Excel International Journal of
Technology, Engineering & Management
Journal homepage: exceljournals.org.in
ISSN ONLINE: https://portal.issn.org/resource/ISSN/2349-8455
ISSN PRINT: https://portal.issn.org/resource/ISSN/2348-0378

Beyond Human-in-the-Loop: Designing Accountable Organizations for AI-Augmented Decision-Making — The HILO-A Framework

Ignacio Bonasa Alzuria
Conceptual Research Paper
Technology Management
Organizational System Design
AI Governance

Corresponding Author: Ignacio Bonasa Alzuria

DOI Link: https://doi.org/10.64971/j.cph.eijtem.v13.i3.18.2026

Abstract
Artificial intelligence is moving from the edge of organizational life into its decision core. Systems that once supported people now classify, rank, predict, recommend, generate, and in some settings execute actions. This shift creates a management problem that cannot be solved by model accuracy alone: organizations must decide where machine intelligence may act, where human judgment must remain consequential, and who is accountable when a decision is produced by a hybrid human–AI chain. To address the gap between high-level governance principles and operational decision rights, this paper proposes the Human-in-the-Loop Organizational Architecture (HILO-A), a conceptual management framework built around five connected layers: Data, Algorithm, Human Judgment, Decision, and Accountability. It also introduces four modes of human–AI interaction—Assisted Intelligence, Augmented Intelligence, Supervised Automation, and Autonomous Execution—and a Decision Criticality Matrix based on impact, reversibility, uncertainty, and ethical sensitivity. The paper is deliberately managerial rather than purely technical. It treats AI adoption as a redesign of decision rights, escalation paths, exception handling, and responsibility. The framework is positioned alongside current governance thinking, including the NIST AI Risk Management Framework, the OECD AI Principles, and the European Union AI Act. The central proposition is that mature AI organizations will not be defined by maximum automation, but by the quality with which they allocate intelligence and responsibility. HILO-A offers a practical vocabulary and implementation protocol for managers who need to make that allocation explicit, auditable, and adaptable.
Keywords: Artificial Intelligence; Human-in-the-Loop; AI Governance; Organizational Design; Decision-Making; Management Architecture; Human–AI Collaboration; Accountability; System Design.

1. Introduction

Artificial intelligence is changing a deceptively ordinary part of management: who gets to decide. For decades, most enterprise technologies were easy to place conceptually. A database stored information. A spreadsheet calculated. An enterprise resource planning system standardized a process. A dashboard displayed indicators. The manager, professional, or employee remained visibly responsible for interpreting the information and deciding what should happen next. AI makes that boundary less obvious. A contemporary AI system can screen a large set of alternatives, predict an outcome, rank candidates, flag anomalies, draft a response, recommend a course of action, or trigger a workflow. Generative AI adds another layer because it can produce persuasive language and apparently coherent reasoning at a speed that encourages reliance. The organizational question is therefore no longer limited to whether AI can improve productivity. It is whether an organization has designed a sensible relationship between computational capability and decision authority.

That relationship matters because technical performance and organizational legitimacy are not the same thing. A model may be statistically impressive and still be used in the wrong context. A recommendation

may be accurate on average and still be inappropriate in a particular case. A human reviewer may technically remain "in the loop" while having neither the time, competence, authority, nor psychological freedom to challenge the machine. Conversely, excessive human checking can erase the productivity gains that motivated automation in the first place.

This paper addresses that design problem. It proposes HILO-A: the Human-in-the-Loop Organizational Architecture. The framework is intended for managers, transformation leaders, governance professionals, system designers, and organizations adopting AI-supported decision processes. Its aim is not to add another abstract list of ethical principles. It is to translate human oversight and accountability into an organizational architecture that can be mapped, discussed, assigned, and reviewed.

The paper makes three contributions. First, it reframes AI transformation as decision transformation. Second, it proposes a five-layer architecture connecting data, algorithms, human judgment, decisions, and accountability. Third, it introduces a criticality-based method for choosing among different levels of automation. Together, these contributions shift the discussion from the binary question "human or machine?" toward the more useful question "what combination of human and machine intelligence is appropriate for this decision, under these conditions, with this accountability?"

2. Literature Survey and Theoretical Background

Organizations often begin AI programmes by identifying use cases: customer service, fraud detection, recruitment, forecasting, maintenance, marketing, knowledge search, coding, document analysis, or internal productivity. This is reasonable, but it can hide a deeper change. Each use case contains one or more decisions. When AI enters the workflow, it changes how information reaches those decisions and may change who—or what—has influence over the outcome.

Consider a simple example. A customer-service model recommends whether a complaint should be escalated. At first sight, the system merely prioritizes work. In practice, the recommendation may influence waiting time, compensation, managerial attention, and customer retention. If employees routinely follow the recommendation, the algorithm has acquired practical decision power even if the formal policy says that humans remain responsible. Decision architecture therefore concerns real influence, not merely the labels written in a procedure.

This distinction becomes more important as AI is embedded in chains of action. One model may summarize information, another score risk, a workflow engine may apply a threshold, and a human may see only the final recommendation. Accountability can become fragmented across data owners, vendors, developers, business units, compliance teams, and frontline users. The danger is not that nobody participated. The danger is that everybody participated partially and responsibility became difficult to locate.

The NIST AI Risk Management Framework treats AI risk management as an organizational activity spanning the lifecycle of AI systems and emphasizes governance, mapping, measurement, and management of risk. The OECD AI Principles likewise stress human agency and oversight, transparency, robustness, and accountability. In the European Union, Article 14 of the AI Act requires effective human oversight for high-risk AI systems and explicitly contemplates the ability of authorized people to understand limitations, avoid over-reliance, interpret outputs, disregard or override outputs, and intervene in system operation. These developments point in the same direction: oversight must be operational, not ceremonial.

HILO-A takes this emerging governance expectation and asks a management-design question: what would an organization look like if meaningful oversight were built into its decision architecture from the beginning rather than added after deployment?

Research gap. Existing approaches provide indispensable guidance on AI risk, human oversight, responsible deployment, human-centred design, and automation–augmentation. Yet they do not, by themselves, provide a compact organizational architecture that makes decision authority traceable across the full chain from data and algorithmic output to human judgment, action, override, and accountability. HILO-A addresses this organizational-design gap. Its intended contribution is not another list of AI principles, but a decision-rights architecture that managers can map, govern, audit, and revise as risk and evidence change.

3. Materials and Methods: Conceptual Research Design

This is a conceptual research paper. It does not claim to report a new empirical experiment, survey, or field study. Its contribution is theory-building through synthesis and managerial design. That distinction is important for research integrity: no interviews, participant data, statistical results, or validation outcomes are invented in this manuscript.

The framework was developed through a transparent conceptual synthesis rather than a claimed systematic review. The source base was purposively organized around four bodies of knowledge directly relevant to the research problem: (1) human–automation interaction and automation bias; (2) augmentation and hybrid intelligence in management; (3) human-centered AI and human–computer interaction; and (4) authoritative AI governance frameworks. Constructs recurring across these bodies—oversight, role clarity, risk proportionality, traceability, intervention, and accountability—were extracted and translated into organizational design questions. The five HILO-A layers were then derived as the minimum decision chain required to connect informational input to organizational consequence and identifiable responsibility. The four interaction modes were used to represent increasing delegation of decision and execution authority. Finally, the framework was stress-tested conceptually against illustrative management scenarios and converted into falsifiable propositions for future empirical research.

The approach is therefore abductive and design-oriented. It begins with an observed organizational tension, uses established governance concepts as constraints, and proposes a parsimonious architecture intended to be useful enough for practice and specific enough for later research. The framework is not presented as a replacement for legal compliance, technical assurance, cybersecurity, data governance, or domain expertise. It is a managerial layer connecting those disciplines to decision design.

3.1 Literature Synthesis: From Automation to Hybrid Intelligence

Research on human–automation interaction predates the current generative-AI wave and provides an important warning against treating oversight as a binary variable. Parasuraman and Riley (1997) distinguished use, misuse, disuse, and abuse of automation, while Parasuraman, Sheridan, and Wickens (2000) showed that automation can be allocated at different stages of information acquisition, analysis, decision selection, and action implementation. This tradition matters because the managerial question is not whether a human is present somewhere in the process, but what cognitive and organizational function that person still performs.

More recent management scholarship reframes the issue as an automation–augmentation paradox. Raisch and Krakowski (2021) argue that automation and augmentation are interdependent rather than mutually exclusive. Jarrahi (2018) similarly describes human–AI symbiosis in organizational decision-making, while Dellermann et al. (2019) use the concept of hybrid intelligence to emphasize systems in which human and machine capabilities complement one another. HILO-A builds on this stream but shifts the unit of analysis from capability complementarity to the architecture of decision rights and accountability.

Human-centered AI research adds a second foundation. Shneiderman (2022) argues for designs that combine high levels of human control with high levels of automation rather than assuming a simple trade-off. Interaction research also shows that the quality of the interface matters: people need understandable feedback, clear system status, appropriate explanations, and routes to correct or contest machine behavior. This makes human oversight partly an organizational-design and partly a human–computer-interaction problem.

A third stream concerns trust and reliance. People can reject algorithms after observing errors, but they can also over-rely on automated recommendations when systems appear authoritative or when workload makes independent evaluation costly. The practical implication is uncomfortable but important: adding a human reviewer can create the appearance of safety without creating meaningful challenge. HILO-A therefore treats information, competence, time, and authority as minimum conditions for consequential human judgment.

Recent evidence makes that reliance problem more nuanced. Jussupow, Benbasat, and Heinzl (2024) show that algorithm aversion and appreciation are not opposite anomalies but context-dependent preferences that vary with the decision configuration. Lu et al. (2024), synthesizing 442 effect sizes from 163 studies, similarly report that willingness to prefer AI depends strongly on perceived capability and the need for personalization. Quasi-experimental evidence on human–generative-AI collaboration also suggests that AI can reduce cognitive burden in information-intensive situations while creating new risks of over-reliance and context loss (2024). The implication for HILO-A is that governance should not assume either automatic trust or automatic resistance: the architecture must design for appropriate reliance.

Organizational research is also moving from abstract responsible-AI principles toward implementation. Recent work on AI governance emphasizes that roles, cross-functional coordination, transparency, and continuous adaptation are necessary to translate external principles into operational practice (2025). Research on human oversight under the EU AI Act further cautions that human intervention can support

fairness and accountability but does not guarantee either outcome by itself. This strengthens the case for treating oversight as a designed organizational capability rather than a final approval step.

Finally, contemporary governance frameworks converge on accountability and proportional oversight. NIST's AI RMF explicitly calls for clear roles and responsibilities for human–AI configurations and oversight. The OECD AI Principles emphasize human agency, oversight, traceability, and accountability. Article 14 of the EU AI Act requires human oversight for high-risk AI systems to be commensurate with risk, autonomy, and context, and includes the ability to understand limitations, avoid automation bias, interpret outputs, disregard or override them, and intervene in operation. These frameworks establish the governance direction; HILO-A seeks to operationalize that direction at the level of a concrete organizational decision.

3.2 The Unresolved Organizational Gap

Existing approaches are strong on risk principles, lifecycle governance, or human–machine interaction, but managers still face a practical translation problem. A governance framework may say that roles must be clear without specifying how a particular credit, hiring, maintenance, compensation, or strategic decision should be partitioned between data, model, reviewer, executor, and accountable owner. A human-centered design principle may call for control without specifying who owns the consequences when control is distributed across several functions.

HILO-A addresses this narrower gap. It is not offered as a competing regulatory framework, technical standard, or universal ethics code. It is a meso-level management architecture: more concrete than a principle, less technical than a model card, and focused specifically on the flow of organizational decisions. Its novelty therefore lies in integration—linking decision criticality, human authority, override, execution, and accountability in one map that can be used before and after deployment.

Table 1. Positioning HILO-A against adjacent approaches

Approach	Primary focus	Human oversight	Decision rights	Override logic	Named accountability
NIST AI RMF	Lifecycle AI risk management	Explicit	Addressed broadly	Supported	Explicit governance roles
OECD AI Principles	Human-centric policy principles	Explicit	Principle-level	Supported conceptually	Explicit
EU AI Act, Art. 14	Legal oversight for high-risk AI	Mandatory in scope	Context/risk dependent	Explicit disregard/override/intervention	Provider/deployer duties
Human-Centered AI	Safe, usable human control	Core design concern	Interaction-oriented	Strong human control	Varies by implementation
Automation–Augmentation	Division/complementarity of work	Analytical focus	Central tension	Not primary focus	Not primary focus
HILO-A	Organizational decision architecture	Operational layer	Explicitly mapped	Logged and governed	Named decision-architecture owner

Note. The comparison describes primary emphasis rather than exclusivity. HILO-A is designed to complement, not replace, legal, technical, or enterprise risk-management frameworks.

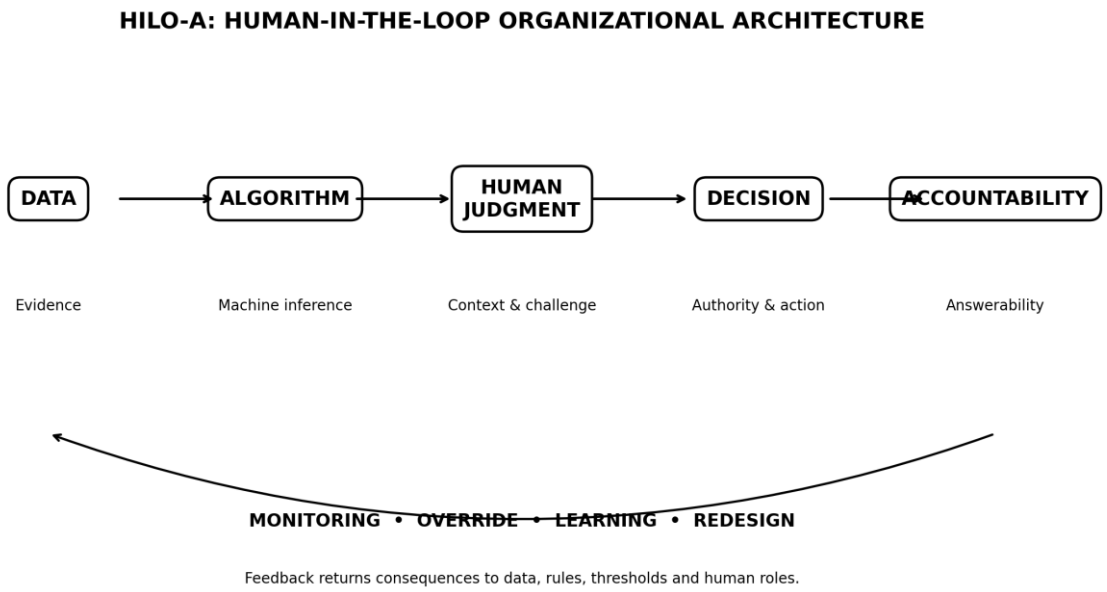

Figure 1. HILO-A organizational decision architecture. Accountability, monitoring, override, and feedback operate across the full decision chain.

4. Results and Analysis: The Proposed HILO-A Framework

HILO-A consists of five connected layers: Data → Algorithm → Human Judgment → Decision → Accountability. The arrow is intentionally simple. In real organizations the flow is iterative, but the sequence forces managers to ask five questions in order: What information is being used? What is the machine doing with it? Where does human judgment genuinely enter? What action follows? Who remains answerable for the result?

Layer	Core question	Typical failure if neglected
1. Data	What evidence enters the system?	Poor quality, bias, missing context, uncertain provenance.
2. Algorithm	What is the system authorized to infer or generate?	Capability silently becomes authority.
3. Human Judgment	Where can a person interpret, challenge or contextualize?	Rubber-stamp oversight or automation bias.
4. Decision	Who/what converts analysis into action?	Unclear decision rights and escalation.
5. Accountability	Who owns the consequences and learning?	Responsibility dissolves across the system.

4.1 Layer One: Data

AI-supported decisions inherit the strengths and weaknesses of their information environment. Data quality is therefore not a preliminary technical detail; it is part of the decision itself. Managers should know which data are decisive, how current they are, what populations or situations they represent, and which relevant realities are absent.

The human contribution begins here. Domain experts can identify whether a clean dataset is nevertheless contextually incomplete. A customer may look risky because a model lacks information about an exceptional event. An employee profile may look unusual because historical data encode yesterday's job design. A forecast may appear precise while its underlying assumptions have become obsolete. Data governance and human judgment are consequently linked before an algorithm produces any output.

4.2 Layer Two: Algorithm

The algorithmic layer concerns both capability and permission. A system may be capable of predicting, ranking, recommending, generating, or executing. Management must separately determine what it is

permitted to do. This separation is crucial because organizations can drift from experimentation into de facto delegation without an explicit decision to delegate.

HILO-A recommends an authorization statement for material AI use cases: the system's intended function, prohibited functions, acceptable inputs, output status, confidence or uncertainty considerations where available, and conditions requiring escalation. Such a statement turns a technical component into a governed organizational actor without pretending that the system itself bears moral or legal responsibility.

4.3 Layer Three: Human Judgment

Human judgment is valuable when it adds something the system cannot reliably capture: context, tacit knowledge, competing values, empathy, strategic interpretation, exceptional circumstances, or responsibility for trade-offs. The purpose of a human-in-the-loop design is not to insert a person everywhere. It is to locate the moments where human contribution is consequential.

Meaningful oversight requires more than a click on an approval button. The reviewer needs sufficient information to understand the output, enough time to evaluate it, competence appropriate to the decision, and authority to disagree. A nominal reviewer who is punished for slowing a workflow or who sees only an opaque score is not meaningfully exercising judgment.

4.4 Layer Four: Decision

The decision layer converts analysis into organizational consequence. HILO-A distinguishes five possible treatments of an AI output: inform, recommend, require human authorization, execute with human supervision, or execute autonomously within defined boundaries. Naming these treatments makes decision rights visible. It also prevents a common ambiguity in which a "recommendation" becomes practically mandatory because nobody is expected to challenge it.

4.5 Layer Five: Accountability

Accountability closes the architecture. The core principle is simple: an organization may automate an action without automating away responsibility. For each material AI-supported decision, there should be an identifiable owner of the decision architecture. That person or governance body does not need to personally review every transaction, but it should own the rules governing delegation, monitoring, escalation, and redesign.

The Accountability Non-Delegation Principle. Organizations may delegate analysis, prediction, recommendation, prioritization, and bounded execution to artificial intelligence, but they cannot delegate organizational accountability to the system itself. The more operational authority a system receives, the more explicit the human or institutional ownership of its mandate, monitoring, escalation, and consequences should become.

This layer also turns failures into learning. When an override occurs, a model behaves unexpectedly, or a decision creates harm, the organization should be able to reconstruct the chain: data, system output, human intervention, action, and accountable owner. Traceability is therefore not merely an audit requirement; it is organizational memory.

5. Four Modes of Human-AI Interaction

Mode	AI role	Human role	Suitable context
I. Assisted Intelligence	Provides information or analysis	Makes the decision	Novel, ambiguous or high-judgment decisions.
II. Augmented Intelligence	Produces a recommendation	Interprets and authorizes	Complex decisions with useful predictive support.
III. Supervised Automation	Decides/executes within rules	Monitors exceptions and performance	High-volume, bounded, reasonably reversible processes.
IV. Autonomous Execution	Decides and executes	Governs ex ante and audits ex post	Low-criticality, standardized, reversible actions.

These modes are not a maturity ladder. Moving toward autonomy is not automatically progress. A sophisticated organization may deliberately keep a high-impact decision at Mode II while automating thousands of low-impact decisions at Mode IV. Maturity lies in fit, not in maximal delegation.

The modes can also change over time. A new AI application may begin as assisted intelligence while the organization learns its limitations. With evidence, monitoring, and stable operating conditions, some decisions may move to supervised automation. Conversely, incidents, environmental change, or deteriorating model performance may justify moving a process back toward stronger human involvement.

6. The HILO-A Decision Criticality Matrix

HILO-A proposes that the degree of human involvement should be determined by decision criticality rather than technological enthusiasm. Four dimensions are assessed: impact, reversibility, uncertainty, and ethical sensitivity. Each can be rated from 1 (low) to 5 (very high). The purpose is not to create false mathematical precision. The score creates a structured conversation and makes assumptions explicit.

Dimension	1 – Low	3 – Medium	5 – Very high
Impact	Minor operational effect	Material business/stakeholder effect	Severe financial, safety, rights or strategic consequence
Reversibility	Easy and cheap to reverse	Correctable with meaningful cost	Difficult or impossible to reverse
Uncertainty	Stable context, known patterns	Some volatility or incomplete information	Novel/ambiguous environment, weak predictability
Ethical sensitivity	Minimal effect on people	Meaningful fairness/trust implications	Rights, access, livelihood, safety or dignity at stake

A simple Criticality Index can be expressed as CI = I + R + U + E, producing a range from 4 to 20. A low score does not prove that autonomous execution is safe, and a high score does not prohibit AI support. Rather, the index acts as a governance trigger. Organizations may define their own thresholds; for example, 4–8 may permit bounded autonomy, 9–13 may favour supervised automation, 14–17 may require human authorization, and 18–20 may require enhanced review, dual control, or executive/governance involvement. The proposed bands are heuristic design aids only; they are not empirically validated cut-off values and should not be represented as predictive thresholds. Sector-specific regulation, safety requirements, and professional judgment take precedence.

Methodological caution: the Criticality Index is a heuristic decision-support device, not a psychometrically validated scale, universal risk score, or empirically established cut-off system. Its proposed thresholds are illustrative governance triggers. Sector-specific validation should precede any use of the index as a formal classification instrument.

The matrix embodies proportionality. It aligns conceptually with the EU AI Act's risk-sensitive approach to human oversight and with risk-management frameworks that call for controls to reflect context and potential harm. More importantly for managers, it prevents the automation discussion from being dominated by the question of what the technology can do.

7. The Human Override Principle

A credible human-in-the-loop system must allow disagreement. HILO-A therefore includes the Human Override Principle: where a human is assigned meaningful oversight, the organization must define the conditions, authority, and recording mechanism through which the person can disregard, reverse, pause, or escalate an AI output.

Override is not a sign that AI has failed. It is evidence that the architecture is functioning as designed when context falls outside normal patterns. The more interesting question is what the organization learns from overrides. Repeated overrides may reveal model drift, poor thresholds, missing data, a badly designed interface, or a mismatch between the use case and the technology.

At the same time, unlimited informal override creates its own risk. A reviewer may systematically reintroduce bias or ignore useful evidence. HILO-A therefore recommends an Override Log containing the AI output, human decision, reason category, decision owner, and eventual outcome when observable. The purpose is not surveillance of employees; it is comparative learning between human and machine judgment.

8. Automation Bias and the Problem of Ceremonial Oversight

Human presence does not guarantee human judgment. One of the most important practical risks is automation bias: the tendency to rely excessively on automated outputs. This risk becomes stronger when

a system is usually correct, when workloads are high, when the interface presents a recommendation with apparent certainty, or when challenging the system requires extra effort.

Ceremonial oversight occurs when an organization can truthfully say that a person approves the decision but the design makes disagreement unrealistic. Examples include reviewers processing hundreds of recommendations per hour, approval screens that hide the evidence behind a score, or performance targets that reward speed while override requires a lengthy justification.

HILO-A treats interface design, workload, training, and authority as governance variables. A useful test is counterfactual: can the assigned reviewer reasonably detect an error, understand why it matters, and change the outcome before harm occurs? If the answer is no, the organization should not describe the arrangement as meaningful human oversight.

9. From Organizational Charts to Decision Maps

Traditional organizational charts show reporting lines. They rarely show how decisions actually travel through data, software, teams, vendors, and approvals. AI makes this limitation increasingly important. HILO-A proposes a complementary artifact: the Decision Map.

A Decision Map begins with a consequential organizational decision and works backward and forward. It identifies information sources, models or AI services, transformations of the output, human reviewers, authorization points, automated actions, escalation paths, monitoring indicators, and the accountable owner. The result is not a technical architecture diagram and not an organizational chart. It is a map of decision influence.

Decision Maps can reveal hidden dependencies. A manager may discover that a supposedly human decision is almost always determined by a score produced upstream. A compliance team may discover that a vendor model is being used for a purpose not covered by the original review. A business unit may find that nobody owns the quality of the data feeding a critical recommendation.

10. HILO-A Decision Map Template

Element	Questions to document
Decision	What real-world action or consequence is produced?
Data	Which sources materially influence the output? Who owns them?
AI system	What does it predict, generate, classify or recommend?
Human checkpoint	Who reviews? What information and authority do they have?
Execution	Who/what acts? Is action reversible?
Escalation	What triggers a pause, second review, or higher authority?
Monitoring	Which indicators reveal error, drift, bias, or unexpected behaviour?
Accountable owner	Who owns the architecture and its periodic review?

11. Implementation Protocol: Seven Stages

Stage 1 — Map decisions. Start with decisions rather than tools. Identify recurring decisions, their consequences, volume, current owners, and existing controls.

Stage 2 — Classify criticality. Rate impact, reversibility, uncertainty, and ethical sensitivity. Record the reasoning, not only the number.

Stage 3 — Define the AI role. State whether AI informs, recommends, decides, or executes. Specify boundaries and prohibited uses.

Stage 4 — Define the human role. Specify who interprets, validates, overrides, monitors, or escalates. Confirm competence, time, information, and authority.

Stage 5 — Assign accountability. Name an owner for the decision architecture. Avoid collective language that makes ownership impossible to locate.

Stage 6 — Record exceptions. Capture material overrides, incidents, unexpected outputs, and cases where people overruled or relied on the system incorrectly.

Stage 7 — Learn and redesign. Periodically review evidence. Increase or reduce automation based on performance, context, risk, and organizational learning.

The protocol is intentionally iterative. AI governance should not be a one-time approval ceremony. Systems change, data distributions move, employees develop new habits, vendors update models, and business contexts evolve. A decision architecture that was appropriate at deployment may become inappropriate six months later.

12. Worked Management Scenarios

12.1 Recruitment Screening

An AI tool ranks applicants for a management role. The decision has material impact on livelihood and opportunity, may contain fairness implications, and is only partly reversible once candidates are rejected. HILO-A would normally resist autonomous rejection based solely on the score. The system can assist or augment recruiters, but humans should have access to relevant evidence, understand limitations, and be able to reconsider candidates. The accountable owner should monitor selection patterns and override behaviour.

12.2 Predictive Maintenance

A manufacturing system predicts likely equipment failure and automatically schedules inspection. If the action is low-cost and reversible, supervised automation may be appropriate. If the system instead shuts down safety-critical equipment or suppresses an alarm, criticality changes substantially. The same underlying model can therefore require different decision architectures depending on the action attached to its output.

12.3 Generative AI in Executive Reporting

A generative AI system drafts a monthly executive narrative from operational data. The text may appear authoritative while containing omissions or unsupported causal explanations. HILO-A would treat the system as assisted intelligence unless robust controls justify more. A human owner should verify material claims, especially explanations of why performance changed. The governance risk is not merely hallucination; it is the possibility that fluent language compresses uncertainty into false confidence.

12.4 Customer Compensation

An AI model recommends compensation for service failures. Small, reversible credits could be automated within a bounded range. Large payments, unusual cases, vulnerable customers, or reputationally sensitive incidents could be escalated. This illustrates a core HILO-A idea: organizations need not choose one automation level for an entire process. Decision rights can be segmented by criticality.

12.5 Strategic Investment Decisions

AI can synthesize market information, model scenarios, and challenge assumptions, but strategic investments combine uncertainty, irreversibility, competing values, and executive accountability. Here the machine should expand the decision space rather than close it. The highest value may come from using AI as a structured adversary—generating alternative hypotheses and identifying weak assumptions—while humans retain authority for the final commitment.

13. Governance Roles and Responsibilities

Role	HILO-A responsibility
Board / governing body	Set risk appetite for material AI-supported decisions and receive assurance on high-criticality uses.
Executive owner	Own business outcomes and the decision architecture.
AI / technology team	Build or integrate systems, document limitations, monitor technical performance.
Data owner	Assure provenance, quality, access, and relevant data controls.
Risk / compliance / legal	Interpret applicable obligations and challenge control design.
Human reviewer	Exercise genuine judgment within defined authority; record material overrides.
Internal audit / assurance	Test whether stated oversight exists in practice and whether controls are operating.
Employees / affected users	Provide operational feedback and surface unexpected behaviour or context.

The table deliberately separates technical ownership from decision ownership. An AI team may be responsible for a model without being the appropriate owner of the business consequence. Likewise, legal or compliance functions can advise and challenge without becoming the default owner of every AI-enabled decision.

14. Managerial Propositions for Future Testing

Decision-right clarity is the first mechanism in HILO-A. When algorithmic influence is visible and authorization is explicit, responsibility is less likely to disappear across technical and managerial actors. This leads to the first proposition.

P1. Explicit allocation of AI and human decision rights will reduce perceived and operational ambiguity in accountability for AI-supported outcomes.

The second mechanism is proportionality. Automation can create speed and consistency, but the value of human intervention rises when consequences are harder to reverse, uncertainty is greater, or ethical sensitivity is material.

P2. Decision criticality will moderate the relationship between automation and decision quality, such that higher-criticality decisions will benefit more from meaningful human intervention than lower-criticality decisions.

Oversight is only consequential when a reviewer can actually challenge the system. HILO-A therefore treats information, competence, time, and authority as jointly necessary design conditions rather than decorative controls.

P3. Human oversight will improve decision quality only when reviewers possess sufficient information, competence, time, and authority to evaluate and, where justified, challenge AI outputs.

The fourth mechanism is organizational learning. A disagreement between a person and a model is valuable evidence only if the organization captures why it occurred and what happened afterwards.

P4. Structured logging and review of material human overrides will improve organizational learning and the calibration of human–AI decision boundaries over time.

Finally, accountability affects trust because users need to know that consequential automated decisions remain contestable and institutionally owned. This connects governance design with adoption rather than treating compliance and use as separate problems.

P5. Greater clarity of accountability and contestability in AI-supported decisions will be positively associated with appropriate user trust and sustainable organizational adoption.

Table. Positioning HILO-A against major governance and human–AI perspectives

Approach	**Risk governance**	**Human oversight**	**Decision rights**	**Override/traceability**	**HILO-A distinction**
NIST AI RMF	Strong	Strong	Indirect	Strong	HILO-A maps authority across a concrete decision chain.
OECD AI Principles	Principle-level	Strong	Principle-level	Accountability principle	HILO-A operationalizes allocation of authority.
ISO/IEC 42001	Management-system level	Governance-level	Context dependent	Audit/governance oriented	HILO-A focuses on decision architecture.
Human-Centered AI	Strong	Strong	Variable	Variable	HILO-A makes accountability and override explicit.
Automation–augmentation	Strong conceptual lens	Implicit/variable	Explains allocation tension	Not central	HILO-A turns the tension into a governance design.
HILO-A	Integrated	Explicit	Explicit	Explicit	Data → Algorithm → Human Judgment → Decision → Accountability.

15. Discussion of the Conceptual Results

HILO-A is based on a simple observation: intelligence and responsibility are different organizational resources. AI can increase the amount of intelligence available to a decision by processing more information, finding patterns, or generating alternatives. Responsibility, however, cannot be meaningfully assigned to a model in the same way it is assigned within an organization. The managerial task is therefore to benefit from machine capability without allowing responsibility to evaporate.

This perspective avoids two unhelpful extremes. The first is technological maximalism: if a machine can perform a task, the organization should automate it. The second is human exceptionalism: consequential decisions must always be made entirely by people. Both positions ignore context. Humans can be inconsistent, biased, fatigued, and poorly informed. Machines can be brittle, opaque, overconfident in presentation, and insensitive to exceptional context. Good design asks which combination produces the strongest decision process.

The framework also changes the meaning of "human-centric." Human-centric AI should not mean that a human clicks last. It should mean that the organization has deliberately preserved human agency where agency matters, has made limitations visible, and has retained a route for contestation, correction, and accountability.

There is also an economic argument. Poorly designed oversight is expensive. If every low-risk output requires manual approval, AI becomes an elaborate productivity detour. If high-risk outputs receive no meaningful challenge, incidents can create legal, reputational, operational, and human costs. Proportionality is therefore not only an ethical principle; it is an efficiency principle.

Finally, HILO-A suggests that AI capability should be treated as dynamic. The correct allocation of decision rights today may not be correct tomorrow. Organizations should be able to move a decision toward greater autonomy when evidence supports it and toward greater human involvement when uncertainty or risk rises. This reversible governance posture is more realistic than a one-way automation roadmap.

16. Relationship to Contemporary AI Governance

HILO-A is compatible with, but distinct from, major governance frameworks. NIST's AI RMF provides a broad risk-management structure and emphasizes governance across the AI lifecycle. Its Generative AI Profile extends that work to risks particular to generative systems. The OECD AI Principles articulate a human-centric international policy baseline, including oversight, transparency, robustness, and accountability. The EU AI Act creates binding obligations in its jurisdiction and, for high-risk systems, makes human oversight an explicit design and deployment requirement. As of 2026, NIST is revising AI RMF 1.0 and has also issued work focused on generative AI and critical-infrastructure applications, reinforcing the need for governance approaches that can adapt as capabilities and deployment contexts change.

HILO-A does not attempt to reproduce these frameworks. Its narrower contribution is organizational: it offers a way to translate principles into a map of decision rights. A company can use NIST to structure risk management, legal analysis to determine regulatory obligations, technical standards to structure management systems, and HILO-A to ask who or what is allowed to influence and execute a particular decision.

This distinction matters because compliance artifacts do not automatically change everyday behaviour. A policy may state that humans are responsible while interface design encourages automatic acceptance. A risk assessment may identify bias while no frontline escalation route exists. Decision architecture connects governance intent with operational reality.

16.1 Counterarguments and Boundary Conditions

HILO-A can itself fail if it becomes bureaucracy rather than architecture. Excessive checkpoints may slow low-risk decisions, create approval fatigue, and encourage employees to treat governance as paperwork. A criticality-based design is therefore intended to remove unnecessary human intervention as deliberately as it adds necessary intervention.

Human override can also reintroduce precisely the biases that an algorithm was designed to reduce. A manager may protect an established preference, resist an unfamiliar recommendation, or use discretion inconsistently. For that reason, HILO-A does not romanticize human judgment: overrides should be explainable, reviewable, and compared with subsequent outcomes.

A third failure mode is defensive accountability. If named ownership is interpreted as personal blame, managers may become excessively risk-averse and block useful automation. Accountability should therefore mean authority to govern and learn, not automatic culpability for every model error. The architecture works best in a culture that distinguishes responsible ownership from scapegoating.

HILO-A should not be read as an argument that human judgment is inherently superior. Humans can preserve outdated routines, discriminate, rationalize intuitive preferences, or override a model precisely when the model is correcting a human bias. A mature architecture therefore audits human overrides as seriously as machine errors. The goal is not to privilege the person; it is to preserve a contestable decision process in which neither human nor machine authority becomes invisible.

Nor should accountability be confused with individual blame. Complex AI systems are socio-technical and often involve shared causal contribution. Naming an accountable decision-architecture owner is a governance mechanism for ensuring that someone has the authority to convene the relevant actors, review evidence, and redesign controls. It does not erase the differentiated responsibilities of providers, deployers, data owners, executives, professionals, or regulators.

A final boundary condition concerns speed. In some safety-critical settings, human intervention can be too slow and can make a system less safe. HILO-A therefore permits autonomous execution when the action and context justify it, while shifting human involvement upstream into rule design and downstream into monitoring. Meaningful human control is not synonymous with real-time manual approval.

17. Limitations

The first limitation is conceptual status. HILO-A has not yet been empirically validated across organizations or sectors. Its propositions should therefore be treated as research hypotheses rather than established causal findings.

Second, the Criticality Index is deliberately simple. Different sectors may require additional variables such as safety severity, regulatory classification, cybersecurity exposure, systemic impact, or effects on vulnerable populations. Organizations should adapt the matrix rather than apply thresholds mechanically.

Third, human oversight itself can fail. People may be biased, inattentive, undertrained, pressured, or overly trusting of automated outputs. HILO-A reduces the risk of purely nominal oversight by emphasizing competence, information, time, and authority, but it cannot eliminate human error.

Fourth, the framework focuses on organizational decisions and does not provide a complete technical assurance methodology. Model evaluation, cybersecurity, privacy engineering, data protection, red teaming, and domain-specific safety remain necessary disciplines.

18. Future Enhancement and Research Agenda

Empirical research can test HILO-A in at least four directions. First, case studies can compare how organizations allocate decision rights across similar AI use cases. Second, experiments can test whether different interface and override designs change automation bias. Third, longitudinal studies can examine how decision architectures evolve as organizations accumulate evidence about model performance. Fourth, survey research can investigate whether accountability clarity predicts employee trust and willingness to use AI systems.

A particularly useful research design would compare two organizational units deploying similar AI support. One would use conventional process governance; the other would explicitly map HILO-A layers, assign decision ownership, classify criticality, and maintain an override log. Outcomes could include decision quality, processing time, incident frequency, override quality, user trust, and perceived accountability.

Future work should also investigate generative AI agents. As systems become capable of planning and executing multi-step tasks, the unit of governance may shift from a single model output to a chain of delegated actions. HILO-A can be extended by mapping checkpoints inside agentic workflows and identifying which actions require pre-authorization, bounded authority, or post-execution review.

19. Practical Checklist for Managers

- Have we identified the real decision behind the AI use case?
- Do we know which data materially influence the decision and who owns them?
- Is the AI system informing, recommending, deciding, or executing?
- Is that role explicitly authorized rather than merely technically possible?
- Have we assessed impact, reversibility, uncertainty, and ethical sensitivity?
- Does the human reviewer have enough information, competence, time, and authority to disagree?
- Can a material AI output be paused, overridden, reversed, or escalated where appropriate?
- Are overrides and significant exceptions captured for learning?
- Is one accountable owner identifiable for the decision architecture?
- Do monitoring indicators cover both technical performance and real-world outcomes?
- Can the organization reconstruct how a consequential decision was reached?
- Is the automation level periodically reviewed as evidence and context change?

20. Conclusions

Artificial intelligence is entering organizations not only as software but as a participant in decisions. That change deserves an organizational response equal to its technological importance. The central challenge is not to defend human decision-making from machines, nor to maximize machine autonomy. It is to design the boundary intelligently.

HILO-A proposes one way to make that boundary visible. Its five layers—Data, Algorithm, Human Judgment, Decision, and Accountability—force a complete view of the decision chain. Its four interaction modes distinguish assistance from augmentation, supervised automation, and autonomy. Its Criticality Matrix makes proportionality operational. Its Override Principle and Decision Map convert human oversight from a slogan into a management mechanism.

The deepest managerial lesson is captured by the Accountability Non-Delegation Principle: organizations may delegate calculation, prediction, recommendation, prioritization, and even bounded execution to artificial intelligence, but they cannot delegate organizational accountability to the system itself. Someone must still own the mandate, the limits, the escalation route, the learning process, and the consequences.

The competitive organization of the AI era may therefore be neither the most automated nor the most cautious. It may be the organization that becomes exceptionally good at allocating intelligence: machine intelligence where speed and scale create value, human judgment where context and responsibility matter, and explicit accountability across the entire system.

The question of the AI era is no longer whether machines will participate in decisions; they already do. The managerial question is whether organizations can still explain why consequential decisions were made, where human intervention remained meaningful, when an automated output could be challenged, and who ultimately answers for the consequences. HILO-A is offered as a testable architecture for keeping those questions visible as artificial intelligence becomes more capable and more deeply embedded in organizational life.

Research integrity statement. This manuscript is a conceptual study. It reports no participant sample, interviews, experiment, survey, or fabricated empirical result. Illustrative scenarios are used only to demonstrate the proposed architecture. Empirical validation of HILO-A remains a future research requirement.

Author Note

Ignacio Bonasa Alzuria works at the intersection of leadership, organizational transformation, management innovation, and human development. In this paper he deliberately extends that work into AI-enabled organizational design, focusing on a practical question for managers: how to combine machine capability with meaningful human judgment and explicit accountability. Correspondence details may be added in the journal's required format before submission.

References

Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.600-1

European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union.

International Organization for Standardization & International Electrotechnical Commission. (2023). ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system. ISO.

National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). https://doi.org/10.6028/NIST.AI.100-1

OECD. (2024). OECD AI Principles (updated 2024). Organisation for Economic Co-operation and Development.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39(2), 230–253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, 30(3), 286–297.

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. Academy of Management Review, 46(1), 192–210.

Shneiderman, B. (2022). Human-Centered AI. Oxford University Press.

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. Business & Information Systems Engineering, 61, 637–643.

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. Business Horizons, 61(4), 577–586.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144(1), 114–126.
Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. Academy of Management Annals, 14(2), 627–660.
High-Level Expert Group on Artificial Intelligence. (2019). Ethics Guidelines for Trustworthy AI. European Commission.
Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50–80.
Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90–103.
OECD. (2024). Recommendation of the Council on Artificial Intelligence: OECD AI Principles, updated 2024. Organisation for Economic Co-operation and Development.
Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of human–AI interaction. Journal of Computer-Mediated Communication, 25(1), 74–88.
Tabassi, E. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1
Jussupow, E., Benbasat, I., & Heinzl, A. (2024). An integrative perspective on algorithm aversion and appreciation in decision-making. MIS Quarterly, 48(4), 1575–1590.
Lu, J. G., et al. (2024). AI aversion or appreciation? Meta-analytic evidence for a capability-personalization framework. Academy of Management Proceedings, 2024(1). https://doi.org/10.5465/AMPROC.2024.18332abstract
Moritz, J. M., Pomrehn, L., Steinmetz, H., & Wehner, M. (2023). A meta-analysis and review about the (un)fairness perceptions of algorithmic decision-making. Academy of Management Proceedings, 2023(1). https://doi.org/10.5465/AMPROC.2023.15483abstract
Schaap, G., Bosse, T., & Hendriks Vettehen, P. (2024). The ABC of algorithmic aversion: Not agent, but benefits and control determine the acceptance of automated decision-making. AI & Society, 39, 1947–1960.
Haupt, M., Freidank, J., & Haas, A. (2025). Consumer responses to human-AI collaboration at organizational frontlines: Strategies to escape algorithm aversion in content creation. Review of Managerial Science, 19, 377–413.
Corrêa, A. M., Garsia, S., & Elbi, A. (2025). Better together? Human oversight as means to achieve fairness in the European AI Act governance. Cambridge Forum on AI: Law and Governance, 1, e29. https://doi.org/10.1017/cfl.2025.10010
OECD. (2025). Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions. OECD Publishing. https://doi.org/10.1787/795de142-en
Embedding AI governance in organizations: A multi-level approach to responsible AI use. (2025). Procedia Computer Science, 270, 4144–4154. https://doi.org/10.1016/j.procs.2025.09.539
Exploring collaborative decision-making: A quasi-experimental study of human and generative AI interaction. (2024). Technology in Society, 78, 102662. https://doi.org/10.1016/j.techsoc.2024.102662
On the (non-) reliance on algorithms—A decision-theoretic account. (2024). Journal of Mathematical Psychology, 119, 102844. https://doi.org/10.1016/j.jmp.2024.102844

How do I cite this article?

Ignacio Bonasa Alzuria, Beyond Human-in-the-Loop: Designing Accountable Organizations for AI-Augmented Decision-Making — The HILO-A Framework, Excel International Journal of Technology, Engineering and Management, 2026; Volume -13, Issue-3_Page_200-213.
DOI Link: https://doi.org/10.64971/j.cph.eijtem.v13.i3.18.2026

This is an open access article under the CC BY-NC-ND license

(http://creativecommons.org/licenses/by-nc-nd/4.0/)