

PERSPECTIVE

Human Oversight Is Not Enough: Rethinking Accountability in AI-Assisted Clinical Decision-Making

Yangyi Ren[†]

Department of Clinical Medicine, Second Clinical Medical College, Shaanxi University of Chinese Medicine, Xianyang 712046, Shaanxi, China

[†]Corresponding author. Email: 2464838079@qq.com

Abstract

Governance of artificial intelligence (AI) in clinical care often places clinicians between AI outputs and care decisions, yet a human signature can coexist with weak practical control. This Perspective asks why participation may fail to produce accountable oversight, how duties should be distributed, and which institutional conditions make review meaningful and correctable. Drawing on human–factors research, sociotechnical governance, public-accountability theory, and recent oversight proposals, we argue that nominal involvement cannot establish accountability when clinicians lack relevant information, time, authority, or workable review routes. We define Accountable Human Oversight (AHO) as a normative institutional arrangement linking task-specific human review to capability, authority, traceability, and redress. These conditions overlap with prior frameworks, including a 2026 model. AHO’s narrower proposed contribution is to connect them to actor–forum answerability and accessible clinical reconsideration, rather than to claim new oversight dimensions. We outline risk–proportionate governance that assigns review ownership, evidence, triggers, and corrective actions across developers and model suppliers, healthcare institutions, clinicians, and regulators, while reserving case-level review for settings in which it can add value. AHO has not been empirically evaluated and is not evidence of improved patient outcomes. Future research should examine its feasibility, workload, equity, and effects.

Keywords: Artificial Intelligence; Clinical Decision Support Systems; Human Oversight; Algorithmic Accountability; Clinical Governance; Patient Safety

1. Introduction

Clinician-facing artificial intelligence (AI) can summarize evidence, estimate risk, or recommend a diagnostic or treatment option. A clinician may retain the final decision while an institution treats review as a safeguard. International guidance emphasizes autonomy and stakeholder accountability, while legal duties vary by product function and jurisdiction (World Health Organization 2021; European Parliament and Council of the European Union 2024). This Perspective focuses on decision support before consequential clinical action, not administrative automation, surgical robots, patient-facing chatbots, or autonomous diagnosis.

Human involvement describes a workflow position, not oversight quality. A clinician may see an output without understanding its intended use, time to check it, or freedom to reject it. Automation-bias research identifies conditions for over-reliance, while recent commentaries warn that “clinician



This article is published in *Health Nexus: Digital Health and Medical AI* under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

in the loop” policies can shift safety burdens to those with little control over design and deployment (Goddard, Roudsari, and Wyatt 2012; Lyell and Coiera 2017; Toro-Tobon et al. 2026).

Prior work treats healthcare AI as sociotechnical and accountability as distributed. McCradden et al. (2023) frame an intervention ensemble; Martinho et al. (2026) map multilevel responsibility. Wang, Freeman, and Magrabi (2026) found that 15 of 77 organizational frameworks included oversight and 10 included all four coded components—principles, assessment methods, lifecycle stages, and oversight mechanisms. These are content counts, not outcome measures; the review also discusses gaps in contestability and patient redress. Van de Sande, Economou-Zavlanos, and van Genderen (2026) propose closely overlapping oversight conditions. AHO claims no new oversight domains.

We ask why participation may fail, how duties should be distributed, and what institutions must provide for review and correction. Accountable Human Oversight (AHO) is a normative synthesis linking practical capability and authority to proportionate records, answerability forums, and patient reconsideration. Its contribution is this operational link, not a new taxonomy or validated model. We draw selectively on scholarship about automation, sociotechnical governance, accountability, clinical implementation, and risk-based regulation; this non-systematic synthesis includes no new clinical data or model testing.

2. The Limits of Human Oversight

2.1 Automation Bias and Overreliance

Automation bias is a possible, context-dependent reliance pattern. Reviews link decision-support errors to workload, task complexity, time pressure, cognitive load, and verification difficulty; training and interface changes may mitigate some risks, although evidence is heterogeneous (Goddard, Roudsari, and Wyatt 2012; Lyell and Coiera 2017). Human-factors theory likewise treats reliance as shaped by operator, task, and system context (Parasuraman and Riley 1997; Lee and See 2004).

Clinical value remains task-dependent. In a randomized vignette trial of 50 physicians, GPT-4 access did not significantly improve diagnostic reasoning over conventional resources. This does not establish that AI assistance is ineffective in practice; it shows that access alone did not improve performance in that task (Goh et al. 2024). The goal is calibrated reliance: review should consider acceptance of erroneous advice and rejection of useful advice, and whether staff can detect errors in the workflow.

2.2 Epistemic Asymmetry and Verification Burden

Clinicians may see a score or recommendation without enough information to judge its relevance to a patient. Developers may know intended populations, validation limits, updates, and failure modes that are not visible in concise outputs. Intended-use disclosure differs from explaining an individual prediction, and neither guarantees correctness. Without information on scope, uncertainty, and local validation, clinicians may miss out-of-scope advice.

Checking complex recommendations may require time and data unavailable in routine care. Verification difficulty and cognitive load can increase reliance, while recent guidance calls for accessible information on system limits and subgroup performance (Lyell and Coiera 2017; van de Sande, Economou-Zavlanos, and van Genderen 2026). A formal right to disagree offers little protection when clinicians cannot identify what to question.

2.3 Organizational Constraints on Clinical Judgment

Oversight depends on surrounding conditions: interface defaults, alerts, staffing, timing, policy, and performance targets can enable or constrain review. Automation can leave clinicians with infrequent, difficult exceptions while they retain responsibility but lose routine practice; task allocation and automation level therefore matter (Bainbridge 1983; Parasuraman, Sheridan, and Wickens 2000).

Sociotechnical models likewise treat people, workflow, rules, infrastructure, and monitoring as interdependent (Sittig and Singh 2010).

Human supervision cannot correct failures in procurement, updates, input data, interface design, or monitoring. In an analysis of 266 FDA-reported events involving machine-learning devices, reported problems were mostly device-related, while use problems were more likely to involve harm. Reports lack a use denominator and cannot establish incidence rates (Lyell et al. 2023). This supports system-level investigation, not clinician-blaming, and review of the whole decision pathway.

3. From Individual Responsibility to Distributed Accountability

Responsibility and accountability are related but distinct. Responsibility assigns duties: clinicians use professional judgment, while institutions provide a safe workflow. Accountability requires an actor to explain and justify conduct to a forum that can question it, judge it, and attach consequences (Bovens 2007). Forums may include clinical peers, an institutional safety body, or a regulator acting within its mandate. Patient reconsideration is a related but distinct route; it does not itself determine legal liability or compensation. Legal liability depends on applicable law, facts, and jurisdiction. Reviews identify unresolved questions across medical-AI supply chains, rather than a uniform rule assigning every error to clinicians (Cestonaro et al. 2023).

Developer and supplier duties should follow control over the base model, clinical application, integration, and updates. Model providers may control general documentation; application developers or vendors configure clinical tasks and interfaces; integrators connect products to local workflows. The responsible actors should provide intended use, validation scope, known limits, updates, and technical support, with contacts and escalation paths documented before deployment. This does not make suppliers responsible for every clinical choice, but clinicians cannot assess undisclosed or unreconstructable risks. WHO guidance calls for stakeholder accountability, and the EU AI Act gives a jurisdiction-specific example for systems within its high-risk categories (World Health Organization 2021; European Parliament and Council of the European Union 2024).

Healthcare institutions control procurement, local adoption, workflow design, training, monitoring, incidents, and patient pathways. They should assess fit to the intended population and care process, allocate time and expertise, name a safety owner, and retain authority to restrict or suspend use. Governance recommendations for AI-enabled clinical decision support likewise emphasize lifecycle responsibilities (Labkoff et al. 2024; Lekadir et al. 2025). Clinicians remain responsible for reasonable encounter-level judgment within available information and resources. A clinical lead can review discordance; a multidisciplinary safety body can assess local performance and incidents and require correction, restriction, or suspension. Route product findings to suppliers; regulators act within statutory mandates. FDA device status depends on software function and statutory criteria, so clinical decision support is not one regulatory category (U.S. Food and Drug Administration 2026). These are proposed arrangements, not universal legal rules. Duties follow control and capacity; shared duties do not create equal legal liability.

4. Toward Accountable Human Oversight

We define AHO as an institutional arrangement in which designated decision-makers can assess or intervene in AI-assisted decisions, while organizations preserve proportionate evidence and provide review, correction, and patient redress. Oversight is both professional work and an organizational capability, linking public-accountability theory's actor-forum relationship to sociotechnical healthcare AI (Bovens 2007; McCradden et al. 2023). The conditions are normative necessities for AHO when human review is selected or required, not sufficient to guarantee safety; they do not imply case-by-case sign-off for every output. Figure 1 summarizes the failure mechanisms, AHO conditions, and institutional feedback path.

AHO builds on, rather than replaces, prior accounts. Van de Sande, Economou-Zavlanos, and van Genderen (2026) identify epistemic capacity and cognitive space, which align with AHO's capability; decisional authority aligns with authority; and intervention effectiveness overlaps with authority to act and redress. AHO treats traceability as proportionate evidence for forum review. Wang, Freeman, and Magrabi (2026) likewise discuss organizational oversight and gaps in contestability and patient redress. We claim neither new oversight domains nor unique review owners, triggers, or corrective mechanisms. AHO's narrower contribution is an explicit pathway from patient-accessible reconsideration through a named forum and documented disposition to correction of care or system practice. This integrative normative proposal overlaps with prior governance work and requires evaluation; it is not a validated model (Abulibdeh et al. 2026; Omeljanić et al. 2026; Toro-Tobon et al. 2026).

4.1 Capability: The Ability to Exercise Critical Judgment

Capability is the ability to understand a tool's intended role and limits, recognize out-of-scope cases, and verify material information when needed. It requires evidence on validation populations, uncertainty and known failure modes, model-specific training, technical support, and realistic review time—not reverse engineering. Otherwise, review may become blind acceptance or indiscriminate rejection. Institutions should define whether a task is contextual-fit review, discordance checking, or escalation and assign it to a role with relevant clinical context. Scenario-based workflow checks before deployment and after material updates should assess comprehension and review time; attendance records alone show exposure, not performance. Developers provide use information, and institutions judge local adequacy (Labkoff et al. 2024; European Parliament and Council of the European Union 2024).

4.2 Authority: The Power to Challenge or Override AI

Authority is the practical ability of clinicians and designated safety staff to decline, override, escalate, or pause advice without policy, incentives, or supervision making justified disagreement impracticable. A nominal "final decision" is insufficient if override is burdensome, penalized, or lacks an escalation route; oversight research and critiques of clinician-in-the-loop policies emphasize institutional authority (van de Sande, Economou-Zavlanos, and van Genderen 2026; Toro-Tobon et al. 2026). Institutions should define override and suspension rights, document discordance, protect reasoned dissent, and review override patterns as signals about local fit rather than compliance scores. Authority does not make every override sound: training should support justified challenge, while developers and leaders remain answerable for systemic defects.

4.3 Traceability: The Evidence Required for Accountability

Traceability is the proportionate ability to reconstruct the system version and intended-use conditions, recommendation shown, and resulting action or escalation. It lets a governance forum examine what happened, but is not explainability: logs show outputs without necessarily explaining model reasoning, while explanations do not establish suitability. Developers and institutions should agree on records before deployment, such as model/configuration versions, relevant outputs, clinician response, escalation status, and rationale when needed. Data minimization, access controls, retention limits, and privacy-preserving audits bound collection; indefinite retention of all patient data is not required. Lifecycle guidance places documentation alongside broader governance (Lekadir et al. 2025; Wang, Freeman, and Magrabi 2026).

4.4 Redress: Institutional Correction and Patient Protection

Redress links concerns to reconsideration, correction, investigation, and patient response. It can prompt prospective controls or retrospective review but does not determine negligence or damages.

Logs without review may document errors without preventing recurrence. Institutions should offer reconsideration through ordinary clinical or complaint channels without requiring AI awareness; when AI materially informs consequential advice, explain how to request accessible review and permit appropriate representatives. Route product defects to suppliers and reportable events to authorities, subject to law. Periodically examine available subgroup performance and whether patients can access and complete reconsideration, considering language, disability, health literacy, and resource barriers. Check pathway visibility and usability through periodic reviews and, where feasible, patient or community feedback; do not infer access from request counts alone. Protect privacy and use lawful data. Record missing data as a limitation, not evidence of equal access (Abulibdeh et al. 2026; World Health Organization 2021). Capability enables appraisal; authority permits intervention; traceability supports review; redress connects findings to correction. These conditions are necessary for proposed AHO, not sufficient for safety.

5. Translating Accountability into Clinical Governance

5.1 Risk-Proportionate Human Oversight

Oversight intensity should reflect intended use, likely consequence, model autonomy, reversibility, and whether a reviewer can detect an error with available information and time. Higher risk does not always mean more clinician sign-offs. It may require stronger local validation, independent checks, rapid escalation, a pause function, or authorization by a clinical safety lead. Where bedside detection is unreliable, strengthen upstream controls rather than add nominal review. Depending on actionability and workload, use case-by-case, sampled, or selective escalation. The EU AI Act illustrates a risk-based legal approach for specified high-risk systems, while FDA guidance distinguishes some clinical decision-support functions from regulated device functions; neither establishes a universal classification for clinical AI (European Parliament and Council of the European Union 2024; U.S. Food and Drug Administration 2026).

5.2 Documentation and Incident Review

Institutions should document intended use, local approval, model version, material updates, training, and review responsibility. Case records should allow reconstruction when concerns arise without requiring duplicative narrative for every routine output. A named safety owner should schedule reviews at intervals set by local risk and trigger additional review after a material update, safety signal, near miss, complaint, or repeated discordance. Findings should lead to a documented disposition, such as correction, workflow change, restriction, or suspension. Override frequency is a contextual signal, not a pass-fail measure; both high and low rates can have different causes (van de Sande, Economou-Zavlanos, and van Genderen 2026). Access, retention, and data minimization rules should keep audit trails proportionate.

5.3 Institutional Responsibilities and Patient Protection

Consider a hypothetical illustration: a hospital adopts a diagnostic support tool, but the vendor provides little information about its limits and the hospital has no convenient second-review route. Capability is weak, authority nominal, and traceability incomplete. A clinical lead could review disputed encounters, while an institutional safety body examines local fit and incident patterns and can require supplier investigation, workflow changes, restriction, or a pause. Patients could request reconsideration through ordinary clinical or complaint channels without identifying AI use. Before deployment, the hospital should obtain use information, test escalation, and keep proportionate records. This is a governance illustration, not a reported event.

Table 1 turns these conditions into reviewable questions and evidence. Governance committees should assign named owners, review evidence at defined intervals, and have authority to require modification, restrict use, or suspend a tool. A procurement checklist alone is insufficient if no group

receives post-deployment signals or can act on them. These recommendations complement existing implementation and lifecycle frameworks; they do not imply that institutions can replace regulatory review or developers' technical obligations (McCradden et al. 2023; Labkoff et al. 2024; Wang, Freeman, and Magrabi 2026).

6. Boundaries and Future Research

AHO cannot remove cognitive or time limits. Confirming every low-consequence output may add delay, alerts, and defensive documentation, leaving less attention for consequential decisions. Proportionate review should consider frequency, intervention type, and whether sampling or escalation is preferable to universal sign-off. Unequal staffing, technical capacity, and procurement leverage limit transfer from large academic centers to under-resourced services. Governance plans should specify whether specialist expertise and audit capacity come from the institution, a shared service, a vendor agreement, or another support structure, rather than leaving these costs implicit.

The framework addresses clinician-facing diagnosis, risk, and treatment recommendations. General-purpose language models, validated scores, and regulated device functions differ in intended use, evidence, updates, and law; one checklist cannot settle their risks. The EU and US examples are jurisdiction-specific, not comparative legal analysis. AHO is a normative synthesis, not a validated construct, scale, or evidence of improved outcomes. Its overlap with recent frameworks makes comparison of definitions and practical use more useful than treating labels as validated alternatives.

Future studies could interview clinicians, patients, developers, and managers; simulate time-pressured decisions; audit governance and privacy; and assess workflows before and after deployment. DECIDE-AI can inform reporting of early clinical evaluation and user interaction (Vasey et al. 2022). Studies should examine workload, erroneous acceptance and unjustified rejection, subgroup access, review latency, complaints, and corrective action.

7. Conclusion

Clinician participation is necessary in many high-stakes applications but cannot by itself make oversight meaningful or establish who must answer when a system contributes to error. AHO offers a normative way to connect practical capability and authority with proportionate traceability, review, and patient redress. Its four conditions synthesize and overlap with existing work; the conceptual increment is their explicit connection to an answerability forum and a route to remedy. Responsibility should follow control and resources across developers, healthcare institutions, clinicians, and regulators, with legal liability determined under applicable law. Meaningful accountability in clinical AI depends not merely on keeping humans in the decision-making loop, but on ensuring that human judgment is supported by institutional capability, authority, traceability, and mechanisms for correction and redress.

Declarations

Author Contributions

Yangyi Ren: Conceptualization, Investigation, Writing—original draft, Writing—review and editing, Visualization.

Funding / Supporting Agencies

This work was supported by Panorama Research Institute under Sponsorship No. PSG-PRI-SPN-2026-000018. The sponsor provided financial support.

Competing Interests

The author reports the financial support disclosed above. No other financial or non-financial competing interests were reported.

Ethics Approval and Informed Consent

Not applicable. This Perspective is based exclusively on publicly available scholarly literature and policy documents and does not involve human participants or identifiable patient data.

Data Availability

No new datasets were generated or analysed for this article. The scholarly and policy sources discussed are identified in the reference list.

Generative AI Use

DeepSeek was used to assist with language polishing of the manuscript. The author takes full responsibility for the accuracy, integrity, and final content of the article.

Acknowledgements

No additional acknowledgements.

References

1. Abulibdeh, Rawan, Gift Osei Agyemang, Leo Anthony Celi, Rahul Gorijavolu, Nai Lee Kalema, Ricardo Kleinlein, Kaushik Madapati, Seyed Reza Salarikia, and Alaa Youssef. 2026. "Who's Really in the Loop? Rethinking Oversight in AI-Assisted Health Care." *The Lancet* 407, no. 10545: 2340–2344. [https://doi.org/10.1016/S0140-6736\(26\)00204-7](https://doi.org/10.1016/S0140-6736(26)00204-7)
2. Bainbridge, Lisanne. 1983. "Ironies of Automation." *Automatica* 19, no. 6: 775–779. [https://doi.org/10.1016/0005-1098\(83\)90046-8](https://doi.org/10.1016/0005-1098(83)90046-8)
3. Bovens, Mark. 2007. "Analysing and Assessing Accountability: A Conceptual Framework." *European Law Journal* 13, no. 4: 447–468. <https://doi.org/10.1111/j.1468-0386.2007.00378.x>
4. Cestonaro, Clara, Arianna Delicati, Beatrice Marcante, Luciana Caenazzo, and Pamela Tozzo. 2023. "Defining Medical Liability When Artificial Intelligence Is Applied on Diagnostic Algorithms: A Systematic Review." *Frontiers in Medicine* 10: Article 1305756. <https://doi.org/10.3389/fmed.2023.1305756>
5. European Parliament and Council of the European Union. 2024. "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)." *Official Journal of the European Union* L, 2024/1689. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
6. Goh, Ethan, Robert Gallo, Jason Hom, Eric Strong, Yingjie Weng, Hannah Kerman, Joséphine A. Cool, Zahir Kanjee, Andrew S. Parsons, Neera Ahuja, Eric Horvitz, Daniel Yang, Arnold Milstein, Andrew P. J. Olson, Adam Rodman, and Jonathan H. Chen. 2024. "Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial." *JAMA Network Open* 7, no. 10: Article e2440969. <https://doi.org/10.1001/jamanetworkopen.2024.40969>
7. Goddard, Kate, Abdul Roudsari, and Jeremy C. Wyatt. 2012. "Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators." *Journal of the American Medical Informatics Association* 19, no. 1: 121–127. <https://doi.org/10.1136/amiajnl-2011-000089>

8. Labkoff, Steven, Bilikis Oladimeji, Joseph Kannry, Anthony Solomonides, Russell Leftwich, Eileen Koski, Amanda L. Joseph, Monica Lopez-Gonzalez, Lee A. Fleisher, Kimberly Nolen, Sayon Dutta, Deborah R. Levy, Amy Price, Paul J. Barr, Jonathan D. Hron, Baihan Lin, Gyana Srivastava, Nuria Pastor, Unai Sanchez Luque, Tien Thi Thuy Bui, Reva Singh, Tayler Williams, Mark G. Weiner, Tristan Naumann, Dean F. Sittig, Gretchen Purcell Jackson, and Yuri Quintana. 2024. "Toward a Responsible Future: Recommendations for AI-Enabled Clinical Decision Support." *Journal of the American Medical Informatics Association* 31, no. 11: 2730–2739. <https://doi.org/10.1093/jamia/ocae209>
9. Lee, John D., and Katrina A. See. 2004. "Trust in Automation: Designing for Appropriate Reliance." *Human Factors* 46, no. 1: 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
10. Lekadir, Karim, Alejandro F. Frangi, Antonio R. Porras, and the FUTURE-AI Consortium. 2025. "FUTURE-AI: International Consensus Guideline for Trustworthy and Deployable Artificial Intelligence in Healthcare." *BMJ* 388: Article e081554. <https://doi.org/10.1136/bmj-2024-081554>
11. Lyell, David, and Enrico Coiera. 2017. "Automation Bias and Verification Complexity: A Systematic Review." *Journal of the American Medical Informatics Association* 24, no. 2: 423–431. <https://doi.org/10.1093/jamia/ocw105>
12. Lyell, David, Ying Wang, Enrico Coiera, and Farah Magrabi. 2023. "More Than Algorithms: An Analysis of Safety Events Involving ML-Enabled Medical Devices Reported to the FDA." *Journal of the American Medical Informatics Association* 30, no. 7: 1227–1236. <https://doi.org/10.1093/jamia/ocad065>
13. Martinho, Domingos, Pedro Sobreiro, Andreia Domingues, Filipa Martinho, and Nuno Nogueira. 2026. "Ethical Responsibility in Medical AI: A Semi-Systematic Thematic Review and Multilevel Governance Model." *Healthcare* 14, no. 3: Article 287. <https://doi.org/10.3390/healthcare14030287>
14. McCradden, Melissa D., Shalmali Joshi, James A. Anderson, and Alex John London. 2023. "A Normative Framework for Artificial Intelligence as a Sociotechnical System in Healthcare." *Patterns* 4, no. 11: Article 100864. <https://doi.org/10.1016/j.patter.2023.100864>
15. Omeljancik, Aleksej, Eimantas Peičius, Aušra Urbonienė, and Gvidas Urbonas. 2026. "Beyond Algorithmic Oversight: Internal Morality of Medicine and Meaningful Human Control in AI-Assisted Care." *Healthcare* 14, no. 12: Article 1638. <https://doi.org/10.3390/healthcare14121638>
16. Parasuraman, Raja, and Victor Riley. 1997. "Humans and Automation: Use, Misuse, Disuse, Abuse." *Human Factors* 39, no. 2: 230–253. <https://doi.org/10.1518/001872097778543886>
17. Parasuraman, Raja, Thomas B. Sheridan, and Christopher D. Wickens. 2000. "A Model for Types and Levels of Human Interaction with Automation." *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans* 30, no. 3: 286–297. <https://doi.org/10.1109/3468.844354>
18. Sittig, Dean F., and Hardeep Singh. 2010. "A New Sociotechnical Model for Studying Health Information Technology in Complex Adaptive Healthcare Systems." *Quality and Safety in Health Care* 19, suppl. 3: i68–i74. <https://doi.org/10.1136/qshc.2010.042085>
19. Toro-Tobon, David, Oscar Ponce Ponte, Victor M. Montori, and Juan P. Brito. 2026. "Clinician in the Loop: A Flawed Solution for AI Oversight." *BMJ* 393: Article e089213. <https://doi.org/10.1136/bmj-2025-089213>
20. U.S. Food and Drug Administration. 2026. *Clinical Decision Support Software: Guidance for Industry and Food and Drug Administration Staff*. January 29. <https://www.fda.gov/regulatory-information/>

search-fda-guidance-documents/clinical-decision-support-software

21. van de Sande, Davy, Nicoleta Economou-Zavlanos, and Michel E. van Genderen. 2026. "Meaningful Oversight of Medical AI beyond Human in the Loop." *npj Digital Medicine* 9, no. 1: Article 569. <https://doi.org/10.1038/s41746-026-02971-1>
22. Vasey, Baptiste, Myura Nagendran, Bruce Campbell, David A. Clifton, Gary S. Collins, Spiros Denaxas, Alastair K. Denniston, Livia Faes, Bart Geerts, Mudathir Ibrahim, Xiaoxuan Liu, Bilal A. Mateen, Piyush Mathur, Melissa D. McCradden, Lauren Morgan, Johan Ordish, Campbell Rogers, Suchi Saria, Daniel S. W. Ting, Peter Watkinson, Wim Weber, Peter Wheatstone, Peter McCulloch, and the DECIDE-AI Expert Group. 2022. "Reporting Guideline for the Early Stage Clinical Evaluation of Decision Support Systems Driven by Artificial Intelligence: DECIDE-AI." *BMJ* 377: Article e070904. <https://doi.org/10.1136/bmj-2022-070904>
23. Wang, Amy, Sam Freeman, and Farah Magrabi. 2026. "Governance for Safe and Responsible AI in Healthcare Organisations: A Scoping Review of Frameworks." *npj Digital Medicine* 9, no. 1: Article 516. <https://doi.org/10.1038/s41746-026-02679-2>
24. World Health Organization. 2021. *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*. Geneva: World Health Organization. <https://www.who.int/publications/i/item/9789240029200>

Table 1. Operationalizing Accountable Human Oversight in AI-Assisted Clinical Decision-Making

Oversight condition	Principal risk addressed	Proposed institutional measure	Potential review evidence	Primary responsible actors
Capability	Uncritical reliance; failure to detect use outside the intended scope	Provide intended use, validation limits, local performance evidence, training, support, and adequate review time	Training materials; scenario checks before deployment and after updates; records of error recognition and review workload	Developers or suppliers; institution; clinical lead
Authority	Nominal approval; barriers to justified disagreement	Define rights to reject, override, escalate, or pause AI use; name a review forum and protect reasoned dissent	Override and suspension policies; tested escalation routes; records of disputed outputs and responses	Institution; clinical leaders; safety body
Traceability	Missing records of system version, AI output, or clinical response	Record model version, relevant inputs and outputs, clinical response, and incidents; set access and retention limits	Version and update records; sampled audits; incident reviews; privacy and security checks	Developers; integrators; health IT; safety owner
Redress	Unresolved concerns; barriers to clinical review	Provide accessible human review and complaint routes; assign owners, response times, corrective powers, and escalation duties	Records of requests, completion, delays, decisions, and corrections; usability checks and assessment of access barriers	Institution; clinical service; patient relations; safety body

Note: These are proposed measures and review indicators, not validated performance thresholds or guarantees of safety. Reviews should occur at risk-appropriate intervals and after material updates or safety signals. Records should be proportionate and protect privacy. Suppliers and regulators participate in investigation and correction within their respective responsibilities and legal mandates; actor allocation does not imply equal legal liability.

Abbreviations: AI, artificial intelligence; IT, information technology.

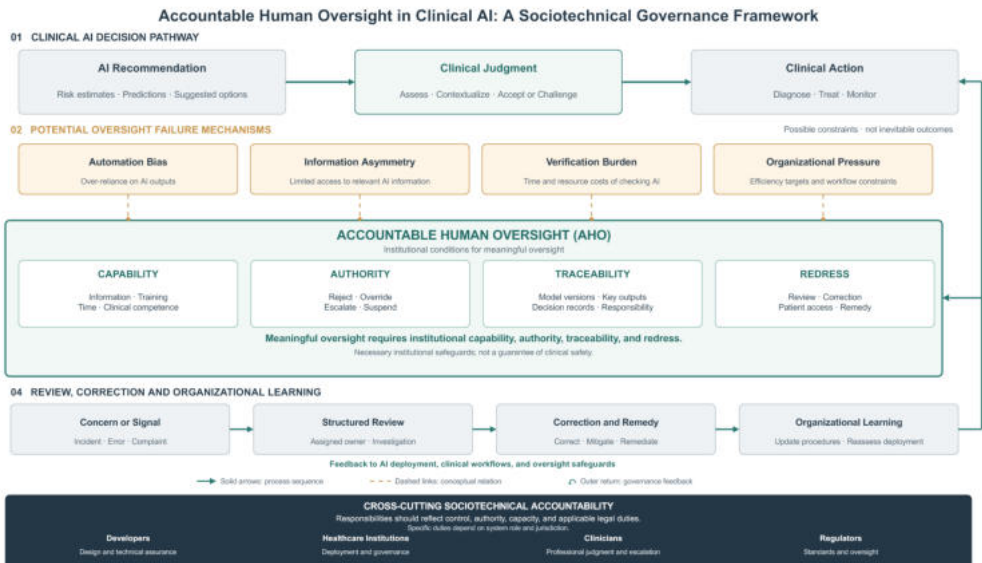


Figure 1. Accountable Human Oversight in Clinical AI: A Sociotechnical Governance Framework. The figure distinguishes the clinical decision pathway from potential constraints on meaningful human oversight. Dashed links mark conceptual relationships between those constraints and the proposed institutional conditions of capability, authority, traceability, and redress; they do not represent a validated causal sequence. The review pathway connects concerns with structured review, correction or remedy, and organizational learning, which feeds back into AI deployment, clinical workflows, and oversight safeguards. The cross-cutting accountability band assigns responsibilities across developers, healthcare institutions, clinicians, and regulators according to control, authority, capacity, and applicable legal duties. This normative framework has not been empirically validated; its proposed conditions are safeguards, not guarantees of clinical safety.