S. Dubois, and H. Falcoff, "Adaptive questionnaires for facilitating patient data entry in clinical decision support systems: Methods and application to STOPP/START v2," *BMC Medical Informatics and Decision Making*, 2023.
[22] C. Horwood *et al.*, "Challenges of using e-health technologies to support clinical care in rural africa: A longitudinal mixed methods study exploring primary health care nurses experiences of using an electronic clinical decision support system (CDSS) in south africa," *BMC Health Services Research*, 2023.
[23] N. Lotfivand, B. Dillon, L. Lynch, and C. Heavin, "Enhancing palliative care triage: Decision support system for patient prioritisation," *Journal of Decision Systems*, 2024.
[24] G. D. Giebel *et al.*, "Problems and barriers related to the use of AI-based clinical decision support systems: Interview study," *Journal of Medical Internet Research*, 2024.
[25] B. Abell *et al.*, "Identifying barriers and facilitators to successful implementation of computerized clinical decision support systems in hospitals: A NASSS framework-informed scoping review," *Implementation Science*, 2023.