International Journal of Scientific Modern Research and Technology
(Volume: 24, Issue: 03, Number: 06) Paper ID: IJSMRT - 26090306
ISSN: 2582-8150

Human Judgment and Ethical Responsibility in Artificial Intelligence Decision-Making: A Theravāda Buddhist Philosophical Perspective

Yen Sokne
Ph.D. Research Scholar
Department of Philosophy, University of Delhi, 110007, India

Abstract. Artificial intelligence can support consequential decisions without supplying the ethical justification for acting on its recommendations. This article examines what makes human judgment responsible in AI-assisted decision-making through a Theravāda Buddhist philosophical perspective. It combines interpretive readings of selected Pāli discourses and monastic disciplinary rules with contemporary research on meaningful human control and AI governance. The analysis distinguishes intention, epistemic care, moral restraint, and procedural answerability. Intention helps explain the moral orientation of human action, but benevolent purposes do not excuse inadequate verification or preventable harm. Wisdom and disciplined conduct support careful assessment of evidence, truthful communication of uncertainty, and correction of mistaken decisions. Monastic dispute procedures offer a limited comparative resource for examining accountable judgment rather than a ready-made regulatory model. Two hypothetical cases involving scholarship screening and recruitment illustrate how these distinctions can inform institutional practice. The article proposes that responsible oversight requires competent reviewers, effective authority to intervene, documented reasons, and accessible correction mechanisms. Its contribution is a normative framework linking the cultivation of judgment to the institutional conditions that enable it. The framework is conceptual and requires empirical evaluation before claims about its effectiveness can be made.

Keywords: Artificial intelligence; Theravāda Buddhism; human judgment; ethical responsibility; meaningful human control; moral agency.

How to cite this article: Yen Sokne. (2026). Human Judgment and Ethical Responsibility in Artificial Intelligence Decision-Making: A Theravāda Buddhist Philosophical Perspective, International Journal of Scientific Modern Research and Technology (IJSMRT), ISSN: 2582-8150, Volume-24, Issue-03, Number 06, Sep-2026, pp. 45-51, URL: https://www.ijsmrt.com/wp-content/uploads/2026/11/IJSMRT-26090306.pdf

Copyright © 2026 by author (s) and International Journal of Scientific Modern Research and Technology Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0)

(http://creativecommons.org/licenses/by/4.0/)

IJSMRT-26090306

I. INTRODUCTION

AI-assisted decisions create a philosophical problem that cannot be resolved by predictive performance alone. A recommendation may accurately estimate an outcome while leaving unanswered whether the institution should pursue that outcome, whether its criteria are defensible, and whether an affected person has received fair consideration. Human participation also leaves these questions open. An employee who confirms a recommendation without understanding its evidential basis may be formally involved while exercising little substantive judgment. The relevant question is therefore what conditions make human judgment ethically responsible when an AI system participates in a consequential decision.

Contemporary research on meaningful human control offers one starting point. Santoni de Sio and van den Hoven relate control to responsiveness to relevant human reasons and facts, together with an appropriate connection to responsible human agents [1]. This account directs attention beyond the presence of an approval button. For a human decision to matter ethically, the surrounding arrangement must make informed intervention possible. The present article approaches that problem through a different philosophical vocabulary: intention, wisdom, moral

discipline, and procedures for addressing disputed conduct in Theravāda Buddhist sources.

This comparison concerns human beings and institutions that design, deploy, supervise, or rely on AI. It does not establish whether an artificial system possesses consciousness or moral agency. Nor does it equate a software objective with Buddhist intention. Instead, it asks how selected Buddhist accounts of action and judgment can illuminate responsibilities that remain with people when technical recommendations influence decisions. References to monastic regulations are interpretive parallels, not statements that ancient norms immediately apply to modern settings.

In this article we argue that responsible judgment cannot be divorced from the institutional capacity we create and the dispositions we cultivate. A beneficial purpose must be accompanied by attention to evidence, restraint in communication and action, and procedures through which errors can be challenged and corrected. These requirements apply differently to a system designer, a deploying organization, and a case reviewer. Treating the final reviewer as the sole responsible person obscures decisions made earlier about objectives, data, thresholds, and working conditions.

The discussion develops a conceptual model, reviews related scholarships, explains its comparative method, and presents analytical findings through two hypothetical cases. It contributes an argument about justified human reliance on AI rather than a technical detection method or an empirical measurement of Buddhist ethical practice. Its central proposition is that neither benevolent intention nor nominal human approval is sufficient for accountable judgment.

II. SYSTEM MODEL

The system considered here is a sociotechnical decision arrangement containing an institution, an AI recommendation process, a human reviewer, and people affected by the resulting decision. The institution specifies the purpose and criteria of a task. The technical process transforms selected information into a classification, ranking, or recommendation. A reviewer interprets the output and decides what action, if any, is warranted. Affected people may supply information, receive an explanation, or seek reconsideration. These relationships are analytically distinct even when one person occupies several roles.

The model distinguishes prediction, justification, and authorization. Prediction concerns an estimated outcome or category. Justification concerns reasons for treating the estimate as relevant and for acting on it. Authorization concerns who is entitled and able to make or revise the decision. An institution can confuse these levels by allowing a high model score to substitute for a defensible selection criterion. It can also assign authorization to a reviewer while withholding the information needed for justification. In either case, responsibility cannot be assessed solely by identifying who clicked the final confirmation.

Four dimensions organize the proposed model. Purpose concerns the ends and motivations shaping human conduct. Epistemic care concerns whether the available information supports the proposed inference in its actual context. Moral restraint concerns whether people avoid harmful practices and communicate limitations truthfully. Procedural answerability concerns whether reasons, decisions, and errors can be examined and corrected by appropriately empowered people. Figure 1 presents these dimensions as a review sequence. The sequence is an original proposal, not an experimentally validated workflow.

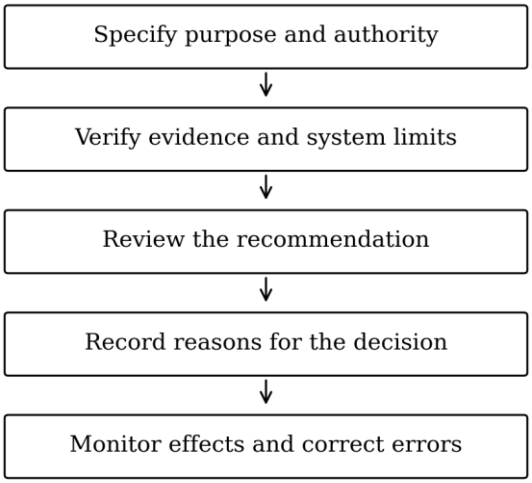

Fig. 1. Proposed sequence for responsible human review (author's conceptual framework).

The dimensions interact. A reviewer may sincerely intend to help applicants but still rely on obsolete records. Accurate records may support a recommendation without justifying an institution's exclusionary objective. A detailed explanation may reveal a mistake while an inaccessible appeal process prevents correction. Ethical responsibility therefore concerns the relationships between these conditions,

rather than the independent completion of four administrative checks.

The degree of review should depend on the potential consequences, the reliability of the evidence, and the ability to reverse a decision. This proposal does not require manual examination of every routine output. It requires institutions to explain why their arrangements provide adequate scrutiny for the decisions being made. Where serious consequences coincide with uncertain evidence or ineffective intervention, the organization has reason to redesign the arrangement or refrain from using it.

III. PREVIOUS WORK

Santoni de Sio and van den Hoven develop meaningful human control through tracking and tracing conditions [1]. Tracking concerns a system's responsiveness to relevant moral reasons and circumstances. Tracing connects system behavior with human agents who have an appropriate understanding of their role. Their framework helps distinguish morally significant control from physical proximity to a technical process. This article draws on that distinction while asking what habits and institutional supports make responsible responsiveness possible.

Cavalcante Siebert and colleagues translate meaningful human control into actionable properties for system development [2]. Their account includes a specified morally relevant domain, compatible understandings of the system, responsibility proportionate to control, and connections between AI actions and responsible humans. These properties support the present distinction between assigning responsibility and enabling its exercise. A reviewer cannot discharge a duty merely because an organizational document names that person as accountable.

Green and Chen's account of algorithm-in-the-loop decision-making centers human decisions within the combined arrangement and reports limits identified through two experiments [3]. Its significance here is methodological: the performance of an algorithm alone does not establish the quality of decisions produced when people use it. This article consequently avoids if human review necessarily improves accuracy or fairness. Those effects require evaluation in the relevant setting.

The NIST AI Risk Management Framework 1.0 identifies governance, mapping, measurement, and management as connected functions [4]. It requires specified duties, documented system restrictions, and provisions for supervision. The oversight of humans, responsibility, transparency, and chances to contest significant judgments is addressed in the Recommendation of UNESCO on the Ethics of Artificial Intelligence [5]. They provide modern institutional reference points. They do not demonstrate that the proposed Buddhist framework is effective; they do serve to pinpoint governance difficulties that its philosophical reasoning can address.

Theravāda sources approach responsibility through the training of persons within a wider ethical and liberative path. The Nibbedhika Sutta focuses on intention in its description of kamma [6]. The Mahāvedalla Sutta deals with wisdom, right view and conditions favorable for knowing [7]. The instruction to Rāhula makes reflection on harmful consequences relevant before, during, and after action [8]. Selected monastic rules add differentiated standards for truthful speech, learning disciplinary obligations, and settling disputes [9]. These sources address distinct subjects and cannot be compressed into a single ancient theory of AI governance.

The present contribution lies in articulating the relationship between moral cultivation and the practical capacity to respond to evidence. The contemporary ideas of oversight illuminate the institutional contexts, and the Buddhist parallel explores the quality of human attention and behavior under those conditions. This claim is a targeted philosophical contribution. It is not that Buddhist literature predicted AI, nor that this is the first study relating Buddhism and technology.

IV. PROPOSED METHODOLOGY

This study uses interpretive textual analysis and comparative normative reasoning. Its primary textual selection is purposive: AN 6.63 addresses intention; MN 43 addresses wisdom and conditions of understanding; MN 61 addresses reflection on action and harm. The disciplinary material concerns Pācittiya 1, Pācittiya 73, and the settlement of monastic issues. These passages were selected because they permit examination of motivation, evidential responsibility, truthful conduct, and correction. They are not presented as a representative survey of the entire Pāli canon.

 ISSN: 2582-8150

The discourses are consulted in published English translations. For disciplinary rules and their procedural explanation, the study uses Ṭhānissaro Bhikkhu's Buddhist Monastic Code [9], distinguishing its modern exposition from the canonical rules it discusses. This reliance on translation and commentary limits claims about Pāli terminology and historical interpretation. No manuscript comparison, philological reconstruction, or fieldwork has been undertaken. Sutta material is distinguished from Vinaya material throughout the analysis.

Each comparison follows three steps. First, a passage is interpreted within the kind of action or procedure it addresses. Second, the ethical distinction it supports is identified without extending the rule beyond its stated conditions. Third, a proposed contemporary application is assessed for both relevance and disanalogy. For example, a disciplinary rule concerning deliberate falsehood can inform an argument about truthful reporting, but it cannot by itself determine the acceptable error rate of a classifier. That determination requires technical evidence and a justified institutional policy.

The comparative assessment uses four questions: What purpose guides the decision? What evidence supports reliance on the recommendation? What restraint is required in acting and communicating? Who can examine and correct the outcome? Answers must identify practical capacities as well as professed values. The analysis evaluates whether the proposed reasons are internally coherent, whether the analogy preserves relevant differences, and whether responsibility is assigned to people who can exercise it.

Two constructed scenarios test the framework's explanatory usefulness. They are hypothetical examples, not observations, simulations, or experimental results. No dataset was collected, no algorithm was benchmarked, and no participants were studied. Section V therefore presents analytical results and discussion. Generative AI assisted with research organization, substantial drafting, and document preparation. The author is responsible for verifying sources, revising the analysis, and approving the final text.

V. ANALYTICAL RESULTS AND DISCUSSION

A. Intention and the limits of benevolent purpose

AN 6.63 associates kamma with intention and action through body, speech, and mind [6]. This locates an important dimension of responsibility in the orientation of human action. Applied comparatively, the question is why people commission and use a system: to assist fair consideration, reduce costs, exercise control, or pursue another purpose. The passage does not imply that a machine's optimization objective is itself cetanā, or that desirable intentions guarantee desirable outcomes.

The proposed inference is that intention should be examined together with foreseeable consequences and opportunities for correction. MN 61 describes reflection on harm before, during, and after action [8]. Its temporal structure provides a resource for criticizing an organization that assesses ethics only at procurement. In the contemporary application developed here, initial review asks whether the task should be automated; ongoing review asks whether actual use remains defensible and retrospective review asks what should change after an error. These are proposed institutional adaptations rather than procedures specified in the discourse.

Consider an organization that introduces automated recruitment to widen access. Its declared purpose may be beneficial. If it later learns that a model is unsuitable for a changed role and continues using it without reassessment, the original purpose does not adequately justify continued reliance. The ethical problem concerns a subsequent human decision under changed evidence. Responsibility must therefore include the willingness to revise a practice that once appeared reasonable.

B. Wisdom and evidential responsibility

MN 43 presents wisdom within an account of liberating understanding and identifies learning and appropriate attention among conditions relevant to right view [7]. Its account cannot be reduced to technical accuracy. The useful comparison concerns the discipline of examining an interpretation rather than treating an appearance of certainty as sufficient understanding. In AI-assisted work, that discipline requires asking what an output represents, what information it omits, and whether its evidential basis fits the present task.

A probability or ranking acquires decision relevance only through additional judgments. An estimate can be accurate on a test dataset but unsuitable for a new

population. A classification can depend on an incomplete record rather than on an applicant's actual circumstances. A reviewer should therefore distinguish the system's reported confidence from justified confidence in a consequential decision. This distinction is philosophical; the adequacy of particular validation methods remains a technical question.

Epistemic responsibility is also distributed. Designers can document assumptions, deploying organizations can provide relevant records, and reviewers can identify inconsistencies within their competence. A case worker cannot personally reconstruct every model. Yet an institution cannot use that limitation to demand unquestioning acceptance. It must provide a route to specialist scrutiny, allow uncertainty to be recorded, and define when a decision should be paused. These are enabling conditions for responsible reliance, not guarantees of correct judgment.

C. Moral discipline and truthful communication

Pācittiya 1 concerns deliberate falsehood by a monk; the rule's treatment distinguishes intentional lying from error [9, ch. 8, rule 1]. Contemporary analogy is correspondingly limited. An inaccurate model output is not automatically an intentional lie. Ethical responsibility arises when people knowingly misrepresent its evidential status, hide relevant limitations, or describe an unverified recommendation as established fact. The immediate object of evaluation is human communication and conduct.

Pācittiya 73 addresses feigned ignorance under specified disciplinary conditions; it should not be read as making every instance of genuine ignorance the same offence [9, ch. 8, rule 73]. Its comparative relevance is to strategic avoidance of an acknowledged responsibility to learn. For AI governance, this suggests examining whether an organization provides training and whether authorized personnel deliberately disregard limitations they have been required and enabled to understand. The analogy does not license blaming employees for information or training withheld by their institution.

Truthful communication requires language that reflects the basis of a decision. An institution should explain the role of an AI recommendation, identify material uncertainty, and distinguish its own judgment from the model's classification. That communication must be useful to the affected person. Providing a technical description that cannot help someone identify or challenge an error may satisfy a reporting habit without supporting answerability.

D. Procedural answerability and meaningful intervention

The seven methods for settling monastic issues apply to different kinds of disciplinary matters; they are not interchangeable ways to resolve every dispute [9, ch. 11]. Procedures involving presence, acknowledgement, or qualified majority decisions operate within particular conditions. The philosophical lesson developed here is that the acceptability of judgment depends partly on appropriate procedure. Communal agreement alone does not establish that a conclusion is properly supported.

A modern AI institution cannot simply reproduce monastic procedure. Disciplinary goals and religious authority are different than public administration and employment. Membership of monastic orders. Rationale is irrelevant to rights and legal duties as they stand. Nevertheless, the comparison illuminates why answerability should include an identifiable issue, an appropriate reviewing authority, an opportunity to address relevant evidence, and a route to correction. These requirements are proposals for present institutions, not assertions that the Vinaya contains a modern administrative code.

Meaningful intervention requires more than permission to disagree. Reviewers need information, competence, time, and effective authority. If a worker faces an unmanageable queue or penalties for justified overrides, formal discretion may provide little practical control. Responsibility should consequently extend to managers who allocate resources and establish incentives. A recorded human signature cannot transfer all responsibility away from those institutional choices.

E. Application to hypothetical cases

In the first scenario, a scholarship system recommends rejecting an applicant because its records omit a recently completed qualification. The recommendation may follow the data supplied to it, but the decision is inadequately grounded. A competent reviewer checks the discrepancy, obtains admissible evidence through the institution's ordinary process, and records a reason for the final decision. If the mistake affects other applicants, case correction should prompt examination of the data process as well. The proposal concerns

warranted review, not a guarantee that every appeal succeeds.

The second case is when an employer reuses a recruitment model after significantly altering the requirements of a post. The recruiter has little time to evaluate the applicants and cannot put the tool on hold. The ethical failure is not adequately described as an individual failure to be compassionate. It concerns a mismatch between delegated responsibility and effective control. An appropriate response may include pausing unsupported use, reassessing suitability, and providing reviewers with an authorized escalation route. Table 1 summarizes the distinctions illustrated by these scenarios.

Table 1. Hypothetical Applications of the Framework

Scenario	Ethical problem	Proposed response
Student aid	Incomplete evidence treated as sufficient	Check missing qualification; record reasons; review data process.
Hiring	Changed task and ineffective reviewer control	Reassess suitability; authorize escalation; pause unsupported use.

These examples support an analytical finding: responsible oversight depends on a defensible connection between purpose, evidence, conduct, and correction. They do not demonstrate improved accuracy, reduced discrimination, or public acceptance. Even man's judgment is faulty and biased. More resources may equal more favorable appeals. Institutions should also analyze initial choices and rectification methods. The Buddhist comparison supplies reasons to examine these conditions while leaving empirical assessment necessary.

VI. CONCLUSION

A Theravāda Buddhist philosophical perspective can illuminate human responsibility in AI-assisted decisions by connecting moral orientation with disciplined judgment. The analysis distinguishes intention, epistemic care, moral restraint, and procedural answerability. Intention is one element in human activity, but it does not free us from the obligation to know facts and to act on predicted difficulties. Wisdom calls for careful attention to the conditions of understanding, and disciplinary examples make obvious the need for precise communication and well-differentiated procedures.

Both personal and institutional control are proposed as part of the structure. There must be opportunities for reviewers to understand, question, document, and revise significant choices. Organizations remain responsible for objectives, resources, incentives, and correction mechanisms that shape those opportunities. Ethical responsibility therefore cannot be discharged merely by inserting a person into an otherwise unexamined decision process.

The argument is limited by its selective textual corpus, dependence on translations, and use of hypothetical cases. It neither establishes that Buddhist practice improves AI governance nor resolves questions about artificial consciousness. Its contribution is a comparative normative account of the conditions under which human reliance on AI can be defended. Technical validation, accessible review, and evaluation of institutional practice remain necessary to determine whether those conditions are actually met.

VII. FUTURE SCOPES

Future research should examine how the proposed distinctions operate in real institutions. Qualitative studies could investigate how reviewers understand responsibility, communicate uncertainty, and experience constraints on intervention. Such research should include affected people alongside developers and managers, since a procedure that appears adequate to its designers may be difficult to use in practice.

Controlled studies could compare different review arrangements utilizing specified activities, documented decision criteria, and independently assessed outcomes. Relevant measures include recognition of incomplete evidence, justified overrides, quality of recorded reasons, consistency of decisions, and access to correction. Evaluations should also measure review time and possible burdens on disadvantaged participants. Findings would need to distinguish improved procedural quality from improvements in predictive performance.

Further textual research could compare translations and commentarial treatments of intention, wisdom, and disciplinary procedure and examine disagreements within contemporary Theravāda interpretation. These

investigations would refine the framework's philosophical basis. They would also help identify where the analogy should stop, preserving the distinctive purposes of Buddhist training while developing publicly assessable reasons for responsible technological practice.

REFERENCES

[1] F. Santoni de Sio and J. van den Hoven, "Meaningful human control over autonomous systems: A philosophical account," Frontiers in Robotics and AI, vol. 5, article 15, 2018. doi: 10.3389/frobt.2018.00015.

[2] L. Cavalcante Siebert et al., "Meaningful human control: Actionable properties for AI system development," AI and Ethics, vol. 3, pp. 241–255, 2023. doi: 10.1007/s43681-022-00167-3.

[3] B. Green and Y. Chen, "Algorithm-in-the-loop decision making," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 9, pp. 13663–13664, 2020. doi: 10.1609/aaai.v34i09.7115.

[4] E. Tabassi, Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1. National Institute of Standards and Technology, 2023 doi: 10.6028/NIST.AI.100-1.

[5] UNESCO, Recommendation on the Ethics of Artificial Intelligence, adopted November 23, 2021. Available: https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence. Accessed: October 13, 2026.

[6] Ṭhānissaro Bhikkhu, trans., "Nibbedhika Sutta: Penetrative (AN 6.63)," Access to Insight, 1997. Available: https://www.accesstoinsight.org/tipitaka/an/an06/an06.063.than.html. Accessed: October 13, 2026.

[7] Ṭhānissaro Bhikkhu, trans., "Mahāvedalla Sutta: The Greater Set of Questions-and-Answers (MN 43)," Access to Insight, 2006. Available: https://www.accesstoinsight.org/tipitaka/mn/mn.043.than.html. Accessed: October 13, 2026.

[8] Ṭhānissaro Bhikkhu, trans., "Ambalaṭṭhika-Rāhulovāda Sutta: Instructions to Rāhula at Mango Stone (MN 61)," Access to Insight, 2006. Available: https://www.accesstoinsight.org/tipitaka/mn/mn.061.than.html. Accessed: October 13, 2026.

[9] Ṭhānissaro Bhikkhu, The Buddhist Monastic Code, Volume I: The Pāṭimokkha Rules, online edition, n.d., ch. 8, rules 1 and 73; ch. 11. Available: https://www.dhammatalks.org/vinaya/bmc/Section0000.html. Accessed October 13, 2026

e-ISSN : 2582-8150