Society of Medical Education & Research

2026, Volume 6, Issue 1, Page No: 134-143
ISSN: 3108-5059

Asian Journal of Ethics in Health and Medicine

Moral Residue After Algorithmic Advice

Wei Zhang[1]*, Chen Hui[2], Michael Tan[1]

[1]Department of Algorithmic Ethics and Moral Residue, Faculty of Medicine, University of Melbourne, Melbourne, Australia.
[2]Department of Clinical Decision-Making and AI Ethics, Faculty of Medicine, National University of Singapore, Singapore.

***E-mail** ✉ wei.zhang@unimelb.edu.au

Abstract

Clinical decision support increasingly places clinicians in a distinctive moral position: they remain answerable for decisions while acting in an informational environment partly structured by algorithmic recommendations. The ethical problem therefore begins not only when an algorithm makes a decision, but after advice has been received and a clinician must follow, modify, delay, seek review of, override, or ignore it. This normative-responsibility analysis examines how those responses alter the grounds on which responsibility should be attributed and how unresolved responsibility may leave moral residue even without proven wrongdoing. It argues that formal decision authority is insufficient to determine moral responsibility because responsibility also depends on epistemic access, meaningful control, causal contribution, professional role, available alternatives, institutional conditions, and the capacity to question algorithmic advice. The article proposes a relational account in which responsibility is distributed across actors according to their actual relationship to the decision rather than assigned automatically to the final human decision-maker. Moral residue is subsequently conceptualized as a morally salient remainder that may persist when value conflict, uncertainty, constrained choice, or unresolved accountability survives an otherwise defensible decision. The proposed framework distinguishes prospective accountability from retrospective blame and treats documentation, explanation, review, acknowledgement, and learning as potential resources for moral repair rather than demonstrated remedies. Its claims are bounded by variation among clinical systems, professions, institutions, patient populations, and jurisdictions. The framework is normative and theory-generating; it does not establish an AI-specific syndrome, causal pathway, liability rule, validated intervention, or universal allocation of responsibility.

Keywords: Algorithmic advice, Moral residue, Clinical decision support, Responsibility, Accountability, Artificial intelligence

Introduction

Responsibility after the recommendation

Ethical analysis of artificial intelligence in healthcare often asks whether an algorithm should be permitted to influence a clinical decision, whether its recommendations are sufficiently accurate or explainable, and whether a human remains "in the loop." Those questions are necessary, but they leave a further problem insufficiently resolved: what happens morally after an algorithmic recommendation reaches a clinician? At that point, the clinician must do something with the advice. Acceptance, modification, delay, escalation, rejection, and non-engagement are not interchangeable responses, and the fact that a human technically retains the final decision does not by itself establish how responsibility should be allocated.

Recent bioethical analysis supports treating personal moral responsibility as more complex than nominal human control. A clinician may remain the formal decision-maker while the conditions relevant to responsibility—knowledge, influence, control, foreseeability, professional role, and opportunities to act

Access this article online https://smerpub.com/

Received: 12 February 2026; **Accepted:** 24 May 2026

How to cite this article: Zhang W, Hui C, Tan M. Moral Residue After Algorithmic Advice. Asian J Ethics Health Med. 2026;6(1):134-43. https://doi.org/10.51847/z27Fe6XlXk

otherwise—have been altered by algorithmic participation [1]. The difficulty is therefore not simply whether a human pressed the final button, signed the order, or communicated the treatment decision.

Algorithmic decision support can also create an attributability problem: a decision may remain formally attributable to a clinician while its epistemic formation partly depends on a system whose reasoning, training conditions, or limitations the clinician does not fully command [2]. More broadly, digital health technologies can redistribute or complicate moral responsibility across clinicians, organizations, developers, and other actors rather than merely transferring an unchanged responsibility burden from one individual to another [3]. The presence of such redistribution does not entail that no one is responsible. It instead makes the grounds and extent of responsibility questions requiring explicit analysis.

This article argues that clinical value judgment remains morally significant even when algorithmic systems supply useful predictive information [4]. It therefore shifts attention from a binary choice between "human" and "machine" decision-making to the moral structure of responding to advice. Its central proposal is that responsibility after algorithmic advice should be evaluated relationally and prospectively as well as retrospectively. On that account, moral residue may arise not only after negligent reliance or unjustified override, but also after defensible decisions in which morally important uncertainty, conflicting obligations, constrained agency, or unsettled responsibility remains.

Following, overriding, and ignoring algorithmic advice

Reliance on clinical AI is not a single behavior. Experimental evidence indicates that clinicians' reliance on AI-enabled decision support can vary in ways that are more or less appropriately calibrated to the circumstances [5]. Recommendation quality matters as well: correct and incorrect AI recommendations can alter human diagnostic performance, demonstrating why neither automatic acceptance nor systematic resistance can be treated as an ethically neutral default [6]. The relevant question is therefore not whether a clinician "used AI," but how the recommendation entered deliberation and what grounds supported the resulting response.

Reliance should also be distinguished from interpersonal trust. Medical AI can be depended upon instrumentally without possessing the reciprocal moral capacities ordinarily associated with trust between persons [7]. This distinction matters because describing a clinician as having "trusted the algorithm" can obscure the more precise questions of what evidence was available, what limitations were known, how much deference occurred, and whether alternative judgment remained practically usable.

Empirical-ethical work likewise indicates that concerns about decision authority and responsibility are related without being identical [8]. Building on that distinction, this article proposes six response modes. Following means substantially implementing the recommendation. Modifying means incorporating it while materially changing part of the proposed course. Delaying means postponing consequential action while uncertainty remains relevant. Seeking review means obtaining additional professional or technical scrutiny before commitment. Overriding means considering the recommendation and deliberately acting contrary to it. Ignoring means proceeding without materially engaging an available recommendation. These are proposed analytic categories, not validated behavioral classifications and not a hierarchy from ethically preferable to ethically inferior responses.

Table 1 presents "Clinician responses to algorithmic advice and their corresponding responsibility demands" for "Moral Residue after Algorithmic Advice: Responsibility for Following, Overriding, and Ignoring Clinical Decision Support."

Table 1. Clinician responses to algorithmic advice and their corresponding responsibility demands

Response	Justification burden	Documentation needs	Patient communication	Uncertainty	Retrospective scrutiny
Follow	Explain why reliance was reasonable	Record material recommendation and basis for reliance	Explain AI role when decision-relevant	Residual concern about deference or error	Was reliance appropriately calibrated?
Modify	Explain accepted and altered elements	Record material departures and reasoning	Clarify that judgment integrated rather than copied advice	Hybrid authorship may remain unclear	Which contribution shaped the outcome?

Delay	Show why postponement was proportionate	Record uncertainty and reason for delay	Communicate consequences of waiting when relevant	Uncertainty is deliberately preserved	Did delay create avoidable risk?
Seek review	Justify escalation where material	Record referral, consultation, or technical review	Explain additional review when clinically salient	Responsibility may become more distributed	Was escalation available and adequate?
Override	Give reasons for acting contrary to advice	Record disagreement and alternative evidence	Explain material conflict where appropriate	Concern may persist that rejected advice was correct	Was override supportable at the time?
Ignore	Explain why non-engagement was justified	Record reason if advice was clinically material	Disclose omission when relevant to deliberation	Salient information may remain unexamined	Was non-engagement reasonable?

Note: The six response categories and their responsibility-demand mappings are original proposed synthesis. Evidence that reliance can be differently calibrated supports the need for response-sensitive analysis [5]. Evidence on recommendation correctness supports attention to the epistemic consequences of following or resisting advice [6]. The distinction between reliance and trust informs the characterization of engagement with AI [7]. Evidence concerning authority and responsibility supports treating those dimensions separately [8].

The table is not a liability matrix. A demanding justification burden does not imply wrongdoing, and extensive documentation does not make a decision morally superior. The proposed classification instead clarifies what questions become salient once a particular response pathway has been chosen.

Why formal decision authority does not settle responsibility

A common solution to concerns about medical AI is to preserve a human decision-maker and assign that person final authority. This safeguard can be valuable, but it does not answer the moral question by itself. Responsibility gaps in medical AI arise partly because the relevant human and institutional contributions cannot always be captured by locating the final formal decision [9]. A clinician may possess nominal authority while lacking meaningful access to system limitations, organizational permission to disregard recommendations, time for independent verification, or a practically available alternative. Conversely, a clinician who meaningfully understands and controls the decision process may retain substantial responsibility even when algorithmic advice is influential.

Descriptive judgments about responsibility are also sensitive to AI participation. Empirical work indicates that perceived responsibility can change when AI contributes to medical decision-making [10]. Such perceptions matter because they may shape expectations, professional behavior, organizational responses, and blame practices. They are not, however, self-validating moral conclusions. People may attribute too much responsibility to the visible clinician, too little to institutional design, or too much to the technology itself. Normative analysis must therefore ask not merely who is perceived to be responsible but why responsibility would be justified.

Ethical analysis of algorithmic decision-making in healthcare similarly shows that accountability cannot be reduced to preserving a nominal human role [11]. This article consequently proposes that responsibility be assessed through several interacting conditions: causal contribution, epistemic access, foreseeability, meaningful control, professional or institutional role, availability of reasonable alternatives, and capacity to contest or escalate. No single condition is proposed as universally decisive.

This account also separates moral responsibility from legal liability. A professional may have a moral reason to explain or revisit a decision without satisfying the elements of legal fault; an institution may bear a moral responsibility for governance even where a jurisdiction places formal liability elsewhere. Likewise, an adverse outcome does not retrospectively prove that following or overriding an algorithm was irresponsible. Responsibility should be judged in relation to what could reasonably have been known, controlled, and justified when the decision was made, rather than reconstructed solely from whether the outcome happened to be good or bad.

The resulting principle is limited but important: formal authority identifies who is empowered to decide; it does not by itself determine who should answer for how the decision became possible, constrained, informed, or executed. That distinction creates the conceptual space in which a relational account becomes necessary.

Relational responsibility in clinical decision support

A relational account rejects two symmetrical simplifications: that responsibility remains wholly with the clinician because the clinician retains final authority,

and that responsibility shifts away from the clinician because an algorithm influenced judgment. Relational approaches to healthcare AI provide a basis for understanding responsibility as emerging from interdependence among actors rather than from isolated individual agency alone [12]. The proposal developed here extends that insight specifically to the period after algorithmic advice has been received.

Responsible AI in healthcare is already treated in the literature as a multidimensional sociotechnical problem rather than only a matter of model accuracy [13]. Clinical decisions are shaped not only by clinicians and algorithms but also by developers who define system capabilities, organizations that select and configure systems, governance structures that determine oversight, teams that provide review, and patients whose values may determine what counts as an acceptable clinical choice. Responsibility therefore cannot be allocated adequately without examining how these actors structure one another's options.

Healthcare professionals themselves report accountability, transparency, and bias as interacting concerns in AI-assisted decision-making [14]. Such findings should not be universalized beyond their empirical context, but they reinforce the plausibility of treating responsibility as institutionally mediated. A clinician ordered to use a poorly understood system under severe time constraints occupies a different moral position from one who has relevant expertise, access to alternatives, meaningful override authority, and institutional support for review.

This article therefore proposes a role–influence–knowledge–capacity–answerability account. Responsibility should track what an actor was positioned to contribute, what influence that actor exercised over the relevant decision environment, what could reasonably be known, what the actor had practical capacity to change, and what forms of answerability appropriately follow. These dimensions may generate overlapping rather than mutually exclusive responsibilities. A developer may be answerable for known system limitations, an institution for deployment and oversight, and a clinician for how a recommendation was interpreted in the patient's circumstances. None of these responsibilities automatically cancels the others.

Relational responsibility also protects against treating AI as a moral sink into which unresolved responsibility can disappear. Whether an AI system itself qualifies as a moral agent is not necessary to the present argument. The immediate normative task is to identify which human and institutional relationships made the consequential decision possible and which actors possessed morally relevant forms of control, knowledge, or capacity.

Figure 1 synthesizes the proposed responsibility pathways that arise once algorithmic advice enters clinical deliberation, showing how following, modifying, delaying, seeking review, overriding, and ignoring each generate distinct demands for justification, outcome interpretation, and subsequent accountability.

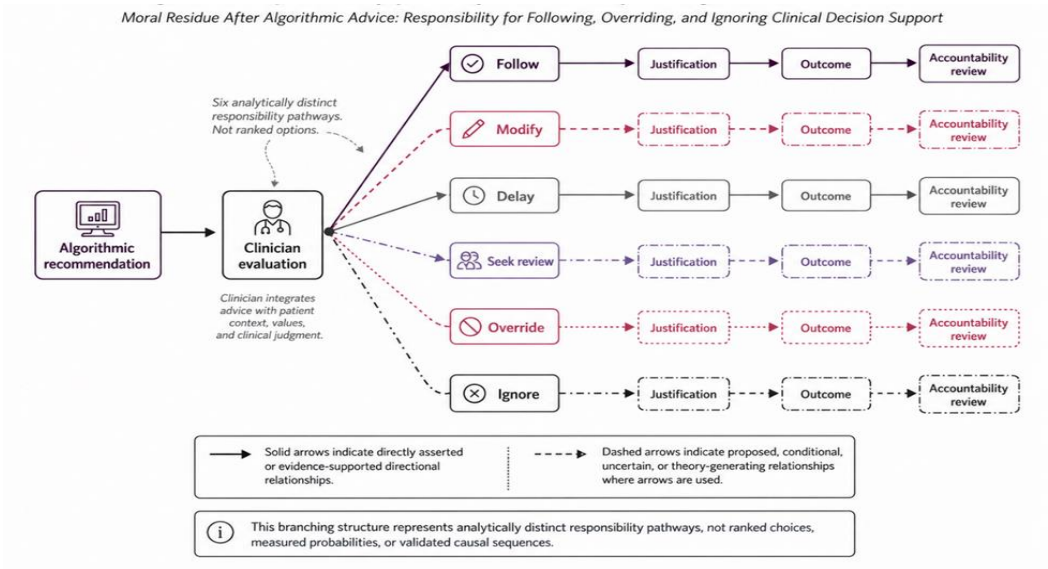

Figure 1. Responsibility pathways after receipt of algorithmic clinical advice

How moral residue arises

Moral residue should not be inferred whenever a clinician experiences discomfort after an AI-supported decision. The more defensible starting point is the literature on morally troubling healthcare work. Qualitative research on machine-learning clinical decision-support development shows that technologically mediated work can generate anticipatory moral distress when participants foresee ethically problematic consequences yet experience constraints on how those concerns can be addressed [15]. This finding establishes plausibility for AI-related moral conflict, but it does not demonstrate an AI-specific residue among clinicians using deployed systems.

Conceptual boundaries therefore matter. Moral injury has been distinguished from moral distress in healthcare scholarship and should not be treated as the inevitable endpoint of a difficult decision [16]. Moral distress itself remains contested and requires greater precision than simply denoting regret, anxiety, dissatisfaction, or professional frustration [17]. Building on these distinctions, this article proposes moral residue after algorithmic advice as a morally salient remainder that persists because significant value conflict, uncertainty about responsibility, constrained choice, or unresolved concern about a decision process or outcome has not been adequately settled.

The proposal deliberately separates residue from proven wrongdoing. A clinician might reasonably follow a recommendation that later proves harmful, defensibly override advice while remaining troubled about uncertainty, or comply with an institutional workflow that leaves little realistic opportunity for independent review. Moral residue can therefore be understood as a remainder of moral involvement, not a retrospective verdict of fault. Its normative importance lies in what remains unsettled: whether the clinician could act on relevant reasons, whether responsibility was fairly distributed, whether the patient's values were adequately represented, and whether the institution acknowledged its own contribution. Empirical work is still needed to determine whether this proposed AI-specific construct is distinguishable from existing measures of moral distress, guilt, regret, burnout, or moral injury.

Prospective and retrospective accountability

Responsibility is often examined after an adverse outcome, when knowledge of what happened can distort judgments about what actors should have done. A more complete account begins prospectively. Expert research on the distribution of responsibility in AI development supports attention to actors' influence and capacity when assigning forward-looking responsibilities [18]. Translated cautiously into clinical decision support, this suggests that institutions, developers, clinical leaders, and users may bear different duties before a particular outcome occurs.

Retrospective accountability is broader than blame. Normative analysis of clinicians' perspectives on black-box AI argues against making physicians automatically absorb all responsibility for tools whose design and opacity they do not control [19]. This does not remove professional duties. It instead requires scrutiny of what the clinician could reasonably know, contest, verify, and change. Responsibility in healthcare may also extend across multiple agents and temporal stages rather than attaching only to the person present at the final decision [20].

This article therefore proposes a two-directional accountability test. Prospectively, each actor should be asked whether they were appropriately positioned to anticipate limitations, maintain competence, create review routes, preserve meaningful disagreement, communicate material uncertainty, and prevent avoidable dependency. Retrospectively, review should reconstruct what information and alternatives were actually available, what influence each actor exercised, why a response pathway was chosen, and what forms of correction or repair are justified. Blame is only one possible result of that inquiry.

The distinction prevents outcome-based moral shortcuts. A poor outcome after reasonable reliance does not establish culpability, while a good outcome after reckless disregard does not erase an irresponsible process. Accountability concerns the quality of participation in a sociotechnical decision, judged against the reasons and capacities available at the relevant time.

Documentation, explanation, and moral repair

If responsibility after algorithmic advice is relational, accountability requires more than identifying a nominal decision-maker. Explainability scholarship identifies several ethical reasons for making AI-supported decisions intelligible while also showing that no single form of explanation is sufficient across all contexts [21]. Documentation can complement explanation by preserving what recommendation was available, what

information shaped its appraisal, what response pathway was chosen, and what unresolved uncertainty remained.

Recommendations for responsible AI-enabled clinical decision support similarly emphasize governance and traceability beyond predictive performance alone [22]. Yet more documentation is not automatically better. Excessive recording can become defensive, burdensome, or performative if clinicians lack meaningful opportunities for review. Multidisciplinary work on explainability further supports tailoring explanation to stakeholder and context rather than assuming maximal technical disclosure is always ethically required [23].

Moral repair is proposed here as a broader process: reconstructing the decision, acknowledging morally significant uncertainty or harm, assigning answerability fairly, enabling review, addressing remediable institutional conditions, and carrying lessons into future decisions. These practices are potential resources for repair, not interventions proven to reduce moral residue.

Repair also has a relational rather than merely intrapersonal object. A clinician may reach personal closure while a patient remains unable to understand why AI advice mattered, or an institution may correct a workflow without acknowledging the burden imposed on staff. Conversely, acknowledgement without structural correction may leave the conditions that generated the conflict unchanged. The proposed account therefore treats repair as potentially operating at clinician, patient, team, and institutional levels. What counts as adequate repair will vary with the kind of unresolved concern: epistemic uncertainty may call for review, avoidable opacity for explanation, unfairly concentrated responsibility for organizational correction, and experienced harm for acknowledgement. None of these responses should be assumed sufficient in isolation.

Table 2 presents "Sources of algorithmic moral residue and opportunities for moral repair" for "Moral Residue after Algorithmic Advice: Responsibility for Following, Overriding, and Ignoring Clinical Decision Support."

Table 2. Sources of algorithmic moral residue and opportunities for moral repair

Source of residue	Residual moral concern	Possible repair practice
Constrained choice	Action did not fully reflect professional judgment	Acknowledge constraint; review whether alternatives were realistically available
Uncertainty	Material doubt remains after action	Document uncertainty; revisit evidence when appropriate
Authority conflict	Formal authority and practical influence diverged	Clarify roles; examine who could contest or alter the decision
Patient harm	Defensible process ended badly	Explain what was known; acknowledge harm without presuming fault
Institutional pressure	Workflow narrowed meaningful choice	Review policy, staffing, escalation, and override conditions
Recommendation disagreement	Clinician and system supported different courses	Preserve reasons for agreement or dissent; seek review when warranted
Unresolved accountability	Contributions remain poorly attributed	Reconstruct actor roles, knowledge, control, and influence
Persistent moral remainder	Concern survives ordinary review	Offer acknowledgement, reflective discussion, learning, and further review

Note: All source-to-repair mappings are original proposed synthesis. Moral-distress literature informs the construct boundary [17]. Explainability evidence informs context-sensitive explanation [21]. Governance recommendations inform traceability [22]. Stakeholder-sensitive accounts inform communication [23]. None establishes that these practices clinically resolve moral residue.

Repeated exposure and cumulative moral residue

A single difficult encounter may resolve quickly; repeated unresolved encounters may not. Recent cross-sectional research treats moral residue as empirically distinguishable and reports associations between moral distress, moral residue, and psychological distress, but it cannot establish temporal causation or an AI-specific mechanism [24]. The article therefore uses accumulation as a hypothesis rather than a demonstrated trajectory.

Repeated algorithmic exposure may also change the environment in which later judgments are formed. Longitudinal observational research in clinical care indicates that algorithmic recommendation tools can alter experiential learning trajectories [25]. Such evidence does not show universal deskilling, nor does it establish

moral residue. It does show why repeated exposure should be considered dynamically: prior advice may influence what clinicians notice, practice, defer to, challenge, or learn from later. Longitudinal evidence from hospital workers likewise shows that moral distress can persist across time and relate to occupational distress, although that evidence arose in a pandemic setting rather than AI-mediated care [26].

This article consequently proposes several qualitative possibilities across repeated decisions: resolution, residual uncertainty, cumulative residue, normalization, moral disengagement, resistance, and moral repair. These are not stages through which clinicians necessarily progress. Normalization may reflect appropriate adaptation or problematic desensitization; resistance may represent justified vigilance or indiscriminate rejection. Longitudinal study is required to distinguish these possibilities.

Figure 2 extends this analysis longitudinally by illustrating how repeated encounters with algorithmic advice may lead to qualitatively different trajectories—including resolution, persistent uncertainty, cumulative moral residue, normalization, moral disengagement, resistance, or moral repair—without implying a fixed sequence or validated causal progression.

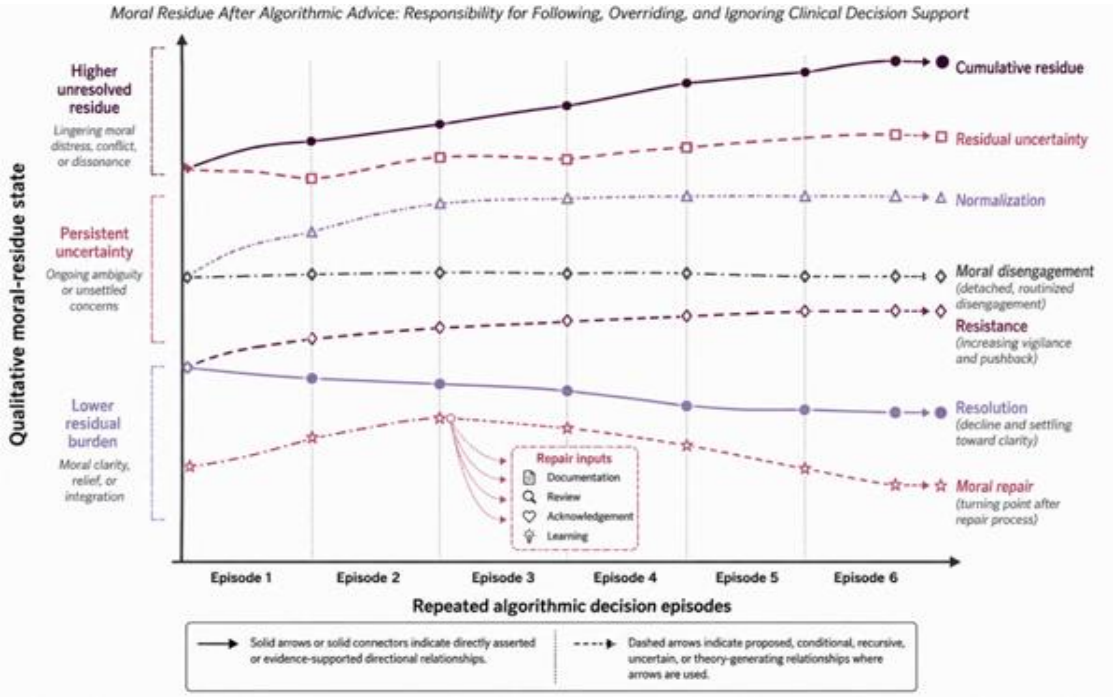

Figure 2. Accumulation and resolution of moral residue across repeated algorithmic decisions

Clinical applications and boundary cases

The proposed account is most useful when a clinical outcome alone gives an incomplete picture of responsibility. Consider an emergency clinician who follows a well-performing recommendation because independent review is impossible, a specialist who overrides advice because patient-specific information conflicts with the model, or a trainee who technically possesses override authority but works in a hierarchy where disagreement is practically costly. These cases differ in epistemic access, meaningful control, alternatives, and institutional support even if each formally ends with a human decision.

The framework also resists treating every departure from AI advice as professional independence. Evidence that reliance can be calibrated in better or worse ways means that following may sometimes be more responsible than overriding [5]. Conversely, a correct recommendation does not eliminate the clinician's obligation to interpret it within the patient's circumstances. Relational responsibility requires examining who shaped the available options, what reasons were accessible, and whether review was realistically possible [12].

Boundary conditions include emergency time pressure, irreversibility of harm, rapidly changing models, trainee status, scarce specialist review, patient disagreement,

opaque systems, resource-limited institutions, and recommendations embedded so deeply in workflow that "advice" becomes functionally difficult to refuse. The proposed analysis is therefore not a universal decision rule. Its practical contribution is diagnostic: it identifies which questions should be asked before moral responsibility, residue, or repair is attributed.

Objections and limitations

One objection is that sufficiently capable AI could make the distinction between advice and authority obsolete. Current evidence does not justify that assumption across clinical tasks: evaluations of large language models identify limitations relevant to autonomous clinical decision-making, while remaining bounded to particular systems and test conditions [27]. The framework must therefore remain revisable as technologies change.

A second objection is that responsibility analysis centered on clinicians could obscure unequal patient risk. Evidence of underdiagnosis bias in chest-radiograph algorithms shows that algorithmic error burdens can differ across underserved patient populations [28]. Equity must consequently constrain judgments about reasonable reliance, institutional governance, and acceptable uncertainty.

A third objection concerns explainability. Not every ethical problem created by clinical AI can be solved by demanding fuller technical explanation; arguments for universal explainability require qualification, including distinctions between opacity and automation [29]. The moral-repair account therefore treats explanation as one resource among documentation, acknowledgement, review, correction, and institutional learning.

More fundamentally, "algorithmic moral residue" may prove unnecessary if existing concepts of moral distress or regret explain the same phenomena. That is a genuine falsification possibility. The proposed construct requires discriminant validation, longitudinal study, cross-professional comparison, patient perspectives, and examination across jurisdictions. Until then, it should function as a normative analytic hypothesis rather than a diagnosis, scale, or established clinical entity.

Conclusion

Algorithmic advice does not end responsibility by leaving a human formally in charge, nor does it automatically transfer responsibility to the system that supplied a recommendation. The morally relevant question is how people and institutions participate in creating, interpreting, constraining, contesting, and reviewing the decision.

This article has proposed a response-sensitive and relational account of that problem. Following, modifying, delaying, seeking review, overriding, and ignoring can each be responsible or irresponsible depending on justification, knowledge, control, alternatives, patient circumstances, and institutional conditions. Moral residue names the possible morally salient remainder when those conditions leave important conflict or responsibility unresolved; it is not evidence of wrongdoing.

The account also distinguishes prospective accountability from retrospective blame and frames documentation, explanation, acknowledgement, review, and learning as possible resources for moral repair. Repeated exposure may resolve, accumulate, normalize, or transform residue, but those trajectories remain hypotheses requiring empirical testing. The contribution is therefore bounded: a conceptual framework for asking better responsibility questions after algorithmic advice, not a validated causal model, liability rule, clinical intervention, or universal prescription.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Smith H, Birchley GM, Ives JCS. Artificial intelligence in clinical decision-making: rethinking personal moral responsibility. Bioethics. 2024;38(1):78-86. doi:10.1111/bioe.13222
2. Zeiser J. Owning Decisions: AI Decision-Support and the Attributability-Gap. Sci Eng Ethics. 2024;30(4):27. doi:10.1007/s11948-024-00485-1
3. Meier E, Rigter T, Schijven MP, van den Hoven M, Bak MAR. The impact of digital health technologies on moral responsibility: a scoping review. Med Health Care Philos. 2025;28(1):17-31. doi:10.1007/s11019-024-10238-3

4. Patil SV, Myers CG, Dai T. Protecting clinical value judgment in the age of AI. npj Digit Med. 2026;9(1):269. doi:10.1038/s41746-026-02561-1
5. Küper A, Lodde GC, Livingstone E, Schadendorf D, Krämer N. Psychological factors influencing appropriate reliance on AI-enabled clinical decision support systems: experimental web-based study among dermatologists. J Med Internet Res. 2025;27:e58660. doi:10.2196/58660
6. Kücking F, Busch DA, Przysucha M, Kutza JO, Hannemann N, Hüsers J, et al. Impact of AI recommendation correctness on diagnostic accuracy in clinical decision-making. Int J Med Inform. 2026;207:106223. doi:10.1016/j.ijmedinf.2025.106223
7. Kerasidou CX, Kerasidou A, Buscher M, Wilkinson S. Before and beyond trust: reliance in medical AI. J Med Ethics. 2022;48(11):852-6. doi:10.1136/medethics-2020-107095
8. Funer F, Liedtke W, Tinnemeyer S, Klausen AD, Schneider D, Zacharias HU, et al. Responsibility and decision-making authority in using clinical decision support systems: an empirical-ethical exploration of German prospective professionals' preferences and concerns. J Med Ethics. 2024;50(1):6-11. doi:10.1136/jme-2022-108814
9. Giubilini A. It Is Not About AI, It's About Humans. Responsibility Gaps and Medical AI. J Bioeth Inq. 2025;22(3):527-37. doi:10.1007/s11673-025-10423-w
10. Krügel S, Ammeling J, Aubreville M, Fritz A, Kießig A, Uhl M. Perceived responsibility in AI-supported medicine. AI Soc. 2025;40(3):1485-95. doi:10.1007/s00146-024-01972-6
11. Grote T, Berens P. On the ethics of algorithmic decision-making in healthcare. J Med Ethics. 2020;46(3):205-11. doi:10.1136/medethics-2019-105586
12. Ferlito B, Segers S, De Proost M, Mertes H. Responsibility Gap(s) Due to the Introduction of AI in Healthcare: An Ubuntu-Inspired Approach. Sci Eng Ethics. 2024;30(4):34. doi:10.1007/s11948-024-00501-4
13. Siala H, Wang Y. SHIFTing artificial intelligence to be responsible in healthcare: A systematic review. Soc Sci Med. 2022;296:114782. doi:10.1016/j.socscimed.2022.114782
14. Nouis SCE, Uren V, Jariwala S. Evaluating accountability, transparency, and bias in AI-assisted healthcare decision-making: a qualitative study of healthcare professionals' perspectives in the UK. BMC Med Ethics. 2025;26(1):89. doi:10.1186/s12910-025-01243-z
15. Whitney C, Preis H, Vargas AR. Anticipatory moral distress in machine learning-based clinical decision support tool development: a qualitative analysis. SSM Qual Res Health. 2025;7:100540. doi:10.1016/j.ssmqr.2025.100540
16. Čartolovni A, Stolt M, Scott PA, Suhonen R. Moral injury in healthcare professionals: A scoping review and discussion. Nurs Ethics. 2021;28(5):590-602. doi:10.1177/0969733020966776
17. Morley G, Ives J, Bradbury-Jones C, Irvine F. What is 'moral distress'? A narrative synthesis of the literature. Nurs Ethics. 2019;26(3):646-62. doi:10.1177/0969733017724354
18. Hedlund M, Persson E. Distribution of responsibility for AI development: expert views. AI Soc. 2025;40(5):4051-63. doi:10.1007/s00146-024-02167-9
19. Lang BH, Kostick-Quenet K, Smith JN, Hurley M, Dexter R, Blumenthal-Barby J. Should Physicians Take the Rap? Normative Analysis of Clinician Perspectives on Responsible Use of 'Black Box' AI Tools. AJOB Empir Bioeth. 2025;16(4):201-12. doi:10.1080/23294515.2025.2497755
20. Brown RCH, Savulescu J. Responsibility in healthcare across time and agents. J Med Ethics. 2019;45(10):636-44. doi:10.1136/medethics-2019-105382
21. Freyer N, Groß D, Lipprandt M. The ethical requirement of explainability for AI-DSS in healthcare: a systematic review of reasons. BMC Med Ethics. 2024;25(1):104. doi:10.1186/s12910-024-01103-2
22. Labkoff S, Oladimeji B, Kannry J, Solomonides A, Leftwich R, Koski E, et al. Toward a responsible future: recommendations for AI-enabled clinical decision support. J Am Med Inform Assoc. 2024;31(11):2730-9. doi:10.1093/jamia/ocae209
23. Amann J, Blasimme A, Vayena E, Frey D, Madai VI; Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20(1):310. doi:10.1186/s12911-020-01332-6
24. Gustavsson ME, Juth N, von Schreeb J, Arnberg FK. Moral distress, moral residue, and associations with

psychological distress: a cross-sectional study. Eur J Psychotraumatol. 2025;16(1):2512677. doi:10.1080/20008066.2025.2512677

25. Sundaresan S, Guler I. Algorithmic Recommendation Tools and Experiential Learning in Clinical Care. Organ Sci. 2025;36(5):1786-802. doi:10.1287/orsc.2022.16738
26. Maunder RG, Heeney ND, Greenberg RA, Jeffs LP, Wiesenfeld LA, Johnstone J, et al. The relationship between moral distress, burnout, and considering leaving a hospital job during the COVID-19 pandemic: a longitudinal survey. BMC Nurs. 2023;22(1):243. doi:10.1186/s12912-023-01407-5
27. Hager P, Jungmann F, Holland R, Bhagat K, Hubrecht I, Knauer M, et al. Evaluation and mitigation of the limitations of large language models in clinical decision-making. Nat Med. 2024;30(9):2613-22. doi:10.1038/s41591-024-03097-1
28. Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nat Med. 2021;27(12):2176-82. doi:10.1038/s41591-021-01595-0
29. Blackman J, Veerapen R. On the practical, ethical, and legal necessity of clinical Artificial Intelligence explainability: an examination of key arguments. BMC Med Inform Decis Mak. 2025;25(1):111. doi:10.1186/s12911-025-02891-2