E-ISSN: 3108-4176

APSSHS

Academic Publications of Social Sciences and Humanities Studies
2026, Volume 7, Issue 1, Page No: 163-172
Available online at: https://apsshs.com/

Annals of Organizational Culture, Leadership and External Engagement Journal

Governing AI-Enabled Organizational Transformation: An Ethical Control Architecture for Human Oversight, Decision Accountability, Stakeholder Transparency, and Responsible Innovation

Wouter De Smet[1]*, Liesbeth Van Dam[1], Pieter Janssens[2]

1. Department of AI Governance and Ethical Control, Faculty of Business, KU Leuven, Leuven, Belgium.
2. Department of Human Oversight and Responsible Innovation, Faculty of Management, Ghent University, Ghent, Belgium.

Abstract

Artificial intelligence is increasingly embedded in organizational decisions, workflows, professional judgment, and transformation programs, yet responsible-AI debates often separate technical risk management from the organizational allocation of authority, accountability, transparency, and stakeholder recourse. This original non-empirical article develops an ethical control architecture for governing AI-enabled organizational transformation. It synthesizes recent management, information-systems, and business-ethics scholarship to distinguish nominal human involvement from meaningful control, technical assurance from organizational responsibility, disclosure from contestability, and one-time compliance from lifecycle governance. The analysis argues that responsible transformation depends on coordinated control across strategic, managerial, operational, and technical layers rather than on a single ethics principle, oversight role, or technical safeguard. The proposed architecture integrates four mutually dependent control functions: preserving substantive human judgment where ethically consequential decisions require it; maintaining traceable responsibility for distributed human–AI decisions; enabling stakeholder transparency, challenge, and redress; and sustaining auditability and revision across the AI life cycle. It further treats escalation, intervention, and accountability as context-sensitive rather than universally fixed. The contribution is a governance-oriented synthesis that makes responsibility allocation and contestability visible as organizational design problems. The architecture is not presented as an empirically validated causal model, a universal implementation template, or evidence of managerial effectiveness. Its usefulness depends on organizational context, decision stakes, institutional setting, system properties, and future empirical testing across levels and over time.

Keywords: Responsible AI governance, Human oversight, Algorithmic accountability, Stakeholder contestability, Organizational transformation, Responsible innovation

How to cite this article: De Smet W, Van Dam L, Janssens P. Governing AI-Enabled Organizational Transformation: An Ethical Control Architecture for Human Oversight, Decision Accountability, Stakeholder Transparency, and Responsible Innovation. Ann Organ Cult Leadersh Extern Engagem J. 2026;7(1):163-72. https://doi.org/10.51847/ZyogEEulsN

Received: 29 March 2026; **Revised:** 12 June 2026; **Accepted:** 16 June 2026
Corresponding author: Wouter De Smet
E-mail ✉ wouter.desmet@kuleuven.be

Introduction

AI-enabled organizational transformation is not reducible to adoption of a more capable information system. Theory-driven management research increasingly treats generative AI as an organizational phenomenon that changes managerial work, coordination, and the assumptions through which firms organize decision processes [1]. The ethical question therefore begins before any individual model output: it begins with how organizations redesign work around AI.

Digital transformation research similarly distinguishes technological adoption from the strategic and structural changes through which organizations respond to digital technologies [2]. That distinction matters ethically because authority, information flows, performance criteria, and stakeholder exposure can change even when the technical system itself appears unchanged. Governance must therefore attend to transformation arrangements as well as model properties.

AI also unsettles simple substitution logic. Management theory characterizes automation and augmentation as interdependent rather than mutually exclusive, creating persistent tensions over which tasks should be delegated, retained, or jointly performed [3]. Ethical control cannot assume that either maximum automation or maximum human involvement is inherently superior; the relevant issue is how roles and responsibility are configured.

Recent responsible-AI governance scholarship accordingly emphasizes structural, relational, and procedural organizational practices [4]. Building on that shift, this article proposes an ethical control architecture that connects meaningful human oversight, decision accountability, stakeholder transparency and contestability, and responsible innovation across the AI life cycle. The architecture is offered as an evidence-bounded synthesis requiring empirical validation, not as a proven governance mechanism.

AI-enabled transformation as an organizational ethics problem

AI-enabled management redistributes tasks, discretion, and interaction between human and computational actors. An integrative management review shows that AI changes managerial roles and interaction patterns across organizational levels rather than functioning only as a detached analytical tool [5]. Ethical consequences therefore arise from organizational redesign choices, including who can question outputs, who bears responsibility, and whose interests enter the decision process.

The normative vocabulary for such choices is relatively mature but not operationally uniform. Comparative analysis of AI ethics guidelines finds recurring principles alongside substantial variation in how those principles are interpreted and implemented [6]. Convergence around fairness, transparency, responsibility, or related values should therefore not be mistaken for convergence in organizational practice.

This gap is not solved merely by issuing additional principles. Principles can orient ethical reasoning, but they do not by themselves guarantee ethical AI practice or specify the institutional mechanisms needed for implementation [7]. The organizational problem is thus translational: values must be converted into decision rights, review routines, documentation, challenge channels, and revisable controls.

Nor are ethical risks exclusively technical. Theoretical work on algorithmic bias shows how formal organizational rationalities and use contexts can help constitute bias rather than merely transmit defects originating inside data or models [8]. AI-enabled transformation is therefore an ethics problem precisely because technical systems become embedded in organizational purposes, classifications, incentives, and authority structures.

From technical risk to organizational responsibility

Technical assurance remains necessary, but realized consequences depend on how organizations combine human and algorithmic capabilities. Field experimental evidence indicates that different human–AI configurations can produce different outcomes [9]. This does not establish a universal superior arrangement; it supports treating configuration itself as an organizational responsibility because organizations choose who receives AI advice, who may override it, and how decisions are reviewed.

Responsible-AI governance research likewise locates governance in structures, processes, and relationships rather than in isolated model checks [4]. Responsibility therefore extends to the design of decision environments: assignment of authority, access to reasons and records, escalation pathways, and mechanisms for learning when controls fail. The proposed architecture treats these arrangements as ethical control points rather than administrative details.

A purely technical framing can also hide the organizational rationalities that shape how algorithmic categories are constructed and used [8]. Consequently, responsibility cannot be discharged by demonstrating model quality alone. Organizations must remain answerable for the purposes for which systems are deployed, the decision rules surrounding them, and the consequences of embedding algorithmic outputs in institutional practice.

Human oversight and meaningful control

Meaningful oversight requires more than placing a person somewhere in a workflow. Evidence on human–AI collaboration distinguishes roles involving automation, augmentation, and different allocations of task authority, showing why the substance of the human role matters [10]. Oversight should therefore specify what judgment the human must exercise, what information is available, and what authority exists to pause, reject, or alter an AI-supported course of action.

Nominal involvement may even coexist with weakened responsibility. Experimental evidence shows that autonomy-restricting algorithms can facilitate displacement of responsibility and unethical behavior under the studied conditions [11]. The finding is individual-level and context-bound, but it directly cautions against treating a human signature or approval step as proof of meaningful control.

Oversight must also accommodate the continuing automation–augmentation tension rather than impose a single fixed balance [3]. In the proposed architecture, meaningful control means preserving decision-relevant human agency, competence, and

intervention authority where ethical stakes warrant them, while allowing lower-friction automation where judgment is not materially displaced. These are governance criteria to be tested, not universally validated thresholds.

Table 1 translates this distinction into an article-specific hazard register showing where nominal human involvement can fail to preserve substantive oversight.

Table 1. Ethical-control hazard register for preserving substantive human oversight in AI-enabled decisions

Decision context or hazard	Ethical risk	Human judgment preserved	Control point	Intervention authority	Transparency / contestability need	Failure signature	Review record
Routine AI recommendation	Rubber-stamping	Independent assessment	Reasoned comparison	Decision owner	Rationale plus challenge path	Automatic acceptance	Output, rationale, override log
Autonomy-restricting workflow	Responsibility displacement	Discretion to refuse	Pause/override gate	Accountable supervisor	Visible authority and challenge	"System required it"	Authority map, override use
High-stakes ranking/classification	Bias or context loss	Exception judgment	Pre-action exception review	Trained reviewer	Reasons plus appeal path	Unexplained adverse action	Inputs, reasons, exceptions
Expert AI-assisted judgment	Hollow professional review	Domain reasoning	Deliberative sign-off	Expert or committee	Reason-giving and query route	Copied model rationale	Dissent, final reasoning

Note: Entries are proposed control translations synthesized from the preceding evidence. They do not constitute empirically validated control-effectiveness claims.

Decision accountability and responsibility allocation

Accountability becomes difficult when organizations describe algorithmic outcomes as though responsibility migrated into the artifact. Business-ethics analysis instead locates accountability for algorithmic decisions in human and organizational actors who design, deploy, authorize, and rely on those systems [12]. The relevant question is therefore not whether the AI is "responsible," but how answerability is allocated among actors with different authority and contributions.

The literature on AI accountability is itself multidimensional. Recent field mapping identifies technical, legal, ethical, organizational, and governance streams that remain only partly integrated [13]. A control architecture should consequently avoid a single-owner fiction and distinguish responsibility for system design, deployment approval, operational use, oversight, stakeholder response, and corrective action.

Importantly, collaboration with AI is not ethically harmful in every configuration. Experimental work in marketing decisions reports conditions in which collaboration with algorithms reduced managers' violations of ethical norms [14]. Such evidence is a useful counterweight to uniformly adverse accounts, but it does not demonstrate organization-wide ethical effectiveness or establish that AI assistance is intrinsically beneficial.

Responsibility can also weaken when authority is restricted. The observed displacement associated with autonomy-restricting algorithms indicates that formal involvement without meaningful discretion may reduce felt ownership of outcomes [11]. The proposed architecture therefore links accountability to actual decision rights, intervention capacity, and traceable reasoning, while allowing responsibility to remain distributed where multiple actors materially shape a decision.

Stakeholder transparency and contestability

Transparency should be treated as a relational governance requirement rather than an undifferentiated demand for more information. Empirical evidence indicates that moral violation can shape transparency expectations and responses to algorithmic decision making [15]. Disclosure requirements should therefore be sensitive to decision stakes and stakeholder vulnerability rather than assuming that the same explanation depth is appropriate in every context.

Process also matters independently of outcomes. Research on perceived legitimacy of algorithmic decisions shows that legitimacy judgments are influenced by both decision process and outcome [16]. Perceived legitimacy is not equivalent to normative legitimacy, but the finding supports designing governance so affected stakeholders can understand relevant procedures, identify responsible actors, and question how a consequential decision was reached.

High-level transparency principles alone do not specify those capacities. Cross-guideline evidence shows variation in the interpretation and implementation of commonly invoked AI ethics principles [6]. This article therefore distinguishes transparency from contestability: transparency supplies intelligible information, whereas contestability requires a viable route to question, challenge, or seek reconsideration of a decision. Neither concept should be equated automatically with effective redress.

Responsible innovation across the ai life cycle

Ethical control must persist after initial approval. A systematic review of AI audits and auditability frames auditability as something that must be made possible through system and governance design, supporting a lifecycle orientation rather than a purely retrospective audit event [17]. Records, responsibilities, and review access therefore need to be created while systems are selected, configured, deployed, monitored, modified, and potentially retired.

Lifecycle control also requires a mode for handling disagreement about what counts as responsible innovation. Deliberative governance scholarship argues that responsible AI innovation can require stakeholder participation and reason-giving around contested choices [18]. Deliberation does not guarantee consensus or remove power asymmetries, but it supports treating disagreement as a governance input rather than a communication failure.

Generative AI sharpens the temporal problem because use patterns and system behavior may evolve after deployment. Conceptualizing GenAI as a complex adaptive system supports governance that remains adaptive and revisable rather than assuming a one-time compliance decision will remain adequate [19]. The proposed architecture accordingly links auditability, stakeholder challenge, monitoring, and redesign as recursive controls, without claiming that any fixed review frequency or intervention threshold is universally optimal.

The ethical control architecture

AI governance becomes organizationally consequential through implementation arrangements rather than through policy language alone. Evidence from public organizations shows that governance is enacted through organizational practices and structures [20]. Although sector transfer requires caution, this supports locating ethical control in concrete allocations of authority, review, documentation, escalation, and responsibility rather than in statements of principle detached from operating routines.

The same concern is acute in expert work, where the ethical value of human involvement depends on preserving substantive judgment rather than ceremonial approval. Business-ethics analysis of AI deployment in auditing emphasizes deliberative approaches that preserve professional reasoning and responsibility [21]. The architecture therefore treats human judgment as a governed capability: it must have information, competence, time, and authority sufficient to matter.

At the institutional level, AI discourse has increasingly moved from a "rush to ethics" toward a "race for governance," while coordination and implementation challenges remain unresolved [22]. This shift supports integrating governance mechanisms, but it does not establish that one architecture will work across organizations, sectors, or jurisdictions.

The article therefore proposes an ethical control architecture organized around four interdependent functions: meaningful human oversight, traceable decision accountability, stakeholder transparency and contestability, and lifecycle responsible innovation. These functions are coordinated across strategic, managerial, operational, and technical governance layers and connected through intervention, escalation, audit, challenge, and revision loops. The integration extends existing responsible-AI governance scholarship [4], but the specific architecture is a proposed synthesis requiring multilevel and longitudinal validation.

Figure 1 shows how the proposed architecture keeps these control functions distinct while linking them across governance layers, intervention points, stakeholder challenge, and lifecycle revision.

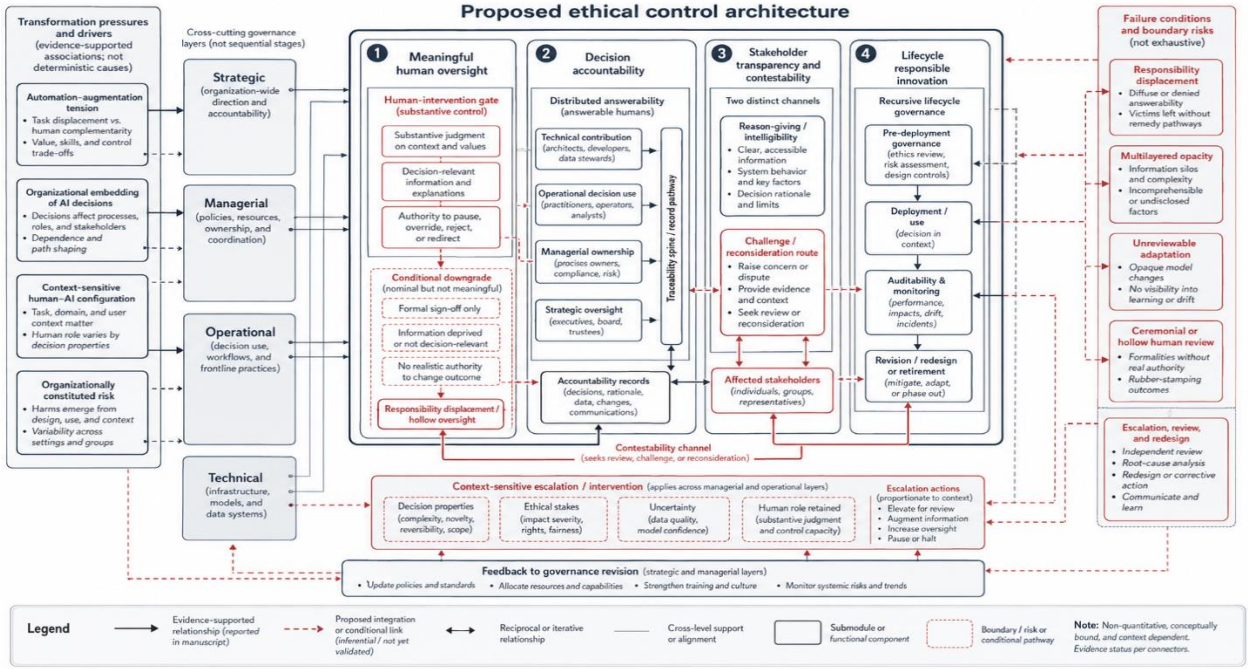

Figure 1. Ethical control links human oversight and accountability across strategic, managerial, operational, and technical governance layers

Governance layers: strategic, managerial, operational, and technical

The proposed architecture treats ethical control as a multilevel organizational responsibility rather than a technical function delegated to one specialist team. AI changes the structure of organizational decision making by altering how information, authority, and judgment can be distributed between humans and computational systems [23]. Strategic governance should therefore define acceptable purposes and accountability ownership, while managerial governance translates those commitments into decision rights, supervision, and escalation responsibilities.

At the operational level, AI-enabled control can reshape autonomy, monitoring, and worker responses. Research on algorithms at work shows that algorithmic systems can reorganize workplace control and become sites of contestation rather than neutral productivity devices [24]. Operational governance consequently needs mechanisms for exception handling, local challenge, and review of how formal controls are actually experienced and enacted.

The technical layer remains indispensable but should not be treated as the sole location of opacity or risk. Evidence from public-sector algorithmic decision making demonstrates that blackboxing can accumulate across technical, organizational, and institutional layers [25]. The proposed allocation therefore links technical records and system constraints upward to managerial and strategic answerability, while preserving operational routes for challenge and intervention. **Table 2** makes those distinct decision rights and traceability obligations visible without implying that one allocation is universally optimal.

Table 2. Accountability allocation across strategic, managerial, operational, and technical AI governance layers

Layer	Decision right	Responsible actor	Oversight / challenge actor	Traceability artifact	Escalation trigger	Stakeholder exposure	Responsibility-gap risk
Strategic	Authorize purpose and risk appetite	Board / executive owner	Board committee / ethics oversight	Mandate, risk rationale	Purpose conflict or material harm	Organization-wide	Accountability without operational visibility
Managerial	Configure roles and control rules	Business / governance lead	Independent risk or compliance function	Role map, control decision	Repeated exceptions or unresolved challenge	Unit / process	Diffused ownership
Operational	Apply, question, override	Decision professional / process owner	Supervisor / review panel	Decision and override record	Uncertainty, anomaly, contested outcome	Directly affected stakeholder	Rubber-stamping or responsibility displacement
Technical	Build, monitor, constrain system behavior	Technical owner	Assurance / audit function	Version, test, monitoring record	Drift, failure, unreviewable change	Indirect to direct	Technical opacity severed from answerability

Note: The allocations are proposed governance translations. Actors and escalation conditions require contextual specification and are not validated as universally effective.

Escalation and human-intervention thresholds

Escalation is most defensible when it is tied to characteristics of the decision rather than to a symbolic requirement that a human approve every output. Strategic reasoning scholarship distinguishes prediction from forms of causal and theoretical reasoning that may demand capabilities not supplied by pattern prediction alone [26]. This does not establish general human superiority, but it supports heightened intervention where decisions depend on causal interpretation, counterfactual judgment, or contestable assumptions.

Behavioral willingness to delegate is an insufficient ethical threshold. Experimental evidence indicates that delegation to AI is shaped by loss aversion and decision framing [27]. An organization should therefore not infer that employees' willingness to delegate proves that delegation is appropriate, legitimate, or responsibility-preserving.

The proposed architecture instead makes escalation conditional on decision stakes, uncertainty, reversibility, stakeholder exposure, evidentiary conflict, and the substantive capabilities retained by human decision makers. Human–AI collaboration research reinforces that "human involvement" encompasses different roles, including automation and augmentation configurations [10]. Intervention should therefore specify what the human can know, question, change, or stop. No universal numerical cutoff is proposed; thresholds remain organizational judgments that should be documented, reviewable, and revised as use conditions change.

Accountability across distributed human–AI decisions

Distributed decision making creates a risk that perceived agency and formal responsibility no longer align. Workplace research on an anthropomorphized AI colleague shows that social and affective framing can influence how employees interpret AI

agency and relationships [28]. Such perceptions do not themselves create a formal accountability gap, but they can complicate who workers experience as acting, advising, or controlling a decision.

AI assistance can also change strategic judgments while humans retain formal decision authority. Evidence from entrepreneurs and investors demonstrates that AI can influence strategic decision making without converting the system into the final accountable decision maker [29]. Governance should therefore distinguish informational contribution from authorization, implementation, and answerability.

Substantive accountability further requires more than identifying a name after harm occurs. Business-ethics scholarship emphasizes engagement and rational discourse as part of managing algorithmic accountability rather than reducing accountability to reputational response [30]. The proposed architecture accordingly links each material AI contribution to a responsible human or organizational role, an inspectable decision record, an escalation route, and a stakeholder-facing challenge mechanism. **Figure 2** represents this as a recursive accountability process rather than a one-way chain.

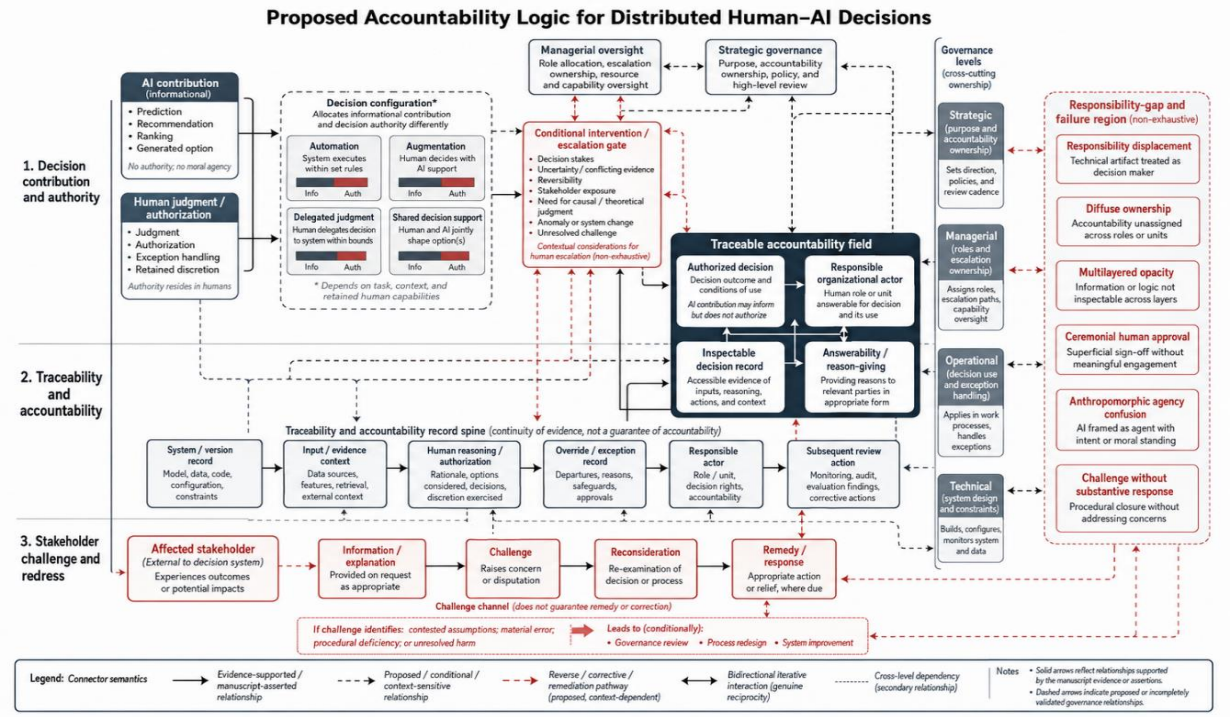

Figure 2. Traceable intervention, challenge, and redress sustain accountability across distributed human–AI decisions

Stakeholder challenge and redress mechanisms

A challenge channel is ethically weak if stakeholders can submit objections but the organization has no obligation to engage with them substantively. Work on corporate–rightsholder remedy argues for dialogical engagement with marginalized stakeholders rather than treating notice or consultation as equivalent to remedy [31]. Applied here, that insight supports a proposed distinction among receiving a complaint, reconsidering a decision, explaining a response, and providing meaningful remedy.

Deliberative governance offers a complementary rationale for reason-giving and stakeholder voice around contested AI choices [18]. Yet deliberation should not be romanticized: access, expertise, institutional power, and the ability to influence outcomes may remain unequal. The architecture therefore treats contestability as a governed capacity rather than proof that a process is legitimate.

Challenge design should also recognize that process affects how algorithmic decisions are judged. Perceived legitimacy is sensitive to both decision process and outcome [16]. Perceived legitimacy is not the same as normative legitimacy, but the evidence reinforces why challenge procedures should be intelligible, timely, attributable to a responsible actor, and capable of producing an answer. Redress remains context-sensitive and should not be inferred merely from the existence of an appeal channel.

Ethical failure modes and responsibility gaps

Ethical control can fail even when formal safeguards appear to be present. Experimental audit evidence shows that providing humans with input can affect their reliance on AI through perceived control [32]. The direction and magnitude are task-specific, but the finding demonstrates why participation itself should not be treated as calibrated oversight. A control can create confidence without securing independent judgment.

Responsibility failure is also plural. Conceptual analysis identifies multiple responsibility gaps associated with AI rather than one generic accountability deficit [33]. For organizational governance, the implication is diagnostic: failures may arise because responsibility is displaced, ownership is diffuse, control is insufficient, records cannot connect actions to actors, or affected stakeholders cannot obtain substantive answers. The specific organizational categories proposed here require empirical testing.

A particularly important failure occurs when human involvement becomes ceremonial. Evidence that autonomy-restricting algorithms can facilitate responsibility displacement shows how a nominally human decision can coexist with weakened ownership of action [11]. The architecture therefore treats hollow review, multilayered opacity, diffuse answerability, and unresponsive contestability as different failure signatures. Distinguishing them matters because adding another approval step would not necessarily address the underlying mechanism.

Boundary conditions

The architecture is deliberately conditional. Human–AI outcomes vary with how decision authority, discretion, and delegation are configured [9, 11, 27]. These studies examine different outcomes and settings, so they do not justify one optimal arrangement. They support the narrower proposition that governance should treat configuration as a contextual variable rather than assuming that either more automation or more human involvement is always ethically preferable.

Institutional setting also matters. Multilayered blackboxing observed in public-sector algorithmic decision making shows that opacity can arise through interactions among technical systems, organizational processes, and institutional arrangements [25]. The mechanism is analytically important, but its prevalence and exact form cannot be generalized to every organization or jurisdiction.

Time is a further boundary. Conceptualizing generative AI as a complex adaptive system emphasizes that system behavior, user practices, and governance conditions may evolve after deployment [19]. Controls that are defensible at authorization may therefore require revision as systems or uses change. The architecture does not specify universal monitoring intervals or escalation thresholds; transfer requires attention to technology type, sector, decision stakes, stakeholder vulnerability, regulatory context, organizational capacity, level of analysis, and temporal change.

Contributions to business ethics and ai governance

The first contribution is integrative. Responsible-AI governance research already identifies multiple structural, relational, and procedural governance practices [4]. This article reorganizes those premises around ethical control: what organizational arrangements preserve meaningful judgment, keep responsibility traceable, make stakeholder challenge consequential, and permit lifecycle revision. The resulting architecture is proposed rather than empirically validated.

Second, it treats accountability as a sociotechnical relationship rather than a search for a single culpable artifact or actor. Existing scholarship rejects the idea that algorithmic systems sever human responsibility and emphasizes answerability, engagement, and differentiated responsibility gaps [12, 30, 33]. The proposed contribution is to connect those insights explicitly to distributed decision rights, escalation, traceability, and redress without asserting one universal responsibility-allocation rule.

Third, the architecture links normative responsibility to inspectable governance capability. Lifecycle auditability provides a practical bridge because accountability claims are difficult to examine when systems, decisions, overrides, and revisions leave no usable record [17]. Auditability is not ethical sufficiency, however; it is one enabling condition within a wider architecture that also requires judgment, responsible actors, stakeholder voice, and context-sensitive intervention.

Implications for boards, managers, and ai governance functions

For boards and senior executives, the principal implication is to govern purpose, authority, and accountability ownership rather than treating responsible AI as a technical compliance exercise. Evidence on organizational AI governance shows that implementation depends on structures and practices, although sector transfer requires caution [20]. Board-level attention should therefore connect approved use purposes to identifiable owners and review routes.

Managers need explicit allocation of decision rights and escalation ownership. Organization-design research on AI-enabled decision structures supports separating human and AI roles rather than leaving authority implicit [23]. The article does not prescribe one centralized or decentralized model.

AI governance functions should also examine whether challenge channels produce substantive responses, particularly where stakeholders face power asymmetries. Dialogical remedy scholarship supports attention to responsiveness rather than formal access alone [31]. These implications are governance considerations, not evidence that the proposed arrangements will necessarily improve ethical or organizational outcomes.

Future research

The architecture creates testable questions rather than validated prescriptions. AI accountability research remains distributed across technical, organizational, legal, and ethical streams [13]. Future studies should test whether integrating these domains improves traceability, perceived responsibility, contestability, or other clearly specified outcomes, while avoiding composite measures that obscure distinct mechanisms.

Field experiments can compare alternative allocations of AI advice, human discretion, review, and escalation. Existing field experimentation demonstrates that human–AI configurations can be manipulated and evaluated [9], but future work should include responsibility and stakeholder-oriented outcomes rather than performance alone.

Multilevel qualitative and process studies are also needed. Research on multilayered blackboxing provides a precedent for tracing how opacity accumulates across organizational and institutional layers [25]. Replication across private, public, professional, and platform settings would clarify transferability.

Finally, longitudinal designs should examine governance adaptation as systems and organizational uses evolve. Dynamic accounts of GenAI governance make temporal change theoretically salient [19]. Such research should allow the architecture to be refined, rejected, or differentiated by technology class rather than assuming stable effectiveness.

Limitations

This article is non-empirical and integrates heterogeneous literatures. Responsible-AI governance scholarship itself remains an evolving field [4], so the proposed architecture should not be interpreted as a validated causal model or an implementation standard.

Normative convergence also has limits. Global AI-ethics guidelines share recurring principles while differing in interpretation and operationalization [6]. Similar terminology therefore cannot establish equivalent organizational enactment across jurisdictions.

The evidence also contains context-specific positive findings. Algorithmic collaboration has reduced ethical-norm violations in one managerial decision context [14], but that result does not establish organization-wide or universal benefit.

Finally, workplace research on anthropomorphized AI shows context-sensitive affective and relational responses [28]. Such evidence can illuminate perceived agency, but it cannot establish that anthropomorphism necessarily produces formal accountability failure. The architecture therefore requires validation across sectors, levels, cultures, technologies, and time.

Conclusion

AI-enabled organizational transformation creates ethical governance problems because authority, judgment, information, and responsibility can become distributed across people, systems, and organizational layers. This article proposes an ethical control architecture that keeps meaningful human oversight, traceable decision accountability, stakeholder contestability and redress, and lifecycle responsible innovation analytically distinct while connecting them through multilevel governance and conditional escalation. Its central claim is organizational rather than technological: ethical control depends on how decision rights, records, intervention authority, and answerability are arranged around AI use. The architecture is neither a universal template nor a validated mechanism. Its boundaries include institutional context, decision stakes, stakeholder vulnerability, technology characteristics, temporal change, and level of analysis. Its value therefore lies in providing a structured set of distinctions and relationships that future empirical research can test, refine, differentiate, or reject.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Brown O, Davison RM, Decker S, Ellis DA, Faulconbridge J, Gore J, et al. Theory-driven perspectives on generative artificial intelligence in business and management. Br J Manag. 2024;35(1):3-23. doi:10.1111/1467-8551.12788
2. Vial G. Understanding digital transformation: A review and a research agenda. J Strateg Inf Syst. 2019;28(2):118-44. doi:10.1016/j.jsis.2019.01.003
3. 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

4. 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
5. Hillebrand L, Raisch S, Schad J. Managing with artificial intelligence: An integrative framework. Acad Manage Ann. 2025;19(1):343-75. doi:10.5465/annals.2022.0072
6. 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
7. Mittelstadt B. Principles alone cannot guarantee ethical AI. Nat Mach Intell. 2019;1(11):501-7. doi:10.1038/s42256-019-0114-4
8. Nishant R, Schneckenberg D, Ravishankar MN. The formal rationality of artificial intelligence-based algorithms and the problem of bias. J Inf Technol. 2024;39(1):19-40. doi:10.1177/02683962231176842
9. Krakowski S, Haftor D, Luger J, Pashkevich N, Raisch S. Human-centered artificial intelligence: A field experiment. Manage Sci. 2026;72(1):57-72. doi:10.1287/mnsc.2022.03849
10. Fügener A, Walzner DD, Gupta A. Roles of artificial intelligence in collaboration with humans: Automation, augmentation, and the future of work. Manage Sci. 2026;72(1):538-57. doi:10.1287/mnsc.2024.05684
11. Jolly AA, Dunlop PD, Parker SK, Kanse L. It's not my responsibility: Working with autonomy-restricting algorithms facilitates unethical behavior and displacement of responsibility. J Bus Ethics. 2026;203(2):405-24. doi:10.1007/s10551-025-06034-5
12. Martin K. Ethical implications and accountability of algorithms. J Bus Ethics. 2019;160(4):835-50. doi:10.1007/s10551-018-3921-3
13. Bartsch SC, Nguyen LH, Schmidt JH, Du G, Adam M, Benlian A, et al. The present and future of accountability for AI systems: A bibliometric analysis. Inf Syst Front. 2025;27(6):2463-84. doi:10.1007/s10796-025-10636-9
14. Gaczek P, Leszczyński G, Wei Y, Sun H. The bright side of AI in marketing decisions: Collaboration with algorithms prevents managers from violating ethical norms. J Bus Ethics. 2026;204(4):771-94. doi:10.1007/s10551-025-06083-w
15. Shah MU, Rehman U, Parmar B, Ismail I. Effects of moral violation on algorithmic transparency: An empirical investigation. J Bus Ethics. 2024;193(1):19-34. doi:10.1007/s10551-023-05472-3
16. Martin K, Waldman A. Are algorithmic decisions legitimate? The effect of process and outcomes on perceptions of legitimacy of AI decisions. J Bus Ethics. 2023;183(3):653-70. doi:10.1007/s10551-021-05032-7
17. Li Y, Goel S. Making it possible for the auditing of AI: A systematic review of AI audits and AI auditability. Inf Syst Front. 2025;27(3):1121-51. doi:10.1007/s10796-024-10508-8
18. Buhmann A, Fieseler C. Deep learning meets deep democracy: Deliberative governance and responsible innovation in artificial intelligence. Bus Ethics Q. 2023;33(1):146-79. doi:10.1017/beq.2021.42
19. Janssen M. Responsible governance of generative AI: Conceptualizing GenAI as complex adaptive systems. Policy Soc. 2025;44(1):38-51. doi:10.1093/polsoc/puae040
20. de Almeida PGR, dos Santos Júnior CD. Artificial intelligence governance: Understanding how public organizations implement it. Gov Inf Q. 2025;42(1):102003. doi:10.1016/j.giq.2024.102003
21. Sinha VK, Derichs D. Reshaping the space of ethics for expert work: Deliberative approaches for deploying artificial intelligence in auditing. J Bus Ethics. 2026;206(2):213-36. doi:10.1007/s10551-025-06166-8
22. Koniakou V. From the "rush to ethics" to the "race for governance" in artificial intelligence. Inf Syst Front. 2023;25(1):71-102. doi:10.1007/s10796-022-10300-6
23. 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
24. 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
25. Kronblad C, Essén A, Mähring M. When justice is blind to algorithms: Multilayered blackboxing of algorithmic decision making in the public sector. MIS Q. 2024;48(4):1637-62. doi:10.25300/MISQ/2024/18251
26. Felin T, Holweg M. Theory is all you need: AI, human cognition, and causal reasoning. Strategy Sci. 2024;9(4):346-71. doi:10.1287/stsc.2024.0189
27. Bockstedt JC, Buckman JR. Humans' use of AI assistance: The effect of loss aversion on willingness to delegate decisions. Manage Sci. 2026;72(1):323-42. doi:10.1287/mnsc.2024.05585
28. Einola K, Khoreva V, Tienari J. A colleague named Max: A critical inquiry into affects when an anthropomorphised AI (ro)bot enters the workplace. Hum Relat. 2024;77(11):1620-49. doi:10.1177/00187267231206328
29. Csaszar FA, Ketkar H, Kim H. Artificial intelligence and strategic decision-making: Evidence from entrepreneurs and investors. Strategy Sci. 2024;9(4):322-45. doi:10.1287/stsc.2024.0190
30. Buhmann A, Paßmann J, Fieseler C. Managing algorithmic accountability: Balancing reputational concerns, engagement strategies, and the potential of rational discourse. J Bus Ethics. 2020;163(2):265-80. doi:10.1007/s10551-019-04226-4

31. Bianchi L, Caruana R, Shivji AK. Engaging marginalized stakeholders: Towards a dialogical theorization of effective corporate-rightsholder remedy. J Bus Ethics. 2025;201(3):605-19. doi:10.1007/s10551-024-05879-6
32. Commerford BP, Eilifsen A, Hatfield RC, Holmstrom KM, Kinserdal F. Control issues: How providing input affects auditors' reliance on artificial intelligence. Contemp Account Res. 2024;41(4):2134-62. doi:10.1111/1911-3846.12974
33. Santoni de Sio F, Mecacci G. Four responsibility gaps with artificial intelligence: Why they matter and how to address them. Philos Technol. 2021;34(4):1057-84. doi:10.1007/s13347-021-00450-x