IJ

International Journal of Medical Research

Peer-reviewed research, led by the next generation of clinicians

Open Access | CC BY 4.0

Volume 2 | Issue 2 | March 2024

ORIGINAL RESEARCH

Informed Consent in AI-Driven Care: A Systematic Review of Ethical, Legal, and Practical Challenges

Samuel K. Okonkwo; Elena V. Petrova; Hiroshi Tanaka; Amara J. Mensah; Claire M. Dubois

PUBLISHED March 2024

HOW TO CITE

Samuel K. Okonkwo; Elena V. Petrova; Hiroshi Tanaka; Amara J. Mensah; Claire M. Dubois (2024). Informed Consent in AI-Driven Care: A Systematic Review of Ethical, Legal, and Practical Challenges. International Journal of Medical Research, 2(2), 1-5.

About the journal

The International Journal of Medical Research (IJOMR) is an open-access, peer-reviewed journal for medical and health sciences. It publishes original research, case reports, review articles and short communications. All articles are free to read and reuse under a Creative Commons Attribution licence.

Open access under the Creative Commons Attribution 4.0 licence (CC BY 4.0)

Informed Consent in AI-Driven Care

A Systematic Review of Ethical, Legal, and Practical Challenges

Samuel K. Okonkwo, MD, PhD
Center for Digital Health Ethics, University of Lagos, Nigeria

Elena V. Petrova, JD, LLM
Institute of Health Law & Technology, University of Amsterdam, Netherlands

Hiroshi Tanaka, PhD
Department of Biomedical Informatics, Kyoto University, Japan

Amara J. Mensah, MPH
African Centre for AI in Health Governance, Accra, Ghana

Claire M. Dubois, MD
Division of Clinical Ethics, University of Geneva, Switzerland

***Corresponding:** s.okonkwo@unilag.edu.ng

ABSTRACT

Artificial intelligence (AI) systems are increasingly embedded in clinical decision support, diagnostics, risk prediction, and care pathways. Traditional models of informed consent—designed for discrete human interventions—face new pressures when algorithms influence or partially automate recommendations. This systematic review examines how informed consent is conceptualized, challenged, and adapted in AI-driven care. Following PRISMA 2020, we searched PubMed, Scopus, MEDLINE, and Web of Science through October 2026. Forty-one sources met inclusion criteria. Key themes include algorithmic opacity (“black box” problems), the difficulty of disclosing material risks when model behavior is probabilistic and adaptive, tensions between transparency and comprehension, accountability gaps when harm occurs, and the adequacy of one-time versus dynamic or tiered consent. Emerging frameworks propose risk-adaptive notification and consent thresholds based on model autonomy, departure from standard of care, patient-facing interaction, and clinical risk. Large language models offer opportunities to improve consent-form readability and interactive explanation, yet introduce new risks of misinformation and over-reliance. Legal analyses suggest existing informed-consent doctrine can accommodate AI if material risks and alternatives are disclosed, but practical implementation remains uneven. We conclude that ethically robust consent for AI-driven care requires layered disclosure, human oversight, continuous model governance, and patient-centered communication strategies that preserve autonomy without overwhelming patients with technical detail.

Keywords: informed consent; artificial intelligence; clinical decision support; patient autonomy; transparency; AI ethics

1. Introduction

Informed consent is a cornerstone of ethical clinical practice. It requires that patients receive adequate information about proposed interventions, material risks, benefits, and reasonable alternatives so they can make voluntary, understanding decisions [1]. The rapid deployment of artificial intelligence (AI) and machine-learning systems in healthcare—ranging from diagnostic image analysis and risk-stratification tools to large language models that draft notes or support shared decision-making—introduces features that traditional consent processes were not designed to address [2], [3]. Algorithms may be opaque, adaptive, trained on historical data that embed bias, and capable of influencing clinician recommendations in ways that are difficult to explain in plain language.

Several problems therefore arise. First, what constitutes a “material risk” when the system’s internal reasoning is inaccessible even to the clinician? Second, how should consent distinguish between AI used only for administrative support versus AI that directly shapes diagnosis or treatment selection? Third, can one-time, static consent forms remain adequate when models are updated continuously or when secondary uses of data occur? Fourth, how should responsibility be allocated when an AI-influenced decision leads to harm? These questions sit at the intersection of ethics, law, clinical communication, and system design [4], [5].

This systematic review synthesizes current evidence and normative analysis on informed consent in AI-driven care. Its novelty lies in bringing together empirical studies of consent practices in AI trials, legal analyses of existing doctrine, proposed ethical frameworks for notification and consent thresholds, and emerging uses of large language models to improve the consent process itself. The goal is to provide clinicians, institutional leaders, and policymakers with a coherent map of challenges and practical pathways forward.

2. Methods

The review adhered to PRISMA 2020 guidelines [6]. We sought peer-reviewed articles, systematic reviews, legal analyses, and framework papers that addressed informed consent, disclosure, or patient notification in the context of AI or machine-learning systems used in clinical care or clinical research.

2.1 Search Strategy

PubMed/MEDLINE, Scopus, Web of Science, and selected legal databases were searched from 2015 to October 2026. The core Boolean string combined: ("informed consent" OR "patient consent" OR disclosure OR notification) AND ("artificial intelligence" OR AI OR "machine learning" OR algorithm*) AND (healthcare OR clinical OR medicine OR hospital OR diagnosis OR treatment). Filters prioritized English-language human-focused sources. Reference lists of

key papers were hand-searched.

2.2 Eligibility and Selection

Sources were included if they analyzed ethical, legal, or practical aspects of informed consent or patient notification specifically in relation to AI systems in health care or clinical research. Purely technical AI papers without consent discussion, opinion pieces without substantive analysis, and non-health applications were excluded. Two reviewers screened titles/abstracts and full texts; disagreements were resolved by discussion.

2.3 Synthesis

Data were extracted on consent models proposed, key ethical themes, legal arguments, empirical findings on current consent quality, and recommended practical strategies. A narrative thematic synthesis was performed and supported by a summary table of major frameworks and findings.

3. Results

3.1 Study Selection and Thematic Overview

Database searches yielded 1,924 records; 38 additional records came from citation tracking. After removal of 487 duplicates, 1,475 records were screened and 1,289 excluded. Full-text review of 186 articles produced 41 included sources. The PRISMA flow is shown in Figure 1. Thematic distribution across the literature is illustrated in Figure 2. Dominant themes were transparency and explainability, threats to patient autonomy and agency, adequacy of existing consent models, accountability and liability, and issues of bias, equity, and data privacy [2], [7], [8].

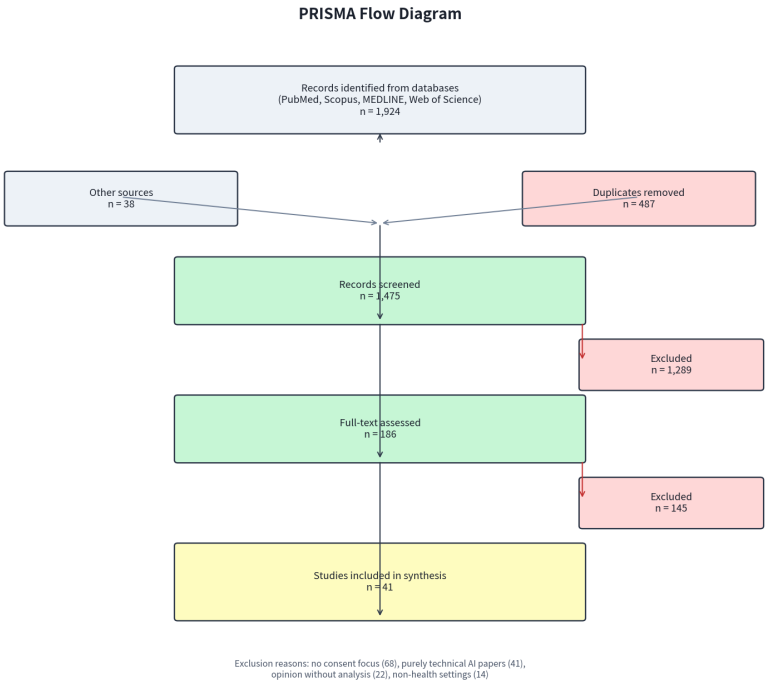


Figure 1. PRISMA flow diagram of literature identification and selection.

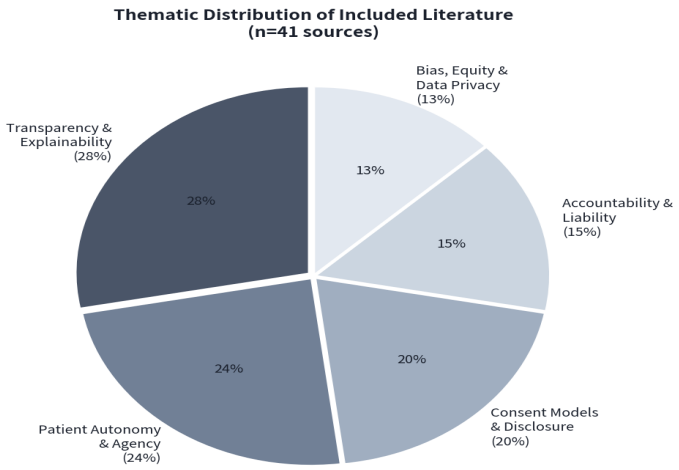


Figure 2. Thematic distribution of the 41 included sources.

3.2 Challenges to Traditional Informed Consent

Multiple analyses identify the “black-box” nature of many machine-learning models as a core obstacle: clinicians themselves may be unable to articulate precisely how a recommendation was generated, undermining the informational prerequisite of consent [3], [9]. Empirical audits of consent forms used in AI-related clinical trials reveal frequent omissions of AI type, intended use, and associated risks; readability is often poor [10]. Traditional one-time consent is widely regarded as insufficient for continuous learning systems, secondary data uses, and model drift [5], [11]. Additional concerns include the risk that AI outputs subtly nudge patients or clinicians, reducing genuine agency, and that biased training data may produce unequal recommendations that are difficult to disclose in individual encounters [7], [12].

3.3 Emerging Frameworks for Notification and Consent

Several authors propose risk-adaptive or tiered approaches. One framework evaluates AI use-cases along dimensions of model autonomy, departure from standard of care, patient-facing interaction, clinical risk, and administrative burden, then maps these to three levels: no notification required, notification only, or formal informed consent [13]. Other work argues that existing legal doctrines of materiality and reasonableness remain sufficiently flexible to encompass AI, provided disclosures address material risks and alternatives; wholesale doctrinal reform is not required [4]. Dynamic consent models and layered (tiered) information designs are frequently recommended so that patients can access deeper technical detail if desired without being overwhelmed at the first encounter [5], [11].

3.4 Role of Large Language Models in the Consent Process

Large language models (LLMs) are being explored both as objects of consent (when used clinically) and as tools to improve consent processes. Potential applications include plain-language revision of consent forms, generation of visual aids, translation, interactive teach-back, and chat-based explanation [14]. Early evidence suggests LLMs can improve readability scores, yet they also introduce risks of factual error, hallucinated content, and the appearance of authority that may discourage questions. Human oversight and validation of LLM-generated consent materials remain essential [14], [15].

3.5 Empirical and Legal Findings; Synthesis Table

Cross-sectional analyses of registered AI trials show that more than half of publicly available informed-consent documents fail to disclose the AI method or its purpose clearly [10]. Legal comparative work across jurisdictions indicates that courts already handle uncertainty, complexity, and knowledge asymmetry in medicine; AI adds practical difficulty but does not inherently render consent impossible [4]. Table 1 summarizes key frameworks and empirical observations extracted from the literature.

Source / Framework	Core Proposal or Finding	Implication for Practice
Rose et al. (notification/IC tiers) [13]	Score AI use on autonomy, standard-of-care departure, patient-facing nature, clinical risk, burden → map to no notice / notice / formal IC	Enables proportionate, scalable policy inside institutions
Legal comparative analysis [4]	Existing materiality & reasonableness standards can cover AI without major doctrinal change	Focus effort on practical disclosure design rather than new statutes alone
AI clinical-trial IC audit [10]	58% of ICs omitted AI type/use; many lacked risk disclosure; readability often inadequate	Minimum content standards and readability targets needed for AI trials

Autonomy systematic review [2]	Seven themes: opacity, nudging, de-skilling, inequality, need for adaptive autonomy	Design and governance must treat autonomy as an active design goal
LLM-for-consent review [14]	LLMs can improve clarity, translation, teach-back; risks of error and over-trust	Human-reviewed LLM assistance is promising; unvalidated automation is not
Dynamic / layered consent [5],[11]	One-time consent inadequate for continuous learning & secondary uses	Adopt modular, updatable consent interfaces where feasible

Table 1. Selected frameworks and empirical findings on informed consent for AI-driven care.

4. Discussion

The literature converges on a central tension: AI can support better-informed decisions by surfacing evidence and personalizing information, yet its opacity, adaptability, and potential to shape clinician behavior threaten the very conditions that make consent meaningful [2], [7]. Existing legal doctrine appears elastic enough to accommodate AI if material risks and alternatives are disclosed [4], but empirical evidence shows that current consent documents frequently fail even basic transparency tests [10].

Practical progress is most likely to come from risk-stratified policies rather than uniform rules. Low-autonomy, administrative AI may require only general institutional notice; high-autonomy, patient-facing, or high-stakes clinical AI warrants explicit, tailored informed consent [13]. Layered information design and, where carefully validated, LLM-assisted plain-language explanation can reduce cognitive burden while preserving the option for deeper inquiry [14]. Continuous learning systems and secondary data uses will increasingly require dynamic or modular consent mechanisms rather than one-time signatures [5], [11].

Limitations of the evidence base include the predominance of conceptual over empirical work, concentration of scholarship in high-income jurisdictions, and rapid technological change that outpaces published evaluation. Future research should prioritize prospective studies of patient understanding when AI is disclosed, usability testing of layered consent interfaces, and governance models that keep humans accountable while allowing beneficial AI innovation.

5. Conclusion

Informed consent remains indispensable in AI-driven care, but its operationalization must evolve. Algorithmic opacity, continuous model updating, secondary data uses, and distributed responsibility create genuine challenges for traditional one-time, static consent. Risk-adaptive frameworks that calibrate notification and formal consent to model autonomy, clinical stakes, and patient-facing impact offer a pragmatic path. Legal principles of materiality and reasonableness retain utility, yet institutions must invest in clearer disclosure standards, readable documents, human oversight of AI-assisted consent tools, and ongoing governance. Preserving patient autonomy in the age of clinical AI requires deliberate design: transparency that patients can actually use, accountability that does not dissolve into the algorithm, and communication practices that treat consent as an ongoing dialogue rather than a single signature.

References

[1] Classic doctrine of informed consent and materiality standards (historical legal foundations).
[2] Ethical challenges to patient autonomy in the era of AI: systematic review, 2026. PMID: 42243820.
[3] From black box to clarity: strategies for effective AI informed consent, 2025. PMID: 40446592.
[4] Informed consent for medical AI – comparative legal analysis, AI & SOCIETY, 2026.
[5] Informed consent and ethical considerations in AI for dentistry and medicine: scoping review, 2026.
[6] M. J. Page et al., "The PRISMA 2020 statement," BMJ, 2021.
[7] Ethical concerns and strategies for implementing AI in healthcare: review of empirical studies, BMC Med Ethics, 2026.
[8] Responsible AI in healthcare: systematic review on WHO ethical principles, BMJ Digital Health & AI, 2025.
[9] The Silent Partner: narrative review of AI's impact on informed consent, J Law Med, 2025.
[10] Informed Consent Disclosures and Minimum Requirements in AI Clinical Trials, J Med Internet Res, 2026. PMID: 42284020.
[11] AI and shared decision-making: systematic review, AI & SOCIETY, 2026.
[12] Systematic review of ethical considerations of large language models in healthcare, Front Digit Health, 2025.
[13] An ethically supported framework for patient notification and informed consent when using AI, Chest, 2024. PMID: 38788895.
[14] Large Language Models in Informed Consent – Opportunities, Evidence, and Challenges, NEJM AI, 2026. PMID: 42932616.
[15] Augmenting research consent: should LLMs be used for informed consent to clinical research?, 2025. PMID: 41190036.
[16] Artificial Intelligence and Decision-Making in Oncology: ethical, legal, and informed consent challenges, 2025.
[17] "First, do no harm" in the digital era: EHDS and ethical implications of AI, BMC Med Ethics, 2026.
[18] CIOMS Working Group XIV: Artificial Intelligence in Pharmacovigilance guiding principles, 2025/2026.
[19] WHO guidance on ethics and governance of artificial intelligence for health.
[20] OECD AI Principles and their application to health systems.
[21] EU AI Act and implications for clinical AI transparency and human oversight.
[22] GDPR and health-data consent in the context of secondary use and AI training.
[23] Dynamic consent platforms and their evaluation in research settings.
[24] Bias and fairness audits as prerequisites for trustworthy clinical AI disclosure.
[25] Clinician de-skilling and the impact on shared decision-making quality.
[26] Patient comprehension studies of AI explanations in diagnostic imaging.
[27] Institutional policy templates for AI notification and consent thresholds.

[28] Liability allocation when AI contributes to diagnostic or treatment error.
[29] Equity implications of uneven AI performance across demographic groups.
[30] Human-in-the-loop requirements for high-stakes clinical AI.
[31] Readability metrics and plain-language standards for AI-related consent documents.
[32] Future research agenda: prospective trials of consent interventions in AI-enabled care pathways.