E-ISSN: 3108-4192

APSSHS

Academic Publications of Social Sciences and Humanities Studies
2025, Volume 5, Issue 2, Page No: 95-103
Available online at: https://apsshs.com/

Journal of Applied Organizational Systems and Behavior

Responsible Artificial Intelligence for Transparent and Contestable Organizational Decision-Making

Carlos Ramirez[1]*, Elena Torres[2], Pablo Ortega[1], Sofia Mendes[3]

1. Department of Responsible AI and Transparency, Faculty of Business, University of Barcelona, Barcelona, Spain.
2. Department of Contestable Decision-Making, Faculty of Management, University of Lisbon, Lisbon, Portugal.
3. Department of AI Governance, Faculty of Psychology, University of Porto, Porto, Portugal.

Abstract

Artificial intelligence increasingly participates in organizational decisions that allocate work, opportunities, resources, evaluation, and other consequential outcomes. Yet responsible artificial intelligence is often framed as a list of abstract principles or technical safeguards rather than as an organizational policy problem involving authority, review, challenge, and continuing accountability. This Original Organizational Policy Framework Article develops a proposed framework for transparent and contestable organizational decision-making. The conceptual approach integrates scholarship on responsible AI governance, algorithmic management, transparency, human–AI judgment, voice, accountability, and auditing while preserving the distinction between established findings and original synthesis. The framework begins with scope and risk classification and then couples four policy domains: transparency requirements, meaningful human review and override, contestability and redress, and accountability allocation. Monitoring and periodic review provide a recursive mechanism for reassessing policy intensity as systems, data, uses, or organizational conditions change. The central proposition is that higher-risk decisions should not trigger isolated safeguards; instead, governance protections should intensify jointly and remain connected through explicit decision rights and escalation pathways. The framework further distinguishes disclosure from intelligibility, human presence from meaningful authority, voice opportunity from consequential contestation, and technological capability from legitimate responsibility. The proposed relationships are not presented as validated causal effects or universal prescriptions. Their value lies in organizing organizational policy choices into testable propositions and observable governance responsibilities. Future research should compare alternative policy architectures, examine how they are enacted in practice, and test whether particular configurations improve decision quality, fairness, accountability, or affected-party influence under specified conditions.

Keywords: Scope and risk classification, Transparency requirements, Human review and override, Contestability and redress, Monitoring and periodic review, Implementation implications

How to cite this article: Ramirez C, Torres E, Ortega P, Mendes S. Responsible Artificial Intelligence for Transparent and Contestable Organizational Decision-Making. J Appl Organ Syst Behav. 2025;5(2):95-103. https://doi.org/10.51847/9SaawmvfX8

Received: 29 October 2025; **Revised:** 25 December 2025; **Accepted:** 26 December 2025
Corresponding author: Carlos Ramirez
E-mail ✉ carlos.ramirez@ub.edu

Introduction

Responsible artificial intelligence has moved beyond a technical question about model design into an organizational question about who may rely on AI, for which decisions, under what information conditions, with what opportunities for review, and with whose continuing responsibility. Recent governance scholarship shows that organizations face a fragmented translation problem: broad principles such as fairness, transparency, and accountability must be converted into structural, relational, and procedural arrangements that operate across the AI lifecycle [1]. At the field level, ethics initiatives show notable convergence around recurring principles while differing substantially in interpretation and implementation [2]. This creates a policy problem rather than merely a principles problem.

The limitation of principle-led approaches is that abstract commitments cannot by themselves specify decision rights, escalation rules, or the institutional conditions under which a person can question an AI-supported outcome. Principles alone therefore cannot guarantee ethical AI [3]. Organizational management research also emphasizes that AI combines autonomy, learning, and inscrutability in ways that make governance recurrent rather than a one-time approval decision [4]. These characteristics may alter how information, expertise, and discretion are distributed even when the system is nominally advisory.

A second limitation is the tendency to imagine human and machine decision-making as substitutable categories. The automation–augmentation paradox instead suggests that automation and augmentation are interdependent organizational choices whose consequences unfold together [5]. Accordingly, this article treats responsible AI as an organizational policy architecture for governing relationships among technology, decision authority, affected parties, and oversight. The persistent gap between ethical commitments and organizational operationalization is consistent with evidence that governance practices remain fragmented, implementation differs across principle systems, and institutional mechanisms are necessary to make principles actionable [1–3].

The article proposes a Responsible Artificial Intelligence Policy Framework organized around five connected domains: scope and risk classification, transparency requirements, meaningful human review and override, contestability and redress, and accountability allocation, with monitoring and periodic review operating across the architecture. The contribution is explicitly non-empirical. It does not claim that the proposed configuration is causally validated, legally sufficient, or universally appropriate. Instead, it offers a policy-level synthesis that identifies governance functions, separates adjacent constructs, specifies original decision principles, and creates a basis for empirical comparison of alternative organizational arrangements.

Scope and risk classification

A responsible AI policy requires an explicit answer to a prior question: which organizational decisions fall within its scope, and which require more intensive safeguards? Trustworthy-AI governance cannot be reduced to scrutiny of model output because the reliability and legitimacy of AI-supported decisions also depend on data, processes, information-sharing arrangements, and the institutional setting in which algorithms operate [6]. Scope should therefore encompass the sociotechnical decision process, not merely the computational component.

Risk classification is especially important when AI affects employment, evaluation, access, or other consequential organizational interests. Analysis of algorithm-based human-resource decisions shows that efficiency-oriented systems may create tensions with personal integrity and employee rights, indicating that organizational significance cannot be inferred from technical sophistication alone [7]. Work on algorithmic control likewise demonstrates that digital systems can redistribute managerial control through directing, evaluating, disciplining, rewarding, or replacing worker activity [8]. Such redistribution makes power asymmetry and the location of discretion relevant to policy classification.

Context also matters. Algorithmic-management research indicates that consequences vary according to the management function involved and may be conditioned by transparency, fairness, and the extent of human influence [9]. It would therefore be misleading to classify all AI-supported decisions according to a single technology label. This article instead proposes a risk-classification logic based on the decision's consequence severity, reversibility or remediability, automation or delegation depth, organizational discretion, power asymmetry, data sensitivity and representativeness, opacity or uncertainty, and scale or frequency.

These dimensions are an original policy synthesis rather than a validated scoring instrument. They are intended to organize attention and determine which safeguards require closer examination. The classification does not imply a universal numerical threshold, and it does not convert contextual evidence into a causal claim that a particular dimension necessarily produces harm. Its function is to make the basis for differentiated governance explicit and revisable.

Transparency requirements

Transparency should be treated as a governance function, not as an undifferentiated demand for more information. Critical work on algorithmic accountability shows that visibility alone does not ensure understanding, scrutiny, or accountability [10]. An organization may disclose that AI is used while leaving affected parties unable to understand what role it played, which information mattered, who can review the decision, or how a challenge can be raised. Disclosure is therefore necessary in some settings but is conceptually narrower than transparency for governance.

Explanation requirements must also be purpose-bound. Stakeholder-oriented research on explainable AI shows that different actors seek different forms of information because their goals, knowledge, and responsibilities differ [11]. A decision subject may need understandable reasons and a challenge route; a reviewer may need provenance, limitations, uncertainty, and access to the evidentiary record; a system owner may require documentation enabling monitoring and audit. The proposed policy implication is not maximal disclosure but information that is adequate for the recipient's governance task.

A further distinction concerns explanation versus interpretability. In high-stakes settings, post-hoc explanations of opaque models should not automatically be treated as equivalent to using interpretable models when interpretable alternatives are feasible [12]. This distinction matters because a plausible explanation may communicate a decision without exposing the actual decision logic sufficiently for independent scrutiny. Transparency policy should therefore ask what must be knowable, by whom, at what stage, and for which decision function.

Table 1 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for definitions, constructs, and conceptual boundaries for responsible artificial intelligence for transparent and contestable organizational decision-making.

Table 1. Definitions, constructs, and conceptual boundaries for Responsible Artificial Intelligence for Transparent and Contestable Organizational Decision-Making.

Construct or Component	Definition	Problem Addressed	Distinction from Adjacent Concepts	Proposed Role	Uncertainty	Boundary Statement
Responsible AI as Organizational Policy	Organizational arrangements governing AI-influenced decisions across their lifecycle	Principles without operational authority	Broader than model ethics or technical assurance	Integrate safeguards and decision rights	Effective configurations remain uncertain	Proposed architecture, not validated governance effect
Scope	Boundary of decisions, systems, actors, and processes governed by policy	Narrow focus on the model alone	Scope identifies what is governed; risk classifies intensity	Define policy coverage	Appropriate boundary varies by context	Does not imply that every AI use is equally consequential
Risk Classification	Proposed categorization of governance intensity using decision consequences and context	Uniform treatment of heterogeneous AI uses	Not a technical model-risk score	Trigger differentiated safeguards	No universal weighting or threshold is established	Original policy synthesis
Transparency	Information enabling relevant scrutiny of AI involvement and decision process	Disclosure without intelligibility	More than notice; less than proof of accountability	Enable review and challenge	Information needs differ by audience	Transparency does not establish fairness or accountability
Explanation	Stakeholder-relevant account of reasons, inputs, or process	One-size-fits-all disclosure	Differs from intrinsic interpretability	Support role-specific understanding	Appropriate form is context-dependent	Explanation is not proof of correctness
Interpretability	Degree to which decision logic can be understood directly	Reliance on post-hoc accounts in consequential settings	Distinct from explaining an opaque model after the fact	Strengthen scrutiny where materially required and feasible	Feasibility varies by task and model	Interpretability is not guaranteed fairness
Human Review	Proposed working definition: deliberate human reassessment of an AI-supported decision	Symbolic human presence	Review requires judgment, not mere workflow inclusion	Provide substantive scrutiny	Performance varies with expertise and context	Proposed construct boundary, not validated safeguard
Override Authority	Proposed capacity to pause, modify, reject, or escalate AI output	Human reviewer without usable decision power	More consequential than advisory consultation	Preserve actionable human discretion	Effects on governance outcomes remain unvalidated	Proposed governance condition
Contestability and Redress	Proposed route for challenge, reconsideration, reasoned response, and remedy where warranted	Voice without consequential influence	Contestability is not identical to transparency or voice opportunity	Make challenge capable of affecting process	Optimal procedure remains uncertain	Original governance mechanism
Accountability Allocation	Proposed distribution of answerability among decision, system, data, review, and governance roles	Responsibility diffusion after delegation	Distributed responsibility must not become unowned responsibility	Preserve identifiable answerability	Legal allocation varies by jurisdiction	Organizational allocation is not legal liability determination

Monitoring and Periodic Review	Proposed recurrent assessment of system and governance conditions	Static pre-deployment approval	Broader than model-performance monitoring	Reassess policy intensity and controls	Trigger criteria require empirical testing	Proposed indicators are not validated thresholds

Transparency is therefore treated here as an enabling condition for scrutiny and challenge rather than as evidence that a decision is fair, comprehensible to every audience, or accountable in practice. The framework later connects transparency to review and contestability, but those connections remain proposed cross-domain relationships requiring testing.

Human review and override

Human review is often invoked as a safeguard without specifying what the human can actually know or do. Field research on professionals using opaque AI for critical judgments shows that engagement depends on situated expertise and on how users manage uncertainty and opacity in practice [13]. This finding cautions against treating human involvement as a uniform control. A reviewer who lacks relevant information or domain competence may formally satisfy a workflow requirement while adding little independent scrutiny.

Human involvement can also create new dependence. Research on human–AI collaboration shows that AI advice may improve individual judgments while simultaneously reducing diversity in human assessments and eroding unique collective knowledge [14]. In organizational settings, this possibility is important because repeated convergence on the same recommendation may weaken the independence that oversight is intended to preserve. Human review should therefore be evaluated not simply by whether a person appears in the process, but by whether the arrangement sustains meaningful capacity to depart from automated advice.

Control over algorithmic output also changes behavior. Experimental work finds that people become more willing to use imperfect algorithms when permitted even limited modification [15]. This supports a narrower claim that modification authority can affect reliance; it does not demonstrate that override rights automatically improve fairness, accuracy, or accountability. The organizational design problem is therefore to distinguish genuine review authority from symbolic permission.

This article proposes that meaningful human review requires five conditions: relevant competence, access to decision-relevant information, sufficient time, legitimate authority to act, and organizational permission to disagree or escalate. Override is defined as usable capacity to pause, modify, reject, or refer an AI-supported decision rather than merely annotate it. Taken together, evidence on situated professional engagement, convergence toward AI advice, and modification rights shows that human involvement is context-dependent rather than inherently protective [13–15]. The proposed conditions require empirical testing across occupations, decision types, and power structures and should not be interpreted as a validated checklist.

Contestability and redress

Contestability concerns whether an affected person can do more than receive an explanation: the person must have a usable route for questioning an AI-influenced decision and obtaining substantive reconsideration. Organizational evidence cautions against treating acceptance as a proxy for legitimacy. Reactions to algorithmic managerial decisions vary with task characteristics; people evaluated algorithmic authority less favorably in some tasks requiring human skills, showing that perceived fairness and trust are context-sensitive [16]. These individual perceptions do not establish whether a decision is substantively fair, but they indicate that the organizational meaning of algorithmic authority differs across decision settings.

Explanation also does not eliminate the need for challenge. Experimental research on personnel selection found that reactions to algorithmic versus human decision makers and the effects of explanations differed across study conditions [17]. A policy that treats explanation as sufficient redress may therefore confuse information provision with a mechanism capable of changing an outcome. Organizational voice research adds a further boundary: opportunities to speak do not themselves guarantee that expressed concerns receive consequential influence [18].

This article consequently proposes a contestability chain comprising accessible notice of the challenge route, an opportunity to present grounds or relevant information, reconsideration by a competent and appropriately empowered reviewer, a reasoned organizational response, and correction, remedy, or escalation where warranted. “Redress” refers to the possible organizational response to a substantiated challenge, not to a guaranteed reversal. The sequence is a proposed governance mechanism rather than an empirically validated appeal design. Its effectiveness may depend on accessibility, power relationships, reviewer independence, time costs, and whether affected parties can challenge information or assumptions without organizational retaliation.

Table 2 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for proposed mechanisms, relationships, and organizational consequences in responsible artificial intelligence for transparent and contestable organizational decision-making.

Table 2. Proposed mechanisms, relationships, and organizational consequences in Responsible Artificial Intelligence for Transparent and Contestable Organizational Decision-Making.

Mechanism or Relationship	Antecedent	Mediator or Process	Moderator	Expected Consequence	Rival Explanation	Validation Requirement
Risk-Proportionate Governance, Proposed	Consequential AI-influenced decision	Classification directs safeguard intensity	Task, power, reversibility	More differentiated scrutiny	Administrative burden rather than risk may drive differentiation	Comparative organizational testing
Purpose-Bound Transparency, Proposed	Need for scrutiny	Role-relevant information supports understanding	Stakeholder purpose, expertise	More usable scrutiny	More information may overload rather than clarify	Experiments and field studies comparing disclosure designs
Independent Review, Proposed	AI recommendation	Human reassessment and possible departure	Expertise, opacity, authority	Reduced symbolic oversight	Human disagreement may introduce error	Compare review structures and decision outcomes
Consequential Contestability, Proposed	Challenge to decision	Reconsideration plus reasoned response	Task type, reviewer authority	Challenge can influence process	Satisfaction may reflect procedure without substantive correction	Trace challenges, responses, and outcomes longitudinally
Accountability Continuity, Proposed	Delegated algorithmic role	Named organizational ownership	Governance structure	Reduced responsibility diffusion	Responsibility may remain ambiguous despite formal assignment	Process studies of enacted responsibility
Adaptive Revision, Proposed	Material governance signal	Reassessment of policy controls	Organizational capability	Revised policy configuration where warranted	Change may reflect politics or resources rather than learning	Longitudinal policy and audit research

The table therefore separates reported mechanisms from cross-domain propositions. It does not imply that any expected consequence has been demonstrated across organizations.

Accountability allocation

Contestability becomes organizationally consequential only if somebody is answerable for what happens after a challenge. Algorithmic delegation can obscure this point because responsibility may be attributed to a model, vendor, technical team, manager, or reviewer without specifying who owns the organizational decision. Ethical analysis of algorithms rejects the idea that delegation removes responsibility for consequential outcomes [19]. The relevant policy issue is therefore not whether responsibility exists, but how answerability is allocated when multiple actors contribute.

Research on accountable AI further identifies gaps created when algorithmic systems complicate traditional arrangements of explanation, oversight, and control, especially in consequential public decisions [20]. Although institutional arrangements differ across sectors, the broader organizational implication is that technical delegation requires countervailing checks rather than a presumption that automated processes are self-accounting. Accountability should also be distinguished from transparency: making a system visible does not specify who must respond or who possesses authority to alter practice.

Conceptual work on AI accountability emphasizes answerability together with the institutional capacity to interrogate and constrain power [21]. Building on this distinction, the article proposes five organizational roles: a decision owner answerable for the governed decision; a system owner responsible for system operation and change management; a data steward responsible for relevant data governance; a human reviewer responsible for substantive reconsideration when review is required; and a governance or audit function responsible for independent scrutiny of the arrangement.

This is a proposed organizational allocation, not a determination of legal liability. Roles may overlap in smaller organizations, and professional or statutory duties may supersede internal design. Accountability allocation should additionally preserve escalation when role conflicts arise, because nominal ownership without authority to correct a system or decision can reproduce the same diffusion the policy is intended to prevent. The central boundary is that distributed responsibility must not become unowned responsibility: the framework requires identifiable decision ownership even when technical and review tasks are distributed.

Proposed responsible artificial intelligence policy framework

The proposed framework integrates principles, decision structures, and organizational safeguards without treating any single literature stream as validation of the whole architecture. Ethical AI frameworks provide normative objectives for socially beneficial and non-maleficent use [22]. Organizational responsible-AI scholarship additionally emphasizes that adverse sociotechnical consequences and organizational enactment must be taken seriously rather than relegated to technical model properties [23]. These foundations motivate policy design but do not determine one universally correct organizational configuration.

Decision authority can also be arranged in different human–AI structures depending on decision contingencies [24]. Hybrid-intelligence scholarship similarly argues that human and machine capabilities may be complementary rather than merely substitutive [25]. Neither proposition establishes that adding a human creates legitimate oversight. The framework therefore makes authority explicit and links it to the safeguards required by the earlier risk classification.

Figure 1 presents the original conceptual synthesis for responsible artificial intelligence policy framework, showing how the article's principal constructs, mechanisms, uncertainties, and decision boundaries are connected.

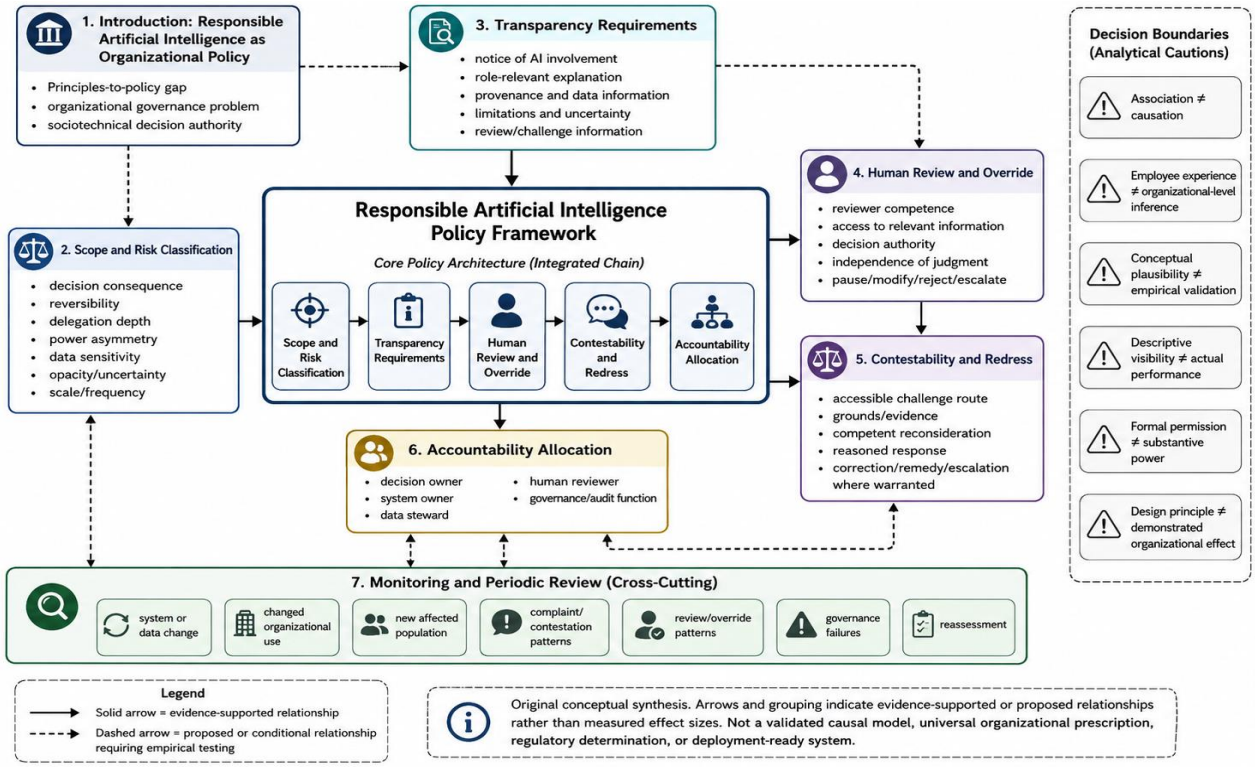

Figure 1. Responsible Artificial Intelligence Policy Framework

The framework's central proposed rule is risk-proportionate coupling. As organizational consequence, power asymmetry, irreversibility, delegation, or opacity increase, transparency, meaningful human review, contestability, and accountability should be considered jointly rather than as interchangeable safeguards. A transparent system without effective review can remain difficult to challenge; a challenge route without empowered reconsideration can be symbolic; review without accountability can leave responsibility diffuse. These are proposed cross-domain relationships, not demonstrated causal effects.

This coupling principle also prevents safeguard substitution. For example, additional explanation should not be assumed to compensate for absent decision authority, and retaining a nominal reviewer should not be treated as a substitute for an accessible challenge process. Which configurations are necessary or proportionate remains an empirical question and may vary substantially by institutional context.

Monitoring forms a recursive outer layer. Material evidence about system change, use context, challenges, overrides, or governance failures may justify reassessment of classification and policy intensity. The architecture thus treats responsible AI policy as revisable organizational governance, while leaving the effectiveness and appropriate strength of each connection open to empirical testing.

Monitoring and periodic review

Responsible AI policy should not end with initial approval because implementation tools and organizational conditions can change after adoption. Scholarship translating AI ethics from principles into practice identifies heterogeneous tools and

methods for operationalization across development and use, while also showing that availability of a tool does not establish its effectiveness [26]. Monitoring therefore concerns whether governance arrangements remain fit for their stated purpose, not merely whether a technical system continues to operate.

Ethics-based auditing offers one mechanism for structured evaluation of automated decision-making systems, but its technical, social, and institutional limitations prevent an audit from functioning as ethical certification [27]. Related work argues for auditing that is continuous, constructive, and oriented toward the AI system and its surrounding processes [28]. These insights support recurrent review while leaving open which indicators or intervals are appropriate for different organizational decisions.

This article proposes review triggers including material changes to a model or data source, changed organizational use, new affected populations, recurring complaints, unusual challenge patterns, repeated overrides, and evidence that assigned governance roles are not functioning as intended. None is a validated alarm threshold. A trigger initiates inquiry; it does not itself establish harm or require a predetermined deployment decision.

Table 3 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for testable propositions, evaluation criteria, and practical responsibilities.

Table 3. Testable propositions, evaluation criteria, and practical responsibilities.

Proposition Or Criterion	Unit And Context	Observable Indicator	Evidence Required	Falsification Condition	Organizational Responsibility	Misuse Risk	Readiness Boundary
P1 Risk-Proportionate Coupling, Proposed	Organizational decision	Safeguards vary with documented decision risk	Comparative policy and outcome data	Isolated safeguards perform equivalently across risk conditions	Decision owner and governance function	Treating dimensions as a score	Not a validated risk threshold
P2 Purpose-Bound Transparency, Proposed	Stakeholder–decision interface	Information matches review or challenge task	Experiments plus field evaluation	Generic disclosure performs as well across stakeholder purposes	System owner and decision owner	Disclosure overload or strategic opacity	Transparency does not prove fairness
P3 Meaningful Review, Proposed	Human–AI decision workflow	Reviewer can access information and depart from AI output	Behavioral and process evidence	Formal human presence performs equivalently without authority	Decision owner and reviewer	Rubber-stamping disguised as oversight	Requires competence, time, and authority
P4 Consequential Contestability, Proposed	Affected party and organization	Challenge receives competent reconsideration and reasoned response	Longitudinal challenge records and qualitative evidence	Channel availability alone produces equivalent influence	Decision owner and review function	Symbolic voice	No guaranteed reversal or remedy
P5 Non-Transferable Accountability, Proposed	Organizational governance	Named answerability persists after delegation	Process and comparative governance research	Responsibility becomes unnecessary without loss of control	Senior decision owner and governance function	Responsibility diffusion	Not a legal-liability determination
P6 Adaptive Policy, Proposed	System and organizational lifecycle	Material change triggers documented reassessment	Longitudinal audits and policy histories	Static classification remains equally adequate after material change	System owner, data steward, governance function	Continuous review becoming ceremonial compliance	Review signals are not deployment criteria

Operational tools and recurrent auditing can support continuing governance, but their existence does not establish responsible outcomes or certify an AI system as ethical [26–28]. Periodic review should therefore be evaluated by whether it produces reasoned reassessment when material conditions change, not by the volume of monitoring activity.

Implementation implications

Implementation requires organizational capability as well as formal policy. Qualitative research on AI readiness identifies assets, capabilities, and organizational commitment as important conditions for adoption [29]. For the present framework, readiness is narrower than successful adoption: an organization may possess technical capability while lacking reviewers with sufficient authority, reliable escalation pathways, or governance capacity to investigate challenges. Implementation should therefore assess whether the roles specified by policy can actually perform their assigned functions.

Human-resource management provides an especially consequential illustration. AI applications in HR confront data limitations, fairness and accountability questions, employee reactions, and institutional constraints that complicate direct transfer of technical solutions into organizational decisions [30]. This does not imply that HR is uniquely risky; rather, it

demonstrates why policy must be tied to decision context, affected interests, and the organizational authority surrounding algorithmic outputs.

Formal adoption also cannot be equated with enacted practice. Field research on algorithm introduction shows how symbolic actions can reshape organizational regimes of knowing and the authority attached to algorithmic knowledge [31]. A written requirement for review may therefore coexist with informal deference; an appeal mechanism may exist while employees regard challenge as futile; responsibility may be assigned while everyday decisions migrate toward technical specialists or model outputs.

Implementation should consequently focus on enactment: resources for review, role clarity, documentation, accessible challenge procedures, governance independence, and opportunities to detect divergence between formal policy and actual practice. Scarce review capacity, conflicting incentives, or excessive workload may weaken these functions even when formal responsibilities are clearly documented, making implementation quality a separate question from policy design. These are proposed implementation implications, not a deployment manual. Organizational readiness does not demonstrate responsible outcomes, and conformity with the framework would not by itself establish fairness, legality, decision quality, or implementation success.

Conclusion

This Original Organizational Policy Framework Article reframes responsible AI as the governance of organizational decision authority rather than the adoption of isolated ethical principles or technical controls. Its principal contribution is a proposed architecture in which scope and risk classification organize coupled transparency, meaningful human review and override, contestability and redress, and accountability allocation, while monitoring enables reassessment as technology and organizational conditions change.

The framework deliberately separates disclosure from intelligibility, human presence from substantive authority, voice opportunity from consequential challenge, and distributed participation from identifiable accountability. Its propositions are intended to make these distinctions empirically testable rather than to present them as demonstrated organizational effects.

The framework remains bounded by the heterogeneity of decision contexts, institutions, occupations, technologies, and power relations from which its conceptual foundations are drawn. Future research should test alternative configurations, examine enactment over time, identify conditions under which safeguards reinforce or undermine one another, and determine whether the proposed architecture improves relevant organizational outcomes. Until such evidence accumulates, the framework should be understood as a theoretically grounded policy proposition rather than a validated causal model, legal standard, or universal implementation prescription.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Papagiannidis E, Mikalef P, Conboy K. Responsible artificial intelligence governance: A review and research framework. J Strateg Inf Syst. 2025;34(2):101885. doi:10.1016/j.jsis.2024.101885
2. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell. 2019;1(9):389-99. doi:10.1038/s42256-019-0088-2
3. Mittelstadt B. Principles alone cannot guarantee ethical AI. Nat Mach Intell. 2019;1(11):501-7. doi:10.1038/s42256-019-0114-4
4. Berente N, Gu B, Recker J, Santhanam R. Managing artificial intelligence. MIS Q. 2021;45(3):1433-50. doi:10.25300/MISQ/2021/16274
5. Raisch S, Krakowski S. Artificial intelligence and management: The automation–augmentation paradox. Acad Manage Rev. 2021;46(1):192-210. doi:10.5465/amr.2018.0072
6. Janssen M, Brous P, Estevez E, Barbosa LS, Janowski T. Data governance: Organizing data for trustworthy artificial intelligence. Gov Inf Q. 2020;37(3):101493. doi:10.1016/j.giq.2020.101493
7. Leicht-Deobald U, Busch T, Schank C, Weibel A, Schafheitle S, Wildhaber I, et al. The challenges of algorithm-based HR decision-making for personal integrity. J Bus Ethics. 2019;160(2):377-92. doi:10.1007/s10551-019-04204-w

8. Kellogg KC, Valentine MA, Christin A. Algorithms at work: The new contested terrain of control. Acad Manage Ann. 2020;14(1):366-410. doi:10.5465/annals.2018.0174
9. Parent-Rocheleau X, Parker SK. Algorithms as work designers: How algorithmic management influences the design of jobs. Hum Resour Manag Rev. 2022;32(3):100838. doi:10.1016/j.hrmr.2021.100838
10. Ananny M, Crawford K. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media Soc. 2018;20(3):973-89. doi:10.1177/1461444816676645
11. Langer M, Oster D, Speith T, Hermanns H, Kästner L, Schmidt E, et al. What do we want from explainable artificial intelligence (XAI)? A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. Artif Intell. 2021;296:103473. doi:10.1016/j.artint.2021.103473
12. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell. 2019;1(5):206-15. doi:10.1038/s42256-019-0048-x
13. Lebovitz S, Lifshitz-Assaf H, Levina N. To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. Organ Sci. 2022;33(1):126-48. doi:10.1287/orsc.2021.1549
14. Fügener A, Grahl J, Gupta A, Ketter W. Will humans-in-the-loop become borgs? Merits and pitfalls of working with AI. MIS Q. 2021;45(3):1527-56. doi:10.25300/MISQ/2021/16553
15. Dietvorst BJ, Simmons JP, Massey C. Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. Manage Sci. 2018;64(3):1155-70. doi:10.1287/mnsc.2016.2643
16. Lee MK. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. Big Data Soc. 2018;5(1):2053951718756684. doi:10.1177/2053951718756684
17. Wesche JS, Hennig F, Kollhed CS, Quade J, Kluge S, Sonderegger A. People's reactions to decisions by human vs. algorithmic decision-makers: The role of explanations and type of selection tests. Eur J Work Organ Psychol. 2024;33(2):146-57. doi:10.1080/1359432X.2022.2132940
18. Morrison EW. Employee voice and silence: Taking stock a decade later. Annu Rev Organ Psychol Organ Behav. 2023;10:79-107. doi:10.1146/annurev-orgpsych-120920-054654
19. Martin K. Ethical implications and accountability of algorithms. J Bus Ethics. 2019;160(4):835-50. doi:10.1007/s10551-018-3921-3
20. Busuioc M. Accountable artificial intelligence: Holding algorithms to account. Public Adm Rev. 2021;81(5):825-36. doi:10.1111/puar.13293
21. Novelli C, Taddeo M, Floridi L. Accountability in artificial intelligence: What it is and how it works. AI Soc. 2024;39(4):1871-82. doi:10.1007/s00146-023-01635-y
22. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. AI4People—an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds Mach. 2018;28(4):689-707. doi:10.1007/s11023-018-9482-5
23. Mikalef P, Conboy K, Lundström JE, Popovič A. Thinking responsibly about responsible AI and 'the dark side' of AI. Eur J Inf Syst. 2022;31(3):257-68. doi:10.1080/0960085X.2022.2026621
24. Shrestha YR, Ben-Menahem SM, von Krogh G. Organizational decision-making structures in the age of artificial intelligence. Calif Manage Rev. 2019;61(4):66-83. doi:10.1177/0008125619862257
25. Dellermann D, Ebel P, Söllner M, Leimeister JM. Hybrid intelligence. Bus Inf Syst Eng. 2019;61(5):637-43. doi:10.1007/s12599-019-00595-2
26. Morley J, Floridi L, Kinsey L, Elhalal A. From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. Sci Eng Ethics. 2020;26(4):2141-68. doi:10.1007/s11948-019-00165-5
27. Mökander J, Morley J, Taddeo M, Floridi L. Ethics-based auditing of automated decision-making systems: Nature, scope, and limitations. Sci Eng Ethics. 2021;27(4):44. doi:10.1007/s11948-021-00319-4
28. Mökander J, Floridi L. Ethics-based auditing to develop trustworthy AI. Minds Mach. 2021;31(2):323-7. doi:10.1007/s11023-021-09557-8
29. Jöhnk J, Weißert M, Wyrtki K. Ready or not, AI comes—an interview study of organizational AI readiness factors. Bus Inf Syst Eng. 2021;63(1):5-20. doi:10.1007/s12599-020-00676-7
30. Tambe P, Cappelli P, Yakubovich V. Artificial intelligence in human resources management: Challenges and a path forward. Calif Manage Rev. 2019;61(4):15-42. doi:10.1177/0008125619867910
31. Pachidi S, Berends H, Faraj S, Huysman M. Make way for the algorithms: Symbolic actions and change in a regime of knowing. Organ Sci. 2021;32(1):18-41. doi:10.1287/orsc.2020.1377