Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

When Caution Blocks Learning: Justifying the Non-Adoption of Medical AI

Tetsuro Ishida

teturoisida@yahoo.co.jp

Sapporo Royal Hospital

Research Article

Keywords: artificial intelligence, evidence generation, precaution, institutional responsibility, learning healthcare systems, non-adoption

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11427381/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

A healthcare institution may reasonably decline to introduce an artificial intelligence (AI) system because evidence of its local safety and value is insufficient. A different question arises when that institution also controls, and declines to permit, a proportionate means of addressing the uncertainty. This article examines the justification of that second decision. It builds on established accounts of AI underuse, adoption-dependent learning, and institutional duties within learning healthcare systems; the dependence of evidence on use is not presented as a new discovery. Through a hypothetical medication-review example, I argue that a reason sufficient to withhold routine clinical use does not automatically justify closing a decision-relevant learning pathway. Where an institution has a relevant care-governance responsibility, controls necessary access, and can make a fair contribution to ethically permissible inquiry, refusing that contribution requires an action-specific justification. This requirement does not establish a duty to adopt the system, compel patient participation, or undertake every informative study. Nor does institutional responsibility for missing evidence make an unvalidated system safer. An institution can remain right to withhold clinical use while being answerable for having obstructed reasonable inquiry. The practical implication is to evaluate the present deployment decision and the maintenance of its evidential basis separately, with responsibility for restoring justified learning opportunities rather than automatically relaxing safety standards.

1 Introduction

A hospital committee declines to introduce a clinical decision-support system based on artificial intelligence (AI). Its reason is familiar and potentially decisive: the available evidence does not establish that the system will benefit this hospital's patients without unacceptable risks. The committee also declines a proposal to evaluate the system under conditions that would not authorize its routine use. When asked why, it repeats that the evidence is insufficient. Has one defensible reason answered two different questions?

The issue is not simply that non-adoption can have costs. Pagallo and colleagues already examine the opportunity costs of underusing AI in healthcare [1]. Nor is liability necessarily an obstacle to improvement. Maliha and colleagues describe competing incentives, including reasons for health systems both to adopt useful technologies and to avoid unsafe ones [2]. It would therefore be premature to infer either that faster adoption is generally desirable or that institutional hesitation is generally self-protective. The question here concerns the justification of a particular refusal of inquiry, even when the decision against routine use is sound.

I defend a limited claim. When an institution bears a relevant responsibility for care quality, controls access needed for a decision-relevant investigation, and could fairly contribute to an ethically permissible and proportionate learning pathway, its refusal requires justification directed to that pathway. Insufficient evidence for ordinary deployment may be part of the justification, but cannot settle the matter merely by being repeated. The institution must explain why the proposed way of learning is

unacceptable, unnecessary, or not a fair use of its resources. This is not a requirement to invent an unrelated reason: the same uncertainty may justify both refusals if its significance for each activity is established.

The scope is institutional governance of a defined clinical application, not a universal obligation to promote AI. The analysis concerns prospective evidence generation rather than retrospective allocation of damages. Its method is normative comparison using an expressly hypothetical case. It neither estimates how often institutions obstruct learning nor establishes the clinical merits of a current product. The contribution is a specification of existing learning obligations at a particular decision point: whether justified non-adoption also justifies maintaining the uncertainty on which non-adoption depends.

2 What the existing literature already establishes

The dependence of knowledge on technological use is established prior work. Gans models how assumptions about learning affect optimal AI adoption: learning through deployment and learning independently of deployment can support different choices, with reversibility also affecting the result [3]. The analysis directly rules out presenting the broad sequence of non-adoption, limited learning, and continued uncertainty as a newly discovered mechanism. Cecchi and colleagues explicitly connect the Collingridge dilemma with healthcare AI and medico-legal assessment [4]. The present argument does not claim priority for applying that dilemma to medicine.

The normative antecedents are equally important. Faden and colleagues ground obligations of continuous learning in healthcare institutions' responsibilities and distinctive access to clinical practice and information, while protecting patients' interests [5]. Hollestelle and colleagues address the transition to operational learning healthcare systems through requirements concerning public benefit, proportionality, equity, engagement, and sustainability [6]. Institutional learning duties and their ethical limits are therefore premises of this article, not its proposed discoveries. The further question is how such duties constrain the reasons offered for closing a particular route to evidence.

Established coverage frameworks also resist a binary choice between unrestricted access and no investigation. Walker and colleagues distinguish coverage decisions connected to further research, including coverage with evidence development and access only within research [7]. Bracic, Callier, and Price identify reinforcing exclusion across medicine, data, and AI [8]. These contributions preclude treating either research-linked access or self-reinforcing informational disadvantage as novel. They also make a more discriminating question possible: what is owed by an institution that invokes an evidential deficit while controlling a reasonable means of addressing it?

A nearby legal question concerns when non-use itself falls below an appropriate standard. Berz argues that superior statistical performance alone does not establish a general duty to use diagnostic AI, emphasizing institutional and epistemic conditions [9]. Her analysis also addresses evidential difficulties and responsibilities associated with opacity. The question developed here is different: whether an

institution should enable the generation of evidence before such a deployment obligation is established. It need not follow that non-use is negligent, or that a specific patient would have benefited from use.

These literatures substantially narrow the contribution. The argument is not that precaution has costs, that healthcare should learn, or that institutions should explain themselves. It identifies the object of explanation when one institution controls both a deployment decision and access to learning about it. The comparison is targeted rather than systematic.

3 A refusal that leaves the question unanswered

Consider a hospital evaluating a system that flags possible discrepancies between medication lists at admission. Flags would be reviewed by a pharmacist; the system would neither prescribe nor change a medication order. The hospital continues its existing medication-reconciliation service. External evaluations suggest that further investigation may be worthwhile, but provide insufficient information about the hospital's frail patients taking multiple medicines, its record formats, and the workload created by false alerts. These are stipulated features of a hypothetical proposal, not findings about an existing device.

The committee reasonably refuses routine deployment. A research team then proposes a bounded initial evaluation: the system would process a defined set of records, its outputs would be concealed from treating staff, and independent reviewers would compare its flags with a prespecified reference assessment. The study would examine whether the system identifies the relevant discrepancies in this setting and whether further investigation is warranted. It would not establish how pharmacists respond to visible alerts or whether patients' outcomes improve. Those questions would require a different design and a separate decision about permissible clinical exposure.

Suppose adequate study funding has been offered, the proposed workload can be accommodated without displacing a higher-priority project, and a suitable review process can assess data access and other risks. These assumptions do not certify every study called silent evaluation as safe. Record processing can raise privacy and security concerns, independent review takes work, and preparation can disrupt care. They make it possible to examine a case in which a limited pathway is genuinely available rather than a vendor's unfunded aspiration.

The institution nevertheless refuses even to assess the study on its own terms. Its stated rule is that no evaluation involving the hospital's records will be considered until the system is sufficiently validated for routine local use. The team explains that the requested local evidence is what the study is intended to help produce. The response repeats the validation requirement. No account is offered of a specific data risk, methodological defect, competing priority, or feasible source of equivalent evidence elsewhere.

The problem is not that the hospital must believe the system will work. A finding that it performs poorly would also be useful: it could support a more definite rejection and prevent resources being spent on an unsuitable tool. Nor does the study need to eliminate every uncertainty. It must have a reasonable

prospect of informing a live decision or determining whether a next stage is warranted. A small, unrepresentative demonstration that could not do either would not satisfy the example.

The contrast with ordinary deployment is substantive rather than terminological. Concealed outputs cannot establish the consequences of clinicians relying on them. Early-stage clinical evaluation is already a recognized subject of the DECIDE-AI reporting guideline [10], but a reporting guideline is not permission to expose patients to risk. Each proposed activity needs an appropriate design, authorization, and assessment. The error in the example is importing the evidential requirement for one activity into another without explaining why the activities should be governed by the same threshold.

4 Why closing a learning pathway needs its own justification

The case might appear to establish only a logical distinction. Permission to refrain from one activity is not permission to refrain from everything associated with it. That observation is insufficient. Hospitals routinely decline useful proposals without wronging their proponents. To explain why an additional justification is owed here, the argument needs a substantive account of the institution's role, its control, and the burdens of the proposed contribution.

The relevant starting point is a responsibility for the quality and fair organization of care. An institution exercising authority over which tools may enter its service is not acting merely as a private purchaser choosing among conveniences. It makes decisions that affect patients' opportunities for appropriate care and the work of professionals within the service. Those decisions may properly protect patients from uncertainty. But where the institution also controls access needed to assess the options, its responsibility cannot be exhausted by identifying uncertainty. It must consider whether that uncertainty can reasonably be addressed within its proper role.

This argument specifies the institutional learning obligations already defended in learning-healthcare ethics [5, 6]. It does not infer a duty simply from possessing data. A hospital may hold records that would interest many researchers without owing each of them access. The duty becomes more specific when an investigation concerns a material question within the hospital's own care-governance remit, the hospital's cooperation is needed, and participation is a fair and proportionate contribution. The request then concerns how the institution exercises a responsibility it already has, rather than a general demand to assist technological progress.

Control matters because it determines which response is practically available. If a committee requires local evidence but can neither authorize access nor obtain another body's cooperation, the immediate responsibility may be to refer the issue to an appropriate authority. If the hospital can grant access, refusing on the ground that no one has produced the inaccessible evidence leaves its own decision out of the explanation. The evidence gap is real, but its persistence is not wholly external to the institution. An account of responsible governance should include the decisions that maintain it.

The normative concern is not inconsistency in belief. A committee can consistently believe that a tool is unvalidated and decline a study of it. The concern is an incomplete justification for exercising institutional power. Patients and clinicians affected by the decision have a claim to reasons that address available ways of resolving a material question, subject to fair limits. They need not possess an entitlement to the technology for that claim to arise. An institution can owe a fair assessment of a proposal without owing the outcome its proponents desire.

Decision relevance limits the claim. A pathway should target an uncertainty that could affect a practicable choice, not generate information merely because information is desirable. Evidence about technical discrimination may be insufficient when the only live objection concerns an unmanageable staffing burden. Conversely, a workflow study cannot compensate for an unresolved data-integrity defect that prevents meaningful interpretation. The institution may reasonably reject a proposed study that addresses the wrong question. What it should not do is leave the relevant question unspecified while invoking its unanswered status indefinitely.

Proportionality and fair allocation provide further limits. The possibility of learning does not override unacceptable risks, patient refusal, or urgent clinical priorities. Even an ethically permissible study may rank below other permissible work. A hospital with limited research capacity can favor a project with greater expected value, including a non-AI intervention. It should then justify its choice as a priority decision, rather than suggest that the absent evidence alone prohibits the rejected investigation. Clarifying the reason makes its merits open to assessment without presuming that the reason is invalid.

These considerations support a conditional requirement. Where institutional responsibility, relevant control, decision relevance, and a fair opportunity for permissible inquiry coincide, the institution should either enable its reasonable contribution or provide a justification for declining it. The contribution might be a bounded study, access under appropriate safeguards, participation in a coordinated evaluation, or support for an adequate alternative. It need not be routine use, sole sponsorship, or endorsement of a vendor. The requirement concerns a defensible response to the learning opportunity, not a predetermined technological destination.

A justification can draw on the same concern that defeated deployment. Suppose uncertainty includes a serious risk that processing the records would expose identifiable information beyond permissible control. That concern may defeat both ordinary use and the proposed study. Alternatively, unknown device behavior may make even a small live evaluation unacceptably dangerous. The argument does not demand different facts for each refusal. It demands an explanation of why the facts justify each refusal, taking account of differences in exposure, safeguards, reversibility, and purpose.

The institutional contribution can also be discharged without local testing. A credible external study may address the relevant population and setting well enough to inform the hospital's decision. Joining it, supporting access through a suitable network, or waiting for a justified period can then be preferable. However, saying that somebody else might eventually supply evidence is not equivalent to identifying a credible pathway. A reasoned decision should distinguish justified reliance on coordinated inquiry from

an expectation that others will bear all the work while every institution reserves an unrestricted option to wait.

5 Present restraint and responsibility for continued uncertainty

The most important implication is that two evaluations can diverge. One asks whether the system should be used now, given the evidence actually available. The other asks whether the institution has exercised its control over evidence generation responsibly. A negative answer to the second does not require a positive answer to the first. The current evidential threshold can remain unmet even after an institution has acted wrongly in obstructing an earlier, justified opportunity to investigate.

Return to the hospital after a year. In one possible history, it supported a well-designed evaluation, but recruitment problems outside its reasonable control left the results inconclusive. In another, it prevented the evaluation without an adequate reason. Suppose the deployment-relevant evidence available at review remains materially similar. Continued non-adoption may be appropriate in both histories. Yet the institution's answer to why the question remains unresolved differs. Conscientious but unsuccessful inquiry and avoidable obstruction are not equivalent exercises of its responsibilities.

This distinction does not make evidence true or false according to its origin. Information missing because of an institutional failure is still missing. It would be a mistake to treat obstruction as proof that the system is safe, or to assume that the obstructed study would have shown benefit. The criticism concerns a foreclosed opportunity to make a better-grounded decision. Its force can survive the possibility that the system would have failed evaluation. Indeed, responsibly identifying failure is one purpose of enabling inquiry.

For the same reason, the appropriate response is not automatic authorization. Lowering the safety threshold because the hospital was responsible for missing evidence would make patients bear the consequences of correcting an institutional failure. The first remedial question is instead what reasonable learning opportunity can now be restored. That may require reconsidering access, redesigning a protocol, arranging cooperation, or explaining why the opportunity is no longer proportionate. Lost time cannot be reversed, and a new study may itself no longer be worthwhile. Responsibility for the earlier decision can remain even then.

Answerability should also be distinguished from culpability and legal liability. An institution may owe an account and a change of procedure without any individual having acted negligently. A committee member may have followed a rule they lacked authority to alter; responsibility for revising that rule may lie elsewhere. Conversely, institutional complexity does not mean that no one can reasonably be assigned the work of reconsideration. The relevant question is who controls the decision or can secure its review, not who can most conveniently absorb blame after an adverse outcome.

Liability anxiety is one possible reason for closing a pathway, but the argument does not depend on diagnosing that motive. Budget arrangements, fragmented authority, or a mistaken interpretation of

safety policy could produce the same pattern. Even sincere concern for patients can sustain it. A satisfactory response must therefore address the institution's reasons and options, rather than infer bad faith from non-adoption. The proposed assessment concerns how a learning refusal is justified; establishing the actual causes and prevalence of such refusals would require empirical research.

6 Limits, objections, and institutional practice

A first objection is that this approach uses future patients' possible benefits to make present patients research resources. That inference is not warranted. Institutional responsibility for enabling inquiry does not establish an individual's duty to enter a trial or authorize every use of their information. Applicable consent, privacy, and research protections remain independent constraints. If an acceptable study cannot recruit enough willing participants, the institution may have fulfilled its responsibilities without obtaining decisive evidence. It should not respond by making access to ordinary care conditional on participation.

A second objection is that permission to investigate is already governed by research ethics, making the argument redundant. Ethical review is necessary where applicable, but does not settle every institutional decision about resources, access, and participation. An ethically permissible study is not automatically one that this institution must host. Conversely, declining routine deployment does not automatically establish that reviewing or hosting it would be wrong. The contribution concerns the relation between those institutional decisions, not a replacement for ethics review or a new approval authority.

A third objection concerns administrative burden. Requiring a fresh inquiry into every rejected AI proposal could consume the resources the argument aims to steward. The requirement should therefore be proportionate at the level of review as well as at the level of research. Institutions may use transparent screening criteria, shared evaluations, and priority-setting procedures. A poorly specified commercial proposal need not trigger an elaborate committee process. The stronger case concerns a defined, material uncertainty and a credible request that meets the conditions described above. Fair rejection can be brief when its basis is clear.

A related concern is endless provisionality. Some questions remain too costly, too dangerous, or too unimportant to investigate, and some technologies become obsolete before evidence can be produced. An institution may close consideration rather than promise perpetual reassessment. What matters is the basis of closure. A clear conclusion that the expected value of further investigation is too low differs from repeating a demand for evidence while declining every reasonable means of obtaining it. A commitment to learning includes the possibility of learning enough to stop, and the possibility of justifiably deciding not to seek more.

Application also depends on what the evidence could transfer to. A study of one model version, patient group, or workflow may not resolve questions about another. The proposed hospital should specify what the initial medication-discrepancy evaluation would establish and what would remain undecided. Concealed-output testing cannot answer how visible recommendations affect clinical judgment; a later

live study would require its own justification. It would be inappropriate to use success in an easier evaluation as permission for a materially different activity. Separating learning decisions should protect against that expansion, not facilitate it.

Recent work in this journal treats health-data access as a calibrated governance practice rather than unrestricted openness [11]. That orientation is compatible with refusing access on adequate grounds. The present claim identifies a particular relation to scrutinize: the unresolved deployment decision cannot, by itself, settle whether to close a pathway designed to inform it.

A practical decision record can expose the relevant reasoning without becoming a new compliance instrument. It should identify the uncertainty doing work in the deployment decision, the investigation or alternative source that might address it, the institution's control over that route, and the principal reasons for permitting or refusing its contribution. Where consideration is deferred rather than closed, the record should specify a meaningful basis for review, such as completion of a relevant study or resolution of a named constraint. A calendar date alone does not produce evidence, and expiry of a review period should not automatically authorize deployment.

The hospital example illustrates why an outcome measure of adoption would misdescribe success. Responsible governance might approve the initial study and then reject the system after unfavorable results. It might prefer a non-AI improvement or participate in external research instead. The important contrast is between a decision whose evidential basis has been responsibly maintained and one that cites uncertainty without addressing its own avoidable contribution to that uncertainty. The account supplies reasons for investigating that contrast, not a validated measure of institutional performance.

7 Conclusion

Withholding routine medical AI use and closing a pathway to evidence are distinct exercises of institutional authority. A reason sufficient for the first is not automatically sufficient for the second. Where relevant responsibility, control, decision relevance, and a fair opportunity for permissible inquiry coincide, declining a contribution requires a justification directed to that inquiry. This specifies established learning obligations without establishing a general duty to adopt AI.

The resulting judgment can be demanding without being technologically permissive. An institution may be right not to use a system today and still be answerable for having obstructed reasonable learning about it. That failure does not fill the evidence gap or license unsafe care. It changes what the institution must explain and, where feasible, repair. The appropriate response is to reopen justified opportunities for inquiry, not to make patients carry the risk of treating missing evidence as if it had already been supplied.

Use of artificial intelligence

ChatGPT (OpenAI) supported literature discovery, drawing graphs, translation and language revision.

Declarations

Author Contribution

T.I. conducted the literature review, drafted the manuscript, revised it, and finalized it.

References

1. Pagallo, U., O'Sullivan, S., Nevejans, N., Holzinger, A., Friebe, M., Jeanquartier, F., Jean-Quartier, C., & Miernik, A. (2024). The underuse of AI in the health sector: Opportunity costs, success stories, risks and recommendations. *Health and Technology, 14*, 1–14. https://doi.org/10.1007/s12553-023-00806-7
2. Maliha, G., Gerke, S., Cohen, I. G., & Parikh, R. B. (2021). Artificial intelligence and liability in medicine: Balancing safety and innovation. *The Milbank Quarterly, 99*(3), 629–647. https://doi.org/10.1111/1468-0009.12504
3. Gans, J. S. (2025). How learning about harms impacts the optimal rate of artificial intelligence adoption. *Economic Policy, 40*(121), 199–219. https://doi.org/10.1093/epolic/eiae053
4. Cecchi, R., Haja, T. M., Calabrò, F., Fasterholdt, I., & Rasmussen, B. S. B. (2024). Artificial intelligence in healthcare: Why not apply the medico-legal method starting with the Collingridge dilemma? *International Journal of Legal Medicine, 138*(3), 1173–1178. https://doi.org/10.1007/s00414-023-03152-5
5. Faden, R. R., Kass, N. E., Goodman, S. N., Pronovost, P., Tunis, S., & Beauchamp, T. L. (2013). An ethics framework for a learning health care system: A departure from traditional research ethics and clinical ethics. *Hastings Center Report, 43*(s1), S16–S27. https://doi.org/10.1002/hast.134
6. Hollestelle, M. J., van der Graaf, R., Sturkenboom, M. C. J. M., & van Delden, J. J. M. (2024). An ethics framework for the transition to an operational learning healthcare system. *Learning Health Systems, 8*(3), Article e10414. https://doi.org/10.1002/lrh2.10414
7. Walker, S., Sculpher, M., Claxton, K., & Palmer, S. (2012). Coverage with evidence development, only in research, risk sharing, or patient access scheme? A framework for coverage decisions. *Value in Health, 15*(3), 570–579. https://doi.org/10.1016/j.jval.2011.12.013
8. Bracic, A., Callier, S. L., & Price, W. N., II. (2022). Exclusion cycles: Reinforcing disparities in medicine. *Science, 377*(6611), 1158–1160. https://doi.org/10.1126/science.abo2788
9. Berz, A. S. (2026). Medical negligence in the age of statistically superior AI. *Medical Law Review, 34*(1), Article fwag007. https://doi.org/10.1093/medlaw/fwag007
10. Vasey, B., Nagendran, M., Campbell, B., Clifton, D. A., Collins, G. S., Denaxas, S., Denniston, A. K., Faes, L., Geerts, B., Ibrahim, M., Liu, X., Mateen, B. A., Mathur, P., McCradden, M. D., Morgan, L., Ordish, J., Rogers, C., Saria, S., Ting, D. S. W., … DECIDE-AI expert group. (2022). Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nature Medicine, 28*, 924–933. https://doi.org/10.1038/s41591-022-01772-9

11. Msheik, A., Peralta, R., Al Mokdad, Z., Alsaad, N., Mohammed, K., & Al-Rumaihi, G. (2026). Ethical governance of open data across biomedical research, healthcare, and public health: Privacy, equity, trust, and controlled access. *Health Care Analysis*. Advance online publication. https://doi.org/10.1007/s10728-026-00598-7