RESPONSIBLE ARTIFICIAL INTELLIGENCE FOR MANAGERIAL DECISION SUPPORT: AN ASSURANCE-ORIENTED CAPABILITY MODEL

Maya Ariyanti[1] Nazhira Nindya Padma Hanuun[2]

[1]Telkom University; [2]Yayasan Kreatif Indonesia Emas

Emails: ariyanti@telkomuniversity.ac.id; nazhiranindya@gmail.com

Abstract

Background: Organizations increasingly depend on responsible artificial intelligence, yet visible activity does not by itself demonstrate reliable capability or sustainable value. Fragmented responsibilities, weak evidence, and locally optimized metrics can cause decisions about managerial decision support to create unmanaged exposure elsewhere. **Aim:** This article develops a governance framework that connects responsible artificial intelligence, managerial decision support, and human judgment through five mutually reinforcing capabilities: data integrity, model and technology controls, human judgment, accountability, and continuous monitoring. **Method:** The paper uses an integrative conceptual review. Established management research, professional standards, and institutional guidance are synthesized through construct clarification, mechanism mapping, risk-control analysis, and proposition development. It does not report respondents, sample statistics, or causal estimates. **Results:** The synthesis indicates that performance becomes more resilient when decision rights, data definitions, controls, escalation paths, and learning routines are designed as one management system. The proposed model links each capability to observable evidence and balanced indicators. **Contribution:** The article offers an auditable implementation sequence and propositions that can be tested in later empirical research.

Keywords: *Responsible Artificial Intelligence; Managerial Decision Support; Data Integrity; Governance; Organizational Capability*

1. INTRODUCTION

The practical problem addressed by this article is not whether organizations should invest in responsible artificial intelligence, but how they can turn that investment into reliable and defensible outcomes. Managers often observe activity through projects, transactions, dashboards, or compliance tasks. These observations are useful, but they can conceal whether the organization has the capability to repeat results under changing conditions.

A second problem is interdependence. Decisions concerning responsible artificial intelligence alter requirements for managerial decision support, while choices about human judgment influence risk, cost, customer outcomes, and accountability. When functions use different definitions and review cycles, the same event can be treated as success by one unit and as an unresolved exposure by another.

Dynamic-capability theory explains why durable performance depends on the ability to sense changes, seize opportunities, and reconfigure resources rather than on static assets alone (Teece, 2007). The resource-based view likewise directs attention to routines and knowledge that are valuable, difficult to imitate, and supported by the organization (Barney, 1991). These perspectives imply that governance should make coordinated capability visible.

For digital-oriented decisions, the quality of evidence is particularly important. Data must be sufficiently complete, timely, consistent, and traceable for its intended decision. Control should not be reduced to after-the-fact checking; it should shape definitions, approval boundaries, exception handling, and the retention of evidence while work is performed.

1. INTRODUCTION (CONTINUED)

This article therefore asks: how can organizations govern responsible artificial intelligence so that it strengthens managerial decision support, supports human judgment, and remains accountable under uncertainty? The objective is to specify mechanisms, roles, indicators, and boundary conditions rather than to claim a universal causal effect.

The contribution is practical and research-oriented. It distinguishes outputs from capabilities, presents an integrated framework, identifies predictable failure modes, and states propositions suitable for survey, case-study, archival, or design-science testing.

The unit of analysis in the proposed model is the management decision rather than the technology, department, or policy in isolation. A decision-centered view makes it possible to connect responsible artificial intelligence with the information used, the authority exercised, the parties affected, and the evidence retained. It also clarifies where coordination is necessary and where local discretion remains appropriate.

The model treats data integrity as an anticipatory capability. It combines environmental scanning, stakeholder signals, operational exceptions, and professional interpretation. The objective is not to collect every available signal, but to distinguish information that could change a material choice from information that merely increases reporting volume.

The design capability represented by model and technology controls converts an identified need into roles, standards, thresholds, and control points. Good design specifies what must be consistent and what may be adapted. This distinction is important because excessive standardization can suppress local knowledge, while unlimited variation makes assurance and organizational learning difficult.

This decision-centered perspective also changes how success is interpreted. An organization may increase activity related to responsible artificial intelligence without improving its ability to make consistent choices. Evidence of capability should therefore include repeatability across people and periods, appropriate handling of unusual cases, and the capacity to explain why a decision was reasonable with the information available at the time.

Stakeholder consequences create an additional analytical requirement. Efficiency or financial value should not be assumed to benefit every affected group equally. Segmented evidence can reveal whether access, service quality, risk, or burden is distributed unevenly. When trade-offs are unavoidable, the decision rationale and mitigating action should be recorded rather than hidden within an aggregate result.

2. LITERATURE REVIEW AND CONCEPTUAL FOUNDATIONS

Capability is treated here as a repeatable organizational ability supported by people, process, information, technology, and governance. In the context of responsible artificial intelligence, this definition prevents a tool, policy, or isolated project from being mistaken for institutionalized practice. A capability exists only when responsibilities and evidence remain workable during routine operations and disruption.

Information quality provides the connective tissue between responsible artificial intelligence and managerial action. The information-systems success literature emphasizes that system quality and information quality shape use and benefits (DeLone & McLean, 2003). For governance purposes, fitness for use must be defined at the level of the decision: the same dataset may be adequate for monitoring but inadequate for approval, estimation, or assurance.

Balanced measurement is also necessary because narrow targets invite local optimization. Financial outcomes should be connected with operational reliability, stakeholder experience, risk exposure, and learning. Kaplan and Norton (1992) showed the value of linking several perspectives; the present framework extends that logic by requiring traceability from a metric to its owner, source, control, and decision consequence.

The main governance tension is between speed and assurance. Excessive control can delay legitimate adaptation, while weak control can normalize poor-quality data and opaque automated decisions. Risk-based design resolves this tension by increasing scrutiny where reversibility is low, potential harm is high, or evidence is difficult to reconstruct.

Execution depends on human judgment. People must understand the purpose of the process, the limits of their authority, and the conditions that require escalation. Training is therefore insufficient when it only explains system steps. It should also develop judgment about uncertainty, competing objectives, vulnerable stakeholders, and the consequences of overriding a control.

The assurance dimension, accountability, provides structured challenge without transferring operating responsibility away from management. Evidence should allow a reviewer to reconstruct the decision, test whether required controls operated, and determine whether an exception was resolved at the appropriate level. Assurance findings then become inputs to learning rather than isolated compliance events.

Finally, continuous monitoring closes the management cycle. Learning requires comparison between expected and observed outcomes, analysis of recurring exceptions, and revision of assumptions or routines. A lesson is institutionalized only when it changes a definition, control, resource allocation, training practice, or decision rule and when that change is subsequently evaluated.

The framework does not assume that more control is always better. Control has direct and opportunity costs, and poorly designed requirements may encourage workarounds. Periodic control evaluation should therefore examine operating effectiveness, user burden, duplicated evidence, and whether the underlying risk has changed. Controls that no longer add decision value should be redesigned or retired through an authorized process.

3. RESEARCH METHOD

An integrative conceptual review was selected because the research question crosses functional and institutional boundaries. The method combines established theories, professional standards, and policy guidance into an explanatory model. It is not presented as a systematic review or meta-analysis, and no fabricated search counts or effect sizes are reported.

Sources were considered when they clarified a construct, described a governance mechanism, established a recognized control principle, or proposed a decision-useful measure relevant to responsible artificial intelligence. Analysis proceeded through four steps: definition of constructs, mapping of mechanisms, classification of risks and controls, and synthesis into capability dimensions and propositions.

The framework was evaluated conceptually against four criteria: internal coherence, traceability of claims to mechanisms, practical observability, and adaptability across organizational settings. These criteria do not establish empirical validity. They are design tests that make future validation more disciplined.

Construct clarification began by separating capabilities, activities, outputs, and outcomes. This step prevents common category errors, such as treating the existence of a dashboard as evidence of decision quality or treating a completed audit as evidence that underlying risk has declined. Each construct was defined in terms of an organizational ability that can be observed through recurring practice.

Mechanism mapping then connected each construct to an expected behavioral or informational pathway. For example, clearer ownership should reduce unresolved exceptions because a named role has both authority and accountability. The mapping also identifies alternative explanations and boundary conditions, making the propositions more suitable for later empirical testing.

Risk-control analysis was used to examine foreseeable failure modes. Risks were framed as conditions that could prevent the intended mechanism from operating, and controls were classified as preventive, detective, or corrective. Evidence requirements were specified so that control operation can be assessed independently rather than inferred from the absence of a reported incident.

The final synthesis organized the constructs into a five-stage cycle. The sequence is analytical rather than strictly chronological: organizations may revisit an earlier stage when new evidence changes an assumption. This recursive structure reflects management under uncertainty and avoids presenting the framework as a one-time implementation checklist.

Time is an important boundary condition. Some decisions require rapid response with incomplete evidence, while others permit extended analysis. Predefined emergency authority, minimum evidence standards, retrospective review, and expiry dates for temporary measures allow organizations to respond quickly without turning exceptional practice into an undocumented permanent routine.

4. Conceptual Results and Discussion

The first result is that data integrity must precede automation or scale. Leaders need an agreed decision purpose, an accountable owner, and a definition of acceptable evidence. Without these elements, expansion magnifies inconsistent practice and makes later remediation more expensive.

The second result concerns model and technology controls. Controls should be embedded in the work rather than added only at reporting time. Validation rules, approval thresholds, segregation of duties, and retained decision logs reduce the likelihood that poor-quality data will remain invisible until consequences are material.

The third result is the continued importance of human judgment. Judgment is not the absence of structure. High-quality judgment states assumptions, considers alternatives, records overrides, and escalates uncertainty. These practices reduce opaque automated decisions and create evidence for later review.

4.1 Integrated capability framework

Stage	Capability	Required practice	Illustrative indicator
Sense	Data Integrity	Translate responsible artificial intelligence into explicit responsibilities, evidence requirements, and decision criteria.	data-quality exceptions
Design	Model And Technology Controls	Translate managerial decision support into explicit responsibilities, evidence requirements, and decision criteria.	override frequency
Execute	Human Judgment	Translate responsible artificial intelligence into explicit responsibilities, evidence requirements, and decision criteria.	control-test completion
Assure	Accountability	Translate managerial decision support into explicit responsibilities, evidence requirements, and decision criteria.	incident recovery time
Learn	Continuous Monitoring	Translate responsible artificial intelligence into explicit responsibilities, evidence requirements, and decision criteria.	decision traceability

The fourth result is that accountability must connect first-line ownership with independent challenge. Assurance should test whether controls operate, whether metrics remain fit for purpose, and whether remediation addresses root causes rather than only closing findings.

Finally, continuous monitoring determines whether the framework remains useful. Incidents, complaints, forecast errors, audit findings, and missed opportunities should feed a common learning backlog. Completion should be measured through changed routines and reduced recurrence, not by the number of meetings held.

Organizational size also affects implementation. Smaller entities may combine roles and use simpler evidence, but they still need clear ownership and challenge for material decisions. Larger entities require stronger coordination across units, common definitions, and mechanisms for consolidating exceptions. The principle is proportional governance, not identical organizational form.

4.2 Governance risks and controls

Material risk	Preventive or detective response	Evidence retained
Poor-Quality Data	Assign ownership, define thresholds, validate inputs, and escalate exceptions linked to data integrity.	Decision log, exception record, control test, and trend for data-quality exceptions.
Opaque Automated Decisions	Assign ownership, define thresholds, validate inputs, and escalate exceptions linked to model and technology controls.	Decision log, exception record, control test, and trend for override frequency.
Privacy And Security Failures	Assign ownership, define thresholds, validate inputs, and escalate exceptions linked to human judgment.	Decision log, exception record, control test, and trend for control-test completion.
Automation Bias	Assign ownership, define thresholds, validate inputs, and escalate exceptions linked to accountability.	Decision log, exception record, control test, and trend for incident recovery time.
Weak Incident Learning	Assign ownership, define thresholds, validate inputs, and escalate exceptions linked to continuous monitoring.	Decision log, exception record, control test, and trend for decision traceability.

A central mechanism is alignment between decision rights and evidence quality. Individuals should not be accountable for outcomes when they lack access to the information or authority needed to influence them. Conversely, broad authority without traceable evidence increases the risk of poor-quality data. A responsibility matrix should therefore be paired with evidence requirements and escalation time limits.

A second mechanism is controlled transparency. Relevant information about responsible artificial intelligence should be visible to those who must act, challenge, or learn, while privacy, confidentiality, and security obligations remain protected. Transparency is not indiscriminate disclosure; it is purposeful access supported by definitions, provenance, retention rules, and documented restrictions.

A third mechanism is balanced feedback. Leading indicators such as data-quality exceptions and override frequency reveal whether the process is functioning before final outcomes are known. Lagging indicators remain necessary, but they should be interpreted together with stakeholder effects, risk exposure, and the cost of remediation. This reduces pressure to optimize one measure at the expense of the system.

A fourth mechanism is proportional control. The depth of approval, testing, and documentation should reflect materiality, reversibility, novelty, and potential harm. Routine low-risk choices can use automated or sampled controls, whereas unusual or difficult-to-reverse choices require explicit challenge and stronger evidence. Proportionality helps preserve speed without normalizing unmanaged exposure.

External partners should be included when they perform material activities or provide consequential information. Contracts can state service levels and audit rights, but operational governance additionally requires data definitions, incident notification, escalation contacts, continuity expectations, and joint learning after failure. Outsourcing an activity does not eliminate accountability for its effects.

4.3 Governance architecture

Governance begins with a materiality statement for responsible artificial intelligence. The statement should identify the decisions in scope, affected stakeholders, principal value drivers, and consequences of error or delay. Materiality should be revisited when scale, regulation, technology, or stakeholder exposure changes; otherwise a once-appropriate control design may become inadequate.

Role design should distinguish ownership, execution, challenge, approval, and assurance. Combining roles can be appropriate in smaller organizations, but the combination should be explicit and supported by compensating review where conflicts could arise. The objective is not bureaucratic separation for its own sake; it is clarity about who decides, who provides evidence, and who can question the decision.

Exception management is a practical test of governance quality. Thresholds linked to control-test completion and incident recovery time should determine when an issue is corrected locally, escalated, accepted temporarily, or stopped. Each material exception needs an owner, due date, rationale, and closure evidence. Repeated extensions should trigger examination of resources or design rather than routine administrative renewal.

Oversight information should preserve context. Aggregated status colors can support attention, but they should not replace trends, assumptions, unresolved disagreements, and the distribution of outcomes across relevant groups. A concise narrative explaining why a measure changed and what management will do next is often more decision-useful than additional dashboard elements.

Independent review should focus on the reliability of the management system. Reviewers can test data lineage, sample decisions, examine overrides, and assess whether remediation addresses causes. Their role is not to guarantee success, but to provide disciplined challenge and evidence about whether governance operates as described.

4.3 Research propositions

P1. Greater maturity of data integrity is positively associated with the consistency of decisions concerning responsible artificial intelligence.

P2. Model And Technology Controls strengthens the relationship between information quality and accountable execution.

P3. The benefit of human judgment is greater when overrides and assumptions are traceable and independently reviewable.

P4. Continuous Monitoring reduces recurrence of material exceptions when remediation addresses causes rather than symptoms.

Maturity can be described in four levels: ad hoc, defined, operating, and adaptive. At the ad hoc level, results depend heavily on individuals. A defined organization documents roles and standards. An operating organization retains evidence that routines work. An adaptive organization uses outcomes and exceptions to revise the system. Progress should be demonstrated through practice, not labels alone.

5. MEASUREMENT AND EVALUATION

Measurement for responsible artificial intelligence should begin with a logic chain linking resources, activities, outputs, intermediate outcomes, and intended value. This structure makes assumptions visible. When an outcome weakens, management can investigate whether the cause lies in execution, an invalid assumption, an external change, or a measure that no longer represents the objective.

The indicator data-quality exceptions can be used as a leading signal when its definition, calculation frequency, data owner, and acceptable range are documented. A target without these attributes is vulnerable to inconsistent interpretation. Historical baselines and segmentation should also be retained so that averages do not conceal deteriorating performance in a material unit or stakeholder group.

The indicator override frequency should be read together with evidence about quality and consequences. Faster completion may represent genuine improvement, reduced scope, deferred work, or weakened challenge. Balanced interpretation requires at least one value measure, one reliability measure, one risk measure, and one learning measure for each material decision process.

Targets should have review triggers. A target may need revision after structural change, but retrospective adjustment merely to preserve apparent success undermines accountability. Changes should record the rationale, approving role, effective date, and effect on comparability. Where continuity matters, both the original and revised series should be shown.

Learning measures such as decision traceability assess whether the organization changes practice after evidence emerges. Closure counts alone are weak because actions can be marked complete without reducing recurrence. Stronger evaluation examines whether the control or routine changed, whether users adopted the change, and whether subsequent outcomes support the intended mechanism.

The mechanisms are mutually reinforcing. Better information without clear authority can leave problems unresolved; authority without challenge can enable override; control without learning can repeat the same failure; and learning without implementation remains anecdotal. Management maturity should therefore be assessed across the complete cycle rather than by the presence of individual practices.

Ethical considerations should be integrated into ordinary governance rather than assigned only to a specialist review at the end. Teams should identify affected parties, foreseeable harms, contestability needs, and acceptable limits before scale. The analysis should be revisited when new uses, data sources, partners, or stakeholder concerns change the risk profile.

6. Implementation Roadmap

Phase	Management task	Minimum evidence	Review question
Diagnose	Select one material decision and map its current owners, inputs, controls, and outcomes.	Current-state map and risk statement	Can the organization reconstruct how the decision is made?
Design	Agree definitions, authority limits, escalation conditions, and balanced indicators.	Approved design and data dictionary	Are speed, value, risk, and stakeholder effects represented?
Pilot	Run the design on a limited scope; record exceptions, overrides, and user feedback.	Pilot log and control evidence	Which assumptions failed and which controls added value?
Scale	Embed responsibilities in routines, systems, training, and assurance plans.	Operating procedures and ownership matrix	Does practice remain consistent across units and partners?
Learn	Review outcomes and recurrence; update thresholds, measures, and training.	Learning backlog and closure evidence	Did remediation change behavior and reduce recurrence?

A ninety-day pilot can translate the framework into routine practice. During the first thirty days, management maps one material decision and agrees the risk statement, roles, definitions, and baseline evidence. The next thirty days test controls and escalation using live cases. The final thirty days review exceptions, user burden, unintended effects, and the quality of retained evidence before scale decisions are made.

Implementation capacity should be planned explicitly. Process owners need time to document and improve routines; data owners need resources for quality correction; control functions need access and analytical capability; and leaders need forums in which unresolved trade-offs can be decided. Assigning responsibilities without capacity creates nominal governance and predictable delay.

Change management should explain why the framework matters to each role. Employees are more likely to follow controls when they understand the risk addressed and can report impractical requirements. Feedback channels should distinguish requests for clarification, proposed improvements, control failures, and protected escalation of misconduct or harm.

Scaling decisions should be based on evidence from the pilot. Management should identify which components are universal, which require contextual adaptation, and which added burden without improving decisions. A controlled rollout with periodic maturity reviews is preferable to declaring enterprise-wide adoption before operating evidence exists.

For managers, implementation should begin with one material decision involving responsible artificial intelligence. The team should document its purpose, owner, inputs, control points, escalation conditions, and outcome measures. This narrow starting point makes gaps visible without launching a broad transformation program before responsibilities are clear.

7. PRACTICAL AND RESEARCH IMPLICATIONS

For boards and oversight bodies, the framework supports better questions. Oversight should ask which assumptions are most consequential, how exceptions are surfaced, which indicators could be gamed, and whether management can reconstruct the evidence behind a significant decision. These questions provide more insight than a dashboard reviewed without context.

For researchers, the propositions can be operationalized with multi-level designs. Case studies can examine how mechanisms interact; surveys can measure capability maturity and perceived decision quality; archival work can connect control evidence with outcomes; and longitudinal research can test whether learning routines reduce recurrence after disruptions.

8. LIMITATIONS AND FUTURE RESEARCH

This article is conceptual and does not demonstrate causal effects. The relative importance of each capability may vary by industry, regulation, organizational size, ownership, and technology maturity. Measures that are decision-useful in one setting may be costly or misleading in another.

Future studies should test the framework with transparent sampling, validated measures, and designs that distinguish adoption from effective use. Researchers should also examine unintended effects, including control overload, exclusion, surveillance, and metric gaming. Comparative studies would help identify when simplified governance is more effective than elaborate formal structures.

9. Conclusion

Reliable performance in responsible artificial intelligence does not arise from technology, policy, or functional expertise in isolation. It depends on an integrated management system that connects data integrity, model and technology controls, human judgment, accountability, and continuous monitoring. The framework presented in this article makes those capabilities observable through responsibilities, evidence, controls, balanced indicators, and learning routines. Its propositions remain to be tested empirically, but the model provides a disciplined basis for implementation and research without presenting conceptual claims as measured findings.

Declarations

Editorial status: Authorship, affiliations, content, and publication readiness were confirmed to the editorial office. Similarity-screening and peer-review records are maintained separately by the journal.

Conflict of interest and funding: Author declarations must be recorded in the journal's editorial file before publication.

Data availability: No empirical dataset was created or analyzed for this conceptual article.

References

Barney, J. (1991). Firm resources and sustained competitive advantage. Journal of Management, 17(1), 99-120. https://doi.org/10.1177/014920639101700108

DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: A ten-year update. Journal of Management Information Systems, 19(4), 9-30. https://doi.org/10.1080/07421222.2003.11045748

International Organization for Standardization. (2018). ISO 31000:2018 Risk management - Guidelines.

Kaplan, R. S., & Norton, D. P. (1992). The balanced scorecard - Measures that drive performance. Harvard Business Review, 70(1), 71-79.

NIST. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology.

NIST. (2024). The NIST Cybersecurity Framework (CSF) 2.0. National Institute of Standards and Technology.

OECD. (2019). Recommendation of the Council on Artificial Intelligence. OECD Legal Instruments.

Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of sustainable enterprise performance. Strategic Management Journal, 28(13), 1319-1350. https://doi.org/10.1002/smj.640

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. Journal of Strategic Information Systems, 28(2), 118-144. https://doi.org/10.1016/j.jsis.2019.01.003