The Ethical Dimensions of AI-Based Diagnostic Systems in Modern Healthcare: A Narrative Review

Dr. Khushbu[1], Ms. Manasi[2], Ms. Neelam[3]

[1]Associate Professor ,Faculty of Nursing, SGT University, Gurugram, India khushbushona24@gmail.com
[2]PG tutor, faculty of Nursing, SGT University, Gurugram, India
[3]PG tutor, faculty of Nursing, SGT University, Gurugram, India

ABSTRACT
Artificial intelligence (AI)–based diagnostic systems are increasingly embedded in modern healthcare to support detection, triage, and interpretation across imaging and non-imaging settings. While these tools promise improved efficiency and, in some contexts, performance comparable to expert clinicians, their adoption raises complex ethical challenges that directly affect patient safety, trust, and equity. This narrative review synthesizes major ethical dimensions of AI-enabled diagnostics, treating these systems as socio-technical interventions whose impacts depend on data provenance, workflow integration, human oversight, and institutional governance. Key themes include the ethics of evidence and clinical validity, transparency and explainability aligned to stakeholder needs, and accountability across developers, healthcare organizations, and clinicians. The review highlights how bias can arise from unrepresentative datasets and problematic proxy outcomes, potentially reinforcing existing health disparities if subgroup performance and downstream impacts are not continuously evaluated. Privacy and data governance are examined as foundational ethical requirements, emphasizing lawful and legitimate secondary use of clinical data, minimization principles, and robust safeguards to prevent misuse and erosion of trust. Issues of informed consent and patient autonomy are discussed in relation to disclosure duties when AI meaningfully influences diagnostic decisions and care pathways. Human oversight is critically analyzed with attention to automation bias, emphasizing that "human-in-the-loop" must be operational and measurable rather than nominal. Finally, the review considers lifecycle governance, post-market surveillance, and procurement ethics, arguing that responsible deployment requires ongoing monitoring for performance drift, contractual auditability, and clear institutional accountability structures. Overall, ethical implementation of AI-based diagnostic systems requires rigorous context-specific validation, transparent communication of limitations, equity-centered evaluation, and continuous governance throughout the product lifecycle to ensure that innovation translates into safe, fair, and trustworthy clinical benefit....

Keywords: Artificial intelligence; Diagnostic systems; Healthcare ethics; Algorithmic bias; Data privacy..
How to cite this article: Khushbu, Manasi, Neelam. The Ethical Dimensions of AI-Based Diagnostic Systems in Modern Healthcare: A Narrative Review. Int J Drug Deliv Technol. 2026;16(63s):1647-1657. DOI: 10.25258/ijddt.16.63s.167
Source of support: Nil.
Conflict of interest: Nil.

INTRODUCTION

Artificial intelligence (AI)–based diagnostic systems are moving from research prototypes into everyday clinical environments, reshaping how diseases are detected, triaged, and monitored. These systems range from image-based algorithms that interpret radiology, dermatology, ophthalmology, and pathology studies to multimodal models that combine laboratory values, clinical notes, and physiological signals to support diagnosis. Their appeal is clear: in controlled settings, AI can match or exceed human performance for certain narrow tasks, potentially reducing delays, standardizing interpretation, and helping clinicians manage escalating volumes of data. Yet "better accuracy" alone does not settle whether a diagnostic AI system should be deployed, how it should be governed, or what duties fall on clinicians, hospitals, and developers. Diagnostic decisions are ethically charged because they can initiate invasive investigations, influence treatment eligibility, affect insurance or employment, and shape patients' understanding of their future. Consequently, the ethical stakes of AI diagnostics are not secondary concerns; they are core determinants of safety, legitimacy, and trust. A key ethical challenge is that diagnostic AI systems are socio-technical interventions rather than purely technical tools. Their real-world impact depends on training data provenance, clinical workflow integration, user behavior, and institutional incentives. A model that performs well in one hospital can degrade when applied elsewhere due to dataset shift, different disease prevalence, or divergent imaging protocols. Furthermore, clinicians may over-trust AI output (automation bias) or under-use it due to lack of interpretability or fear of liability. The result is that ethical evaluation must extend beyond algorithmic performance and include how the tool shapes responsibility, patient autonomy, and equitable outcomes. Global and national

Author for Correspondence: Dr. Khushbu

bodies have increasingly framed "trustworthy AI" in terms of principles such as safety, transparency, accountability, privacy, and equity. In healthcare, these principles become operational questions: What evidence is sufficient before using AI for diagnostic triage? What disclosures do patients deserve when AI influences clinical decisions? Who is accountable when AI contributes to harm—the clinician who used it, the institution that procured it, or the manufacturer that designed it? Guidance specific to AI in health highlights these tensions and emphasizes that humans should retain meaningful control over medical decisions, that privacy and data stewardship require heightened protections, and that fairness must be assessed in context rather than assumed from overall accuracy metrics [1]. Regulatory discussions also stress that AI-enabled software can be iteratively updated, raising lifecycle concerns that differ from static medical devices, and creating a need for continuous monitoring, documentation, and governance rather than one-time approval [2].

Ethical analysis is also complicated by the heterogeneity of AI diagnostic systems. Some are "assistive" tools that provide probability scores, heatmaps, or recommendations; others are "autonomous" or near-autonomous systems that triage, flag critical findings, or generate preliminary reads. The more an AI system substitutes for human judgment, the more stringent the ethical and governance demands become. Even when AI remains assistive, the power asymmetry between vendor and healthcare organization matters: proprietary models may limit transparency about training data, error modes, or subgroup performance. This opacity can impede informed procurement decisions, constrain auditing, and weaken patients' ability to challenge decisions. In parallel, the data foundation of diagnostic AI—often large-scale datasets assembled from electronic health records, images, and biobanks—raises questions about consent, secondary use, and representativeness. Privacy law provides a baseline, but ethical practice often requires more than legal compliance, especially when data are repurposed for model development and commercial products [3]. Finally, the field is evolving rapidly: health systems are experimenting with generative AI and large multimodal models for summarization, reasoning assistance, and diagnostic support. These tools can introduce new classes of risk—hallucinated outputs, uncertain provenance of generated text, and challenges for validation—further heightening the need for robust ethical frameworks and governance structures [4]. This narrative review examines the ethical dimensions of AI-based diagnostic systems through a thematic lens, focusing on patient safety, fairness, transparency, privacy, accountability, human oversight, and real-world governance, and synthesizes best-practice directions for responsible deployment.

Materials and Methods

This narrative review synthesized peer-reviewed literature, international guidance, and regulatory/standards documents related to ethical issues in AI-based diagnostic systems. Sources were identified through targeted searching of major medical journals and institutional repositories, prioritizing (i) widely cited empirical studies on harms (e.g., bias, automation bias), (ii) reporting and evaluation standards for AI diagnostic research, and (iii) authoritative governance documents from health and standards bodies. Included materials addressed diagnostic AI used in clinical contexts (imaging and non-imaging) and focused on ethics-relevant domains including safety, equity, transparency, privacy, accountability, and lifecycle oversight; opinion pieces were used sparingly when they clarified emerging issues and were published by reputable medical outlets. Findings were organized into ten thematic headings reflecting recurrent ethical dimensions across the literature and mapped to practical implications for clinicians, health systems, and developers.

Patient Safety, Clinical Validity, and the Ethics of Evidence

Ethically, the first obligation of any diagnostic AI system is not novelty but demonstrated clinical benefit with acceptable risk. Safety in AI diagnostics is often discussed as "accuracy," yet ethically relevant safety is broader: it includes calibration, robustness across settings, clinically meaningful error patterns, and downstream consequences of false positives and false negatives. A tool that slightly improves average accuracy can still be harmful if its errors cluster in high-risk subgroups or if it triggers cascades of unnecessary testing. Moreover, many AI diagnostic studies remain retrospective and use curated datasets that may not reflect real clinical workflows, leading to optimistic performance that fails to translate into improved outcomes. This is why reporting and evaluation standards matter ethically: they structure what evidence is visible to clinicians and decision-makers, and they reduce the risk that systems are adopted based on incomplete or biased evidence. A central ethical concern is that diagnostic AI systems may be deployed as "workflow accelerators" without adequate prospective evaluation, especially when they fit into routine image reading or triage pipelines. In such cases, harm may not appear as a single catastrophic error but as subtle shifts: delayed diagnoses in some patients, reduced clinician vigilance, or distorted resource allocation. Ethical practice therefore demands a high bar for validation aligned to the tool's intended use—screening, triage, second-read support, or autonomous decision-making—and requires that claims about performance match the context in which the system will be used. Reporting guidance for trials evaluating AI interventions underscores the need to clearly describe the AI's role, the human-AI interaction, and the clinical setting, enabling readers to judge whether benefits are plausible and transferable [5].

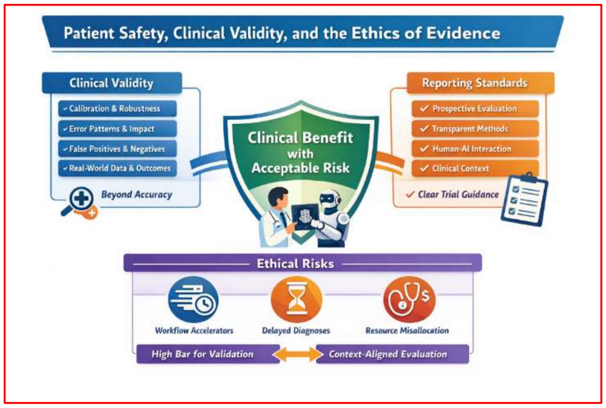

Figure : 1 Patient Safety, Clinical Validity, and the Ethics of Evidence

Transparency, Explainability, and the Right to Understand

Transparency is often invoked as a remedy for the "black box" nature of machine learning, but ethically it is better understood as a set of stakeholder-specific needs. Clinicians may need interpretability to calibrate trust and identify error modes; patients may need understandable explanations about how AI influenced their diagnosis; and institutions may need documentation to audit performance, manage risk, and meet regulatory duties. Explainability is not a single property of a model—saliency maps and feature importances can be helpful, misleading, or irrelevant depending on the clinical task. The ethical goal is not necessarily to fully "explain" a deep neural network, but to ensure that enough information is available to support safe use, accountability, and contestability. A practical ethical approach is to link transparency requirements to the risk level of the diagnostic application. High-stakes uses (e.g., cancer detection, stroke triage, sepsis alerts) warrant stronger transparency and evidence obligations than low-risk administrative support. Transparency also includes disclosing limitations: known failure modes, uncertainty estimates, data sources, and conditions under which the model should not be used. Without such disclosures, clinicians may overgeneralize AI output and inadvertently expose patients to harm. Additionally, transparency interacts with professional integrity: clinicians may feel ethically compromised if they are asked to act on outputs they cannot reasonably interrogate. Governance frameworks increasingly position transparency as part of broader "trustworthiness," emphasizing documentation, traceability, and lifecycle risk management rather than purely technical interpretability. Risk management guidance highlights the need for context-aware evaluations and clear communication of system capabilities and limitations so that users can make informed decisions about reliance on AI outputs [6].

Accountability and Liability in Human–AI Diagnostic Decisions

Accountability is ethically central because diagnostic harm can arise from complex interactions between model design, deployment decisions, clinician behavior, and institutional policies. When AI contributes to a wrong diagnosis, multiple actors may share causal responsibility: developers who chose training targets and data, institutions that procured and integrated the tool, and clinicians who relied on it. Ethical accountability requires that roles are clarified before deployment—who monitors performance, who can pause or roll back the system, who communicates AI-related risks to patients, and who investigates adverse events. Without clarity, responsibility can diffuse, leading to "accountability gaps" where harm occurs but no party feels obligated to act. Legal liability is not identical to ethical accountability, but it shapes behavior. If clinicians fear that using AI exposes them to blame, they may ignore potentially beneficial tools; if vendors disclaim responsibility entirely, incentives for rigorous monitoring and transparency weaken. A balanced ethical approach assigns responsibilities across the lifecycle: developers should provide truthful performance claims, documentation, and post-market support; institutions should ensure appropriate procurement, local validation, training, and monitoring; clinicians should apply judgment and follow policies for escalation and documentation. Accountability also extends to data practices: if a system is built on ethically questionable data acquisition or insufficient privacy protections, harm includes not only misdiagnosis but also erosion of trust. Regulatory frameworks are increasingly relevant to accountability because they create obligations for high-risk AI systems, including requirements around risk management, transparency, and oversight. In the European context, the AI Act establishes a risk-based approach to AI governance and includes obligations that directly affect healthcare deployments of high-risk systems, shaping how accountability is operationalized across stakeholders [7].

Bias, Fairness, and Health Equity

Few ethical issues in diagnostic AI are as consequential as bias and inequity. Bias can emerge from unrepresentative data, flawed outcome labels, missingness patterns, and structural inequities embedded in healthcare delivery. A model may appear accurate overall while systematically underperforming in certain racial, socioeconomic, age, or gender groups. In diagnostics, underperformance can mean missed cancers, delayed stroke treatment, or inaccurate risk stratification—harms that compound existing disparities. Importantly, "fairness" cannot be inferred from a single metric; it must be defined relative to clinical goals and justice considerations, and evaluated across relevant subgroups. An ethically instructive example is the documented racial bias in a widely used health algorithm where the system predicted healthcare costs rather than health needs, resulting in Black patients being less likely to be identified for additional care despite comparable illness burden. The core lesson is not only that bias exists, but that it can arise from seemingly "neutral" design choices such as selecting a proxy outcome that reflects unequal access or spending patterns [8]. In diagnostic AI, similar proxy problems occur when training labels reflect clinician behavior or access to confirmatory testing rather than true disease prevalence. Equity-focused governance requires more than one-time subgroup testing. Health systems should demand evidence of performance across diverse populations and settings, conduct local validation, and monitor for drift that could reintroduce inequities over time. Mitigation strategies include improved dataset representativeness, careful label selection, fairness-aware evaluation, and workflow designs that prevent the AI from becoming a gatekeeper that restricts access to diagnostic resources.

Privacy, Confidentiality, and Data Governance for Diagnostic AI

Diagnostic AI depends on data—often highly sensitive imaging, clinical notes, and longitudinal records. Ethical data governance involves more than preventing breaches; it requires legitimate purpose, minimal necessary use, secure processing, and meaningful oversight of secondary uses. In many systems, data used for model development were originally collected for clinical care, and patients may not have anticipated their information being used to train commercial AI products. Even when de-identified, imaging and genomic data can carry re-identification risks, particularly when combined with other datasets. Thus, privacy protections must be treated as core safety mechanisms, not administrative add-ons. Legal frameworks such as the GDPR emphasize principles of lawful processing, purpose limitation, data minimization, and accountability, and they encourage risk-based assessments (e.g., data protection impact assessments) when processing may pose high risks to individuals. These legal requirements align with ethical expectations that organizations should proactively evaluate and mitigate privacy risks, especially in contexts like health where harms from misuse can be severe [3]. A practical ethical approach to privacy in diagnostic AI includes robust contractual controls with vendors, strict access controls, audit logs, encryption, and clear governance on data sharing and retention. It also includes transparency to patients about data use and safeguards, and mechanisms for addressing concerns. Emerging privacy-preserving techniques—federated learning, secure enclaves, differential privacy—may reduce some risks, but they do not eliminate governance responsibilities; they change the threat model rather than solving the ethical issue of legitimacy. Ultimately, trustworthy diagnostic AI requires that patients and clinicians can reasonably believe that data are handled with respect, restraint, and accountability.

Informed Consent, Patient Autonomy, and Disclosure Duties

In diagnostic practice, patient autonomy is supported through consent, shared decision-making, and respect for patient values. AI introduces new disclosure questions: should patients be told when AI is used, how it influenced decisions, and what its limitations are? Ethically, the answer depends on materiality—whether knowledge of AI involvement could reasonably influence a patient's choices or trust. If AI meaningfully shapes diagnostic conclusions or triage priority, nondisclosure can undermine autonomy and weaken the clinician–patient relationship. Conversely, overly technical disclosures may confuse patients, suggesting that disclosure should be clear, contextual, and linked to clinical impact. Consent is also relevant upstream in data use for model development. Even when law permits certain secondary uses, ethical legitimacy may require stronger patient engagement, governance by ethics committees, and community consultation—particularly when data are used for commercial products or when benefits are unlikely to return to the populations providing data. In global health contexts, concerns about extractive data practices and inequitable benefit sharing are prominent, and ethical governance should address who benefits from AI tools developed using local patient data. Clinical autonomy also includes the clinician's professional autonomy and judgment. AI systems can create subtle pressures: institutional policies may encourage reliance on AI to reduce costs; vendors may market systems as "objective" and "superhuman"; administrators may treat AI output as a standard to which clinicians must conform. These pressures can reduce clinicians' willingness to deviate from AI recommendations, even when clinical context warrants it. Ethical guidance for AI in health emphasizes preserving human decision-making authority and ensuring that AI supports, rather than replaces, clinician responsibility and patient-centered care [1].

Human Oversight, Automation Bias, and the Ethics of Reliance

Human oversight is often presented as a simple safeguard: keep a clinician "in the loop" and ethical concerns diminish. In reality, oversight can fail if clinicians are pressured by time, lack training, or develop over-trust in AI outputs. Automation bias—the tendency to favor suggestions from automated systems and ignore contradictory information—has been documented as a risk in clinical decision support, particularly when systems appear authoritative or when users are cognitively overloaded. In diagnostics, this can manifest as clinicians accepting an AI's "normal" classification without sufficient scrutiny, or disregarding subtle signs that contradict the model's output. Ethically, the problem is not only individual cognitive bias but system design and organizational context. If a hospital deploys AI to speed up reads without adjusting staffing or workflow, clinicians may be forced into reliance patterns that make harmful errors more likely. Similarly, if AI interfaces present outputs without uncertainty information, or if they provide recommendations without rationale, clinicians may be nudged toward uncritical acceptance. The ethical responsibility to prevent automation bias is shared: designers should build interfaces that support appropriate calibration and encourage verification; institutions should train users and design workflows that preserve meaningful review; clinicians should maintain vigilance and document when AI influenced decisions. The ethics of reliance also includes questions about responsibility: if clinicians are expected to override AI, they need clear grounds for doing so without fear of punishment, and they need institutional support when they challenge AI output. A foundational discussion of automation bias in clinical decision support emphasizes that understanding these mechanisms is necessary to design safer systems and to prevent over-reliance that can degrade decision quality [9].

Research Integrity, Reporting Standards, and Reproducibility

Ethical deployment depends on trustworthy evidence, which in turn depends on research integrity. AI diagnostic studies can be difficult to interpret when authors omit key details about data selection, preprocessing, missingness, model tuning, reference standards, and subgroup performance. Without transparent reporting, clinicians and policymakers cannot judge whether results are reliable, generalizable, or relevant to real-world practice. The ethical consequence of poor reporting is not merely academic—it increases the risk that unsafe or ineffective systems will be adopted and that patients will be exposed to unrecognized harms. Reporting standards provide a practical ethical tool: they make invisible decisions visible and encourage consistent disclosure of the AI's role in the clinical pathway. For diagnostic AI used in trials, reporting extensions require authors to clarify how the AI is integrated into care, what training and expertise users require, how performance is measured, and what happens when the AI output conflicts with clinician judgment. Such requirements support patient safety by enabling critical appraisal and replication, and they support justice by making it harder to conceal subgroup harms behind aggregate metrics. From an institutional perspective, reporting standards also help procurement and governance. Hospitals often evaluate vendor claims using published studies; if these studies are incomplete or biased, health systems may purchase tools that do not deliver promised benefits. Ethically responsible procurement should therefore demand not only "peer-reviewed evidence" but evidence reported in a way that allows assessment of bias, applicability, and workflow impacts. The CONSORT-AI extension explicitly aims to improve the completeness and transparency of clinical trial reporting for interventions that include an AI component, supporting more reliable evaluation and safer translation into practice [5].

Lifecycle Governance, Post-Market Surveillance, and Continuous Learning

Unlike many traditional diagnostic devices, AI systems can change over time. Even "locked" models can degrade due to dataset shift, while adaptive or continuously learning systems may update parameters based on new data. Ethically, this creates a lifecycle responsibility: demonstrating safety at time of deployment is not enough; ongoing monitoring is required to detect performance drift, emergent bias, or new failure modes. Post-market surveillance in AI diagnostics should therefore be treated as a core ethical obligation aligned with patient safety and accountability. Lifecycle governance includes defining triggers for review (e.g., changes in imaging protocols, population shifts, new scanners), setting thresholds for acceptable performance, and establishing rapid response pathways when harm is suspected. It also requires careful change management: if a vendor updates a model, health systems should know what changed, why it changed, and whether local validation is needed before the update affects patient care. Documentation should track model versions used in decisions to enable audit and incident investigation. Regulatory and standards discussions increasingly emphasize "total product lifecycle" thinking for AI-enabled medical software, including planned change control, real-world performance evaluation, and feedback loops that support safe improvement. The FDA's AI/ML SaMD Action Plan highlights a lifecycle approach and the need to address ongoing learning, monitoring, and transparency as part of regulating AI-enabled software used in medical contexts [2].

Procurement Ethics, Conflicts of Interest, and Commercial Pressures

AI diagnostic tools are frequently developed and sold by commercial entities, creating ethical tensions around marketing claims, conflicts of interest, and vendor lock-in. Procurement decisions can be distorted by glossy demonstrations, selective evidence, or "benchmark" performance that does not reflect local populations. Ethical procurement requires rigorous due diligence: verifying evidence quality, demanding subgroup performance data, assessing cybersecurity and privacy controls, and evaluating integration costs and workflow impacts. It also requires transparency about financial relationships and incentives—both at the organizational level (contracts, revenue sharing) and at the clinician level (consulting or advisory roles). A common procurement risk is over-reliance on proprietary systems that cannot be independently audited. When vendors treat training data, model architecture, or performance details as trade secrets, hospitals may be unable to evaluate safety adequately or to respond effectively when harms occur. Ethical contracting should therefore include audit rights, clear performance guarantees, incident reporting obligations, and provisions for access to necessary technical documentation.

Procurement ethics also intersects with justice: if AI tools primarily serve high-resource institutions and widen diagnostic gaps between settings, then adoption decisions have distributive consequences. One response is to ground procurement expectations in widely recognized principles for trustworthy AI, including transparency, accountability, robustness, and fairness. International principles emphasize that actors across the AI lifecycle should be accountable for outcomes and should manage risks such as privacy, safety, and bias using appropriate processes and governance [10].

Discussion

This narrative review highlights that ethical evaluation of AI-based diagnostic systems must be framed as governance of a clinical intervention rather than appraisal of a standalone algorithm. While performance metrics remain important, ethical acceptability depends on how diagnostic AI changes clinical decision-making, redistributes responsibility, affects equity, and reshapes patient trust. Across themes, a consistent pattern emerges: many risks arise not from malicious intent but from misalignment between technological capability and the clinical-social context in which AI is deployed. Therefore, the most practical ethical stance is "systems ethics"—designing and governing AI diagnostics as part of a broader safety and accountability ecosystem. A key forward-looking issue is the rise of generative AI and large multimodal models in diagnostic support, including tools that synthesize clinical histories, propose differential diagnoses, or generate interpretations. These systems introduce distinctive hazards: plausible-sounding errors, uncertain provenance of generated text, and difficulty establishing stable performance under changing prompts and contexts. Governance guidance specific to generative AI in health underscores the need for heightened caution, transparency, and evaluation because these models can produce outputs that appear authoritative even when unsupported [11]. Ethically, this increases the responsibility to implement safeguards such as constrained use cases, strong oversight, and explicit communication of uncertainty.

Another priority is operationalizing ethical principles into enforceable institutional practice. High-level values (safety, privacy, fairness) become meaningful only when translated into measurable requirements: subgroup performance thresholds, monitoring intervals, documentation standards, and escalation pathways. Here, regulatory and standards ecosystems matter because they provide "shared languages" for what evidence, documentation, and monitoring should look like. Emerging guidance for good machine learning practice emphasizes lifecycle thinking, quality systems, and risk management that extend beyond initial development to real-world use and continuous evaluation [12]. Equity remains the most ethically demanding challenge because it requires both technical and structural responses. Subgroup testing and fairness metrics are necessary but not sufficient if the healthcare environment remains inequitable. Diagnostic AI can worsen disparities if it is trained on data reflecting unequal access, or if it is deployed as a gatekeeper that limits diagnostic resources for marginalized populations. Ethical governance must therefore include participatory approaches—engaging patient groups and communities, auditing real-world outcomes, and ensuring that benefits are shared. International ethics frameworks stress human rights, inclusion, and the need to avoid reinforcing discrimination through AI systems [13]. In practical terms, this suggests that fairness assessments should extend from model performance to downstream impacts such as access, timeliness, and quality of diagnostic pathways. Accountability also requires renewed attention. As institutions adopt AI, there is a temptation to treat AI output as a "neutral second opinion" and to leave responsibility with individual clinicians. Ethically, this is insufficient. Hospitals should develop explicit accountability maps: who owns model monitoring, who adjudicates disputes, who reports incidents, and how patients can seek explanations and remedies. Broader policy frameworks, including rights-based approaches, emphasize that people should be protected from unsafe or discriminatory automated systems, receive notice and explanation, and have access to human alternatives when appropriate [14]. Finally, research and evaluation norms must continue to mature. Standards like CONSORT-AI and TRIPOD+AI raise the baseline for reporting, but health systems also need pragmatic evaluation approaches to measure real-world benefit: clinical outcomes, workflow effects, unintended consequences, and cost-effectiveness. A helpful framing is comparative and contextual: AI should be evaluated against the actual baseline of care and the real constraints clinicians face, not against idealized conditions. Recent medical commentary has emphasized that measuring AI "against the health system" requires careful attention to what AI replaces, what new risks it introduces, and what outcomes matter most to patients and clinicians [15].

Strengths and Limitations of This Narrative Review

This narrative review integrates ethical, clinical, and governance perspectives to provide a structured thematic synthesis of key issues surrounding AI-based diagnostic systems in modern healthcare. A strength of the review is its use of authoritative sources (international guidance, regulatory documents, and high-impact peer-reviewed literature), supporting a balanced discussion that is relevant to both policy and clinical practice. The thematic organization helps clarify complex concepts such as accountability, bias, transparency, and lifecycle oversight in an accessible way. However, as a narrative (non-systematic) review, the search and selection process may be subject to selection bias and may not capture all relevant studies or the newest evidence in a rapidly evolving field. The review also does not quantitatively compare outcomes across AI systems or clinical specialties, limiting the ability to generalize findings to every diagnostic setting. Finally, many ethical challenges are context-dependent, so recommendations may require adaptation to local regulations, health system capacity, and population needs.

Conclusion

AI-based diagnostic systems offer significant opportunities to enhance detection, triage, and clinical decision-making, but their adoption raises substantial ethical challenges that directly influence patient safety, trust, and equity. This review emphasizes that ethical implementation requires more than strong model performance; it depends on transparency, accountability, privacy protection, fairness evaluation, and meaningful human oversight throughout the system lifecycle. Bias and real-world performance drift remain critical risks, underscoring the need for continuous monitoring, subgroup audits, and governance mechanisms capable of responding to harm. Healthcare organizations should embed ethical requirements into procurement, clinical workflows, and post-deployment surveillance to ensure that AI strengthens rather than undermines professional responsibility and patient autonomy. Ultimately, responsible diagnostic AI must align technical innovation with patient-centered values, regulatory standards, and institutional accountability to deliver benefits safely and fairly at scale..

REFERENCE

1. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance [Internet]. Geneva: World Health Organization; 2021. Available from: https://www.ncdirindia.org/Downloads/WHO_AI_Ethics.pdf
2. U.S. Food and Drug Administration. Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan [Internet]. Silver Spring (MD): FDA; 2021. Available from: https://www.fda.gov/media/145022/download
3. European Parliament; Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation) [Internet]. Official Journal of the European Union; 2016. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX%3A32016R0679
4. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med [Internet]. 2019;25(1):44-56. Available from: https://www.nature.com/articles/s41591-018-0300-7
5. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. BMJ [Internet]. 2020;370:m3164. Available from: https://www.bmj.com/content/370/bmj.m3164
6. National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1) [Internet]. Gaithersburg (MD): NIST; 2023. Available from: https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf
7. European Parliament; Council of the European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [Internet]. Official Journal of the European Union; 2024. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ%3AL_202401689
8. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science [Internet]. 2019;366(6464):447-453. Available from: https://eliassi.org/obermeyer_science_2019.pdf
9. Goddard K, Roudsari A, Wyatt JC. Automation bias: a hidden issue for clinical decision support system use. Stud Health Technol Inform [Internet]. 2011;164:17-22. Available from: https://pubmed.ncbi.nlm.nih.gov/21335682/
10. Organisation for Economic Co-operation and Development (OECD). Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449) [Internet]. Paris: OECD; 2019. Available from: https://legalinstruments.oecd.org/api/print?ids=648&lang=en
11. World Health Organization. Ethics and governance of artificial intelligence for health: large multi-modal models. WHO guidance [Internet]. Geneva: World Health Organization; 2025. Available from: https://www.who.int/publications/i/item/9789240084759
12. International Medical Device Regulators Forum (IMDRF). Good machine learning practice for medical device development: guiding principles (IMDRF/AIML WG/N88 FINAL:2025) [Internet]. IMDRF; 2025. Available from: https://www.imdrf.org/documents/good-machine-learning-practice-medical-device-development-guiding-principles
13. United Nations Educational, Scientific and Cultural Organization (UNESCO). Recommendation on the Ethics of Artificial Intelligence [Internet]. Paris: UNESCO; 2021. Available from: https://unesdoc.unesco.org/ark%3A/48223/pf0000380455
14. The White House Office of Science and Technology Policy. Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People [Internet]. Washington (DC): OSTP; 2022. Available from:

https://data.aclum.org/storage/2025/01/OSTP_www_whitehouse_gov_ostp_ai-bill-of-rights.pdf
15. Kohane IS. Compared with What? Measuring AI against the Health Care We Have. N Engl J Med [Internet]. 2024;391(17):1564-1566. Available from: https://www.nejm.org/doi/full/10.1056/NEJMp2404691.