What AI Transformation Actually Requires

Abstract

Artificial intelligence is increasingly easy to acquire. Organizations can license models, embed copilots into existing software, connect applications to AI services, and give thousands of employees access in a comparatively short period of time, yet none of those actions constitutes transformation. AI transformation begins when the organization changes how work is designed, how authority is assigned, how risk is controlled, how evidence is preserved, and how results are measured.

The distinction matters because much of the language surrounding enterprise AI describes the presence of controls without establishing whether those controls can work. A "human in the loop" may hold no real authority, and an audit trail may record activity without reconstructing a decision. Model governance may govern a model while leaving the business process around it untouched, and adoption metrics may count users without showing whether anything improved.

This paper examines fifteen commonly cited requirements for AI transformation and reframes each as an operational management requirement. It distinguishes Checkpoint-Based Governance (CBG), which places named human authority over both the process and the decision, from Responsible AI automation and from human-in-the-loop observation. It argues that automation and observation are unacceptable as the sole control on consequential work. CBG is in production use in the author's own work, including three books, the HAIA frameworks, and published content, with documented operational evidence. It has not yet been independently validated, peer-reviewed, or deployed at enterprise scale, and it is one implementation path among those that may emerge. The paper is a management synthesis of published standards, regulation, and research, read alongside the author's governance framework, and it is neither a systematic review nor a legal opinion. The argument is simple. AI products can be purchased, but transformation cannot, and management has to put the conditions for it in place.

Keywords: AI transformation, AI governance, Checkpoint-Based Governance, human oversight, accountability, decision traceability, automation bias, enterprise AI

1. The Purchase Is Not the Transformation

The easiest part of AI transformation may be the part an organization can put on an invoice. A company selects a provider, purchases licenses, and gives employees access, and then training begins while internal announcements describe a new era of productivity. Usage numbers rise, and leadership waits for the transformation to arrive.

The technology has real value. The difficulty lies elsewhere, because access to capability and organizational capability are different things. A large language model can draft a document, but the model doesn't determine whether the document should be drafted with AI, what information may be provided to it, or who verifies the result. Nor does it settle what evidence must support the document, who approves its use, what happens when the model is wrong, or how the organization will know whether the new process is better than the old one. Those are management questions.

Established AI governance thinking already draws this distinction. The National Institute of Standards and Technology (NIST) treats AI risk management as an organizational activity. Its AI Risk Management Framework describes governance as a cross-cutting function that informs and runs through mapping, measurement, and management (NIST, 2023). The same framework calls for documented roles and responsibilities and assigns responsibility for AI risk decisions to executive leadership (NIST, 2023). NIST states that AI RMF 1.0, the version cited throughout this paper, is being revised (NIST, n.d.). ISO/IEC 42001 similarly treats artificial intelligence as a management-system matter, addressing the organizational policies, objectives, processes, risk management, and continual improvement that surround AI use (International Organization for Standardization [ISO] & International Electrotechnical Commission [IEC], 2023).

This leads to the first principle of AI transformation. The technology is what the organization buys, while transformation is what the organization changes. The question for leadership therefore moves past "Which AI should we buy?" to a harder one: what has to be in place around the AI before it becomes a reliable way of doing business? There are at least fifteen answers, and Table 1 lists them in the order this paper examines them, with each commonly cited label beside the operational requirement it becomes.

Table 1. The fifteen requirements

Section	Commonly cited requirement	Operational requirement
2	Data quality and ownership	Data Quality, Provenance, and Accountability
3	Use case prioritization	Purpose Before Capability
4	Process redesign	Workflow Design Around Human and AI Roles
5	Change management	Human Adoption and Role Transition
6	Adoption tracking	Outcome-Based Adoption Measurement
7	Risk appetite	Risk Boundaries Before Deployment
8	Model governance	AI-Enabled Process Governance Across the Lifecycle
9	Human in the loop	Human Checkpoints for Oversight and Accountability
10	Audit trail	Evidence and Decision Traceability
11	Escalation paths	Defined Escalation Triggers and Authority
12	Decision rights	Explicit Human Decision Authority
13	Vendor management	Third-Party AI Accountability
14	Capability uplift	AI Governance Competence
15	Value realization	Evidence-Based Value Realization
16	Continuous iteration	Continuous Review, Challenge, and Improvement

2. Data Quality and Ownership Becomes Data Quality, Provenance, and Accountability

"Good data" is an easy requirement to state and a difficult one to operationalize. Organizations often begin by asking whether data is accurate, accessible, and owned by somebody, and AI adds another layer to that inquiry. Management also needs to understand where information originates, how it has changed, what it was collected for,

whether its current use is appropriate, and who is responsible when it shapes an AI-assisted outcome.

Ownership alone does not answer those questions. A department may own a dataset while nobody can establish its provenance, and a vendor may process information while contractual ownership sits elsewhere. An AI system may also rely on information assembled from several systems, each with its own standards of quality and its own limitations.

The requirement should therefore be broader: **Data Quality, Provenance, and Accountability.** NIST connects provenance directly to both transparency and accountability, noting that maintaining the provenance of training data and supporting attribution of system decisions can assist with each (NIST, 2023). The framework also treats data, organizational practices, models, human behavior, and system context as parts of a broader socio-technical risk environment.

The management test becomes practical. Can the organization identify the material information behind an AI-enabled process, determine whether it suits that purpose, establish its origin where necessary, and name who is accountable for its use? If it can't, the organization has data access, which is a different thing from a data foundation for transformation.

3. Use Case Prioritization Becomes Purpose Before Capability

AI creates a temptation that accompanies almost every powerful technology, which is to begin with what the technology can do and then search for somewhere to use it. That order is backwards. The first question should be what problem the organization is solving, and "Where can we use AI?" comes second.

A legitimate AI use case begins with a defined operational need and an expected outcome. Only then should management determine whether AI improves the process compared with reasonable alternatives. That produces a stronger requirement: **Purpose Before Capability.**

For each proposed use, the organization should be able to define the business problem, expected value, affected parties, foreseeable risks, required evidence, decision authority, and measure of success before deployment. This turns prioritization from a technology contest into a management decision. A use case with impressive technical capability and no meaningful business outcome should not outrank a modest application that produces

measurable value. Transformation means applying AI deliberately where it improves the work. A longer list of places where AI appears is not evidence of that improvement.

4. Process Redesign Becomes Workflow Design Around Human and AI Roles

The fastest implementation is often to place AI inside an existing process without changing much else. That may also be the easiest way to preserve the inefficiencies and ambiguities of the old process.

AI changes the allocation of work. Tasks that once belonged entirely to a person may become partially automated, research may accelerate, and drafting may move earlier in the sequence. Review becomes more important, and verification can become harder precisely because output arrives faster.

The requirement is therefore **Workflow Design Around Human and AI Roles.** For every significant AI-enabled workflow, management should be able to explain what the AI does, what the human does, what remains prohibited, where verification occurs, when approval becomes necessary, and who owns the resulting work. The key is intentional allocation. The organization should not let human and AI responsibilities emerge by accident from employee habits, since a process is transformed only when the division of labor between people and machines has been designed on purpose.

5. Change Management Becomes Human Adoption and Role Transition

AI transformation changes more than tools because it changes expectations. An employee who once produced the first draft may now review one, and a manager who once evaluated an employee's work may also need to evaluate the reliability of that employee's AI-assisted process. A specialist may become responsible for challenging an automated recommendation, and a person once responsible for execution may become responsible for supervision. Those are role changes.

Traditional change management still matters, but the requirement should be stated more directly as **Human Adoption and Role Transition.** Employees need to know how to operate a tool, and they also need to know how their responsibilities change when AI enters their work. That includes what they may delegate, what they may not delegate, when they must verify, what they are expected to know, where they retain authority, and what remains their responsibility even when AI performs part of the task.

An organization hasn't managed AI change if employees know where the button is but not what they are responsible for after pressing it.

6. Adoption Tracking Becomes Outcome-Based Adoption Measurement

Enterprise software has trained organizations to measure adoption through activity: how many licenses are active, how many employees opened the application, how often they use it, and how many prompts they submit. These are useful operational statistics and weak evidence of transformation. A person can use an AI assistant every day and become less productive, while another can use it twice a month and eliminate hours of high-value work.

The stronger requirement is **Outcome-Based Adoption Measurement.** AI adoption should connect to changes in how tasks actually get done. Depending on the use case, management may measure time to completion, error rates, rework, throughput, quality, cost, customer outcomes, employee capacity, risk events, or another business-specific result.

The principle holds across every use case. Usage measures interaction, and outcomes measure transformation, so a rising user count should never stand in for evidence that the organization is getting better at something.

7. Risk Appetite Becomes Risk Boundaries Before Deployment

Organizations routinely state that they have a certain risk appetite, and the concept becomes useful for AI only when it produces boundaries people can act on. Those boundaries answer concrete questions. They settle what information may enter a public model, which decisions may use AI assistance, when AI may act on its own, when a person must intervene, what level of uncertainty triggers escalation, and which uses are prohibited regardless of potential efficiency.

The requirement is **Risk Boundaries Before Deployment.** NIST ties governance to organizational risk tolerance, calling for processes that set the needed level of risk management activity based on that tolerance and for policies, procedures, and other controls grounded in organizational risk priorities (NIST, 2023). No organization operates without risk, so the purpose of boundaries is narrower and more practical: they

keep individual employees, project teams, and vendors from inventing the organization's AI risk tolerance one prompt at a time.

Risk boundaries also decide how much human control each use requires, and that choice belongs before deployment, while the workflow is still being designed. Automation and observation remain acceptable for work whose outputs are reversible and whose stakes are low, provided the organization discloses the automation honestly and samples its outputs on a schedule it sets in advance. They are unacceptable as the sole control on consequential work, and Table 2 sets out the tiers that follow from that line.

Table 2. Control tiers by consequence

Tier	Work it covers	Minimum control
Lower stakes	Reversible outputs, internal drafts, and high-volume tasks with low consequence	Automation with honest disclosure and scheduled sampling review
Consequential	Irreversible consequences, high-stakes decisions, and regulated domains	A named human checkpoint on both the process and the decision, with a reconstructable record (Section 9)
Unclassified	Work whose stakes have not yet been assessed	Treated as consequential until it is classified

Classifying work as consequential requires criteria set before deployment, so classification does not rest on judgment formed in the moment. The relevant dimensions include the severity of plausible harm, whether the outcome can be reversed, and the effect on legal rights or access to opportunity. They also include the number of people affected, how readily an error would be detected, and how much time remains to reverse an action once taken. Each organization sets its own thresholds on those dimensions according to its risk tolerance.

Checkpoint-Based Governance also recommends a practical order of adoption: "The recommended starting point is the highest-risk decisions in the current workflow," and "The practice expands as institutional capacity develops" (Puglisi, 2026a).

8. Model Governance Becomes AI-Enabled Process Governance

Model governance is important and too narrow to carry the entire governance burden. An AI failure may originate in the model, and it may just as easily originate in bad data, improper instructions, an inappropriate use case, or a poorly designed workflow. AI failures can also arise from automation bias, weak verification, vendor changes, downstream systems, or human decisions. Governing only the model can therefore produce the appearance of control while the actual business process stays exposed.

The requirement should become **AI-Enabled Process Governance Across the Lifecycle.** NIST describes governance as a continual and intrinsic requirement over an AI system's lifespan and across the organization's hierarchy, and it frames governance as a cross-cutting function (NIST, 2023). ISO/IEC 42001 likewise uses a management-system approach built on establishing, implementing, maintaining, and continually improving how an organization manages AI (ISO & IEC, 2023).

Management should therefore govern the complete system of work, which includes the model, data, human participants, workflow, integrations, outputs, controls, vendor dependencies, monitoring, changes, and eventual retirement. The unit of governance is the business process in which AI operates, and the model is only one component inside it.

9. Human in the Loop Becomes Human Checkpoints for Oversight and Accountability

"Human in the loop" has become one of the most reassuring phrases in AI governance, and it is also dangerously incomplete. A human can sit inside a process without understanding the model, without enough time to review its output, without access to supporting evidence, without authority to reject the result, and without any obligation to intervene. That human is present, but presence alone governs nothing.

Research on automation bias reinforces the concern. A systematic review of 35 peer-reviewed studies documents that over-reliance on automated recommendations turns on several interacting factors, including AI literacy, professional expertise, cognitive profile, trust dynamics, task verification demands, and explanation complexity (Romeo & Conti, 2026). Placing a person in the process doesn't by itself neutralize automation risk.

The European Union's AI Act makes the distinction concrete for high-risk systems. Article 14 requires that the people assigned to oversight can understand the system's capacities and limitations, remain aware of automation bias, interpret its output correctly, decide to disregard, override, or reverse that output, and interrupt the system (Regulation (EU) 2024/1689, 2024, art. 14). Article 26(2) complements it by requiring deployers to assign human oversight to natural persons who have "the necessary competence, training and authority, as well as the necessary support" (Regulation (EU) 2024/1689, 2024, art. 26(2)). Fink (2025) examines how far those provisions can carry the weight placed on them.

For one category, remote biometric identification, Article 14(5) goes further and bars action on the system's identification unless at least two natural persons have separately verified and confirmed it, a stricter rule than any single checkpoint.

The Digital Omnibus on AI, Regulation (EU) 2026/1744, entered into force on 27 July 2026 and postponed application of the relevant Chapter III requirements, including the deployer duties in Article 26. Those requirements now apply from 2 December 2027 for stand-alone high-risk systems and from 2 August 2028 for high-risk AI embedded in regulated products (Cloud Security Alliance, 2026; White & Case LLP, 2026). The dates moved, and the description of what effective oversight must look like when they arrive remains the useful part for management.

Taken together, these sources describe what oversight must be able to do. NIST assigns responsibility for AI risk decisions to executive leadership, ISO/IEC 42001 builds AI management around leadership commitment, and Article 14 lists the capacities an overseer needs. None of them specifies the point at which a named person takes ownership of a specific output, and that is the design gap this paper addresses. These sources state what oversight must make possible. They do not specify the architecture that makes oversight evidenced and attributable, and the method set out below is the architecture this paper proposes for that purpose, not one these sources prescribe.

Organizations tend to reach for one of three methods when they claim their AI is under control, and only one of them attaches a named person to the result. The first is Responsible AI, which the author has defined in published work (Puglisi, 2026b) as automation: machines validate machine output and AI agents run the pipeline, with no human checkpoint. That method raises the question of who answers when an output fails, and it cannot answer that question, because no named person stands behind any individual output.

The second method is a human-in-the-loop arrangement in which a person observes the automation as it runs. Puglisi (2026a) states the limit of that arrangement directly in Checkpoint-Based Governance: "Human In The Loop means the human is present and participating, but does not require that the human holds authority or bears accountability for the outcome." An observer can watch a failure develop and still have no standing to stop it.

The term itself is loose, and Crootof, Kaminski, and Price (2023) document the wide range of roles it is used to describe. They warn against the reflex of inserting a human into a decision process as a remedy for the limits of an algorithm. Their conclusion, that policymakers should regulate the human-in-the-loop system rather than simply placing a human in the loop, is the direction this paper takes.

Either Responsible AI automation or observation-only human involvement can serve lower-stakes work within the boundaries set in Section 7, but automation and observation are unacceptable as the sole control on consequential work, because neither one attaches a person to the result.

The third method is CBG, which is built for that gap. In the framework's own terms, "CBG is AI Governance. It provides human oversight and accountability for AI-assisted work," and it is "the mechanism that converts human presence into human authority, and participation into documented accountability" (Puglisi, 2026a). The framework also names the failure it exists to prevent: "The governance failure mode is not absence of humans. It is presence without accountability" (Puglisi, 2026a). CBG is in production use in the author's own work, including three books, the HAIA frameworks, and published content, with documented operational evidence. It has not yet been independently validated, peer-reviewed, or deployed at enterprise scale. Other governance approaches exist, and CBG represents one implementation path among those that may emerge.

The stronger requirement is therefore **Human Checkpoints for Oversight and Accountability.** A checkpoint is a defined point at which a competent person reviews the work and holds real authority over what happens next. That authority operates on two layers, and treating them as one is where many oversight designs go wrong.

The first layer is the process, where human accountability lives. Here the person directs the AI, challenges what it produces, and verifies the evidence behind it, and the same person controls movement through each step. That control includes approving the interaction to continue to its next step, stopping the process before an output reaches a

decision, and escalating the work to someone with greater authority. Every direction, challenge, and verification is an act a named person performs and can answer for.

The second layer is the decision, where ownership lives. At this layer the checkpoint closes on one of three outcomes: accept, modify, or reject. Reject ends the process, while modify sends the output back into the process layer to be directed, challenged, and verified again before it returns for decision. Approving an interaction and accepting an output are therefore different acts, since the first lets the task proceed and the second makes a person the owner of what it produced.

Some consequential work runs faster than any person can review, and the checkpoint moves upstream for that work rather than disappearing. Under CBG, the organization places a checkpoint before execution, where "the human governor establishes scope, criteria, and constraints," and another during it, where "the governor monitors with authority to intervene, redirect, or terminate" (Puglisi, 2026a). Automation is not the sole control when a named person sets the operating limits and retains authority to halt the system.

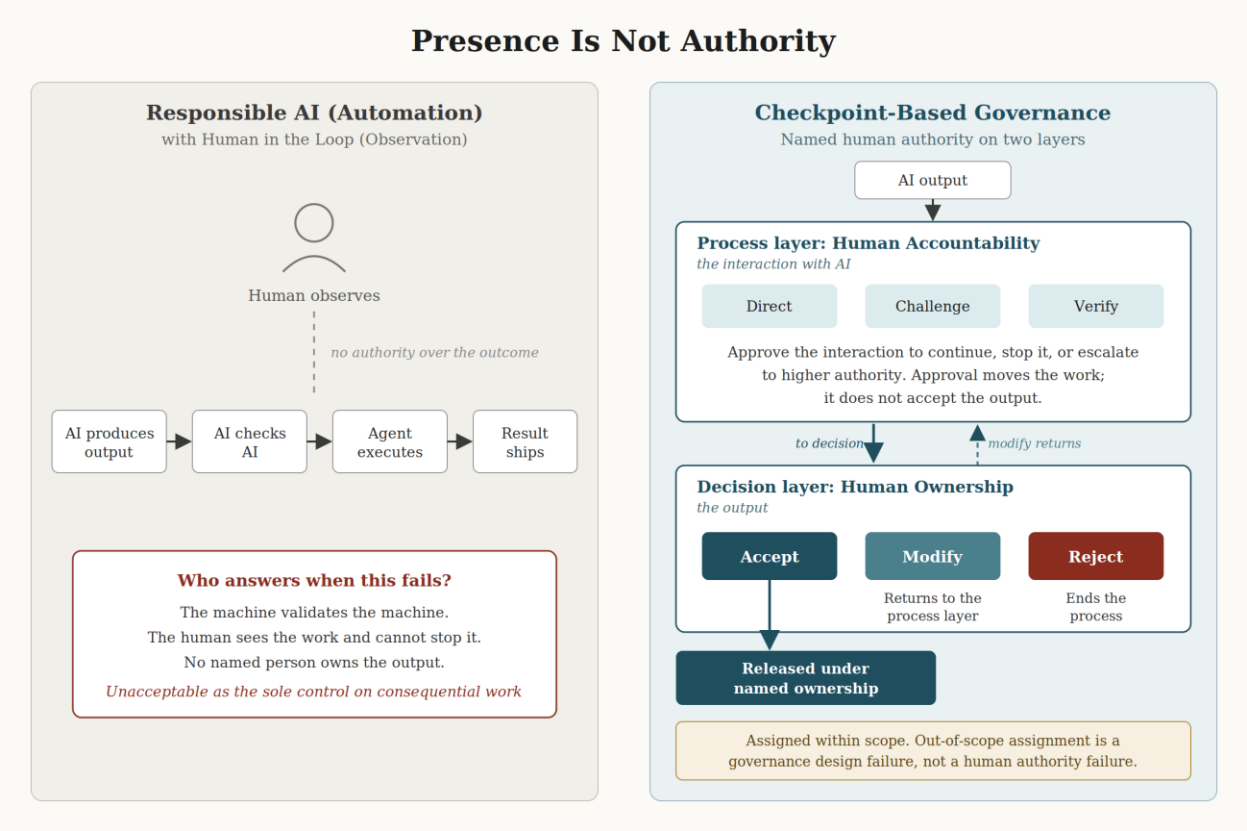

Figure 1. Presence is not authority. On the left, automation runs from input to action while a human observes without authority over the outcome. On the right, AI output

enters a process layer where a named human directs, challenges, and verifies it, and a decision layer where the output is accepted and released under named ownership, modified and returned to the process layer, or rejected.

This structure also separates two concepts that organizations too often blend together. Oversight is the ability to observe, evaluate, challenge, and intervene, while accountability establishes who remains answerable for the consequential decision or action. The phrase "human in the loop" establishes neither one. A properly constituted checkpoint makes both explicit and auditable, although its effectiveness still depends on the human, organizational, and technical conditions under which it operates.

Where Checkpoints Fail

A framework that rejects ceremonial oversight must also account for the ways a checkpoint can become ceremonial, and the human-factors literature has documented that problem for decades. Bainbridge (1983) showed that automation can expand rather than eliminate problems with the human operator, who is left with the tasks the designer could not automate and fewer chances to practice them. Parasuraman, Sheridan, and Wickens (2000) found that automation changes human activity rather than simply replacing it, and that it imposes new coordination demands on the operator. Goddard, Roudsari, and Wyatt (2012) reviewed the frequency, mediators, and mitigators of automation bias, and Romeo and Conti (2026) carried that line of research into modern AI settings.

Two critiques go further and deserve a direct answer. Green (2022) surveyed 41 policies that require human oversight of government algorithms and concluded that people often cannot perform the oversight those policies demand, which lets the policies legitimize flawed systems. He proposes institutional oversight as the alternative, under which agencies must justify an algorithm and any proposed human oversight with evidence before the plan receives public review. Elish (2019) introduced the moral crumple zone, in which responsibility is misattributed to a human actor who had limited control over an automated system.

A third finding cuts closer still. Green and Chen (2019) found that participants given a risk assessment still underperformed it, could not evaluate the accuracy of their own predictions or the tool's, and produced disparate outcomes by race. Their findings are evidence that people holding real decision authority can exercise it badly.

Green and Elish both describe responsibility without authority, which is the arrangement this paper rejects, and Green and Chen show that authority alone does not

make judgment sound. A checkpoint supplies answerability, which differs from correctness, and a checkpoint is meaningful only when three conditions hold. The first is that the person holds real authority on both layers, including the power to stop the process and to reject the output. The second is a record that shows what that person was given, what they directed, challenged, and verified, and what they decided, so that responsibility follows actual control rather than proximity. The third is that passive acceptance must be detectable, since a reviewer's habitual approval without genuine review counts as a governance failure under CBG (Puglisi, 2026a).

CBG also places the failure where it belongs when an organization asks too much of the person at the checkpoint: "Human authority at the checkpoint is always valid within an appropriate scope. CBG checkpoint assignments must be calibrated to the governor's developmental stage and life experience. A governor is assumed capable of decisions within the scope that matches their experience. Placing a governor at decisions outside that scope is a governance design failure, not a human authority failure" (Puglisi, 2026a). Under that rule, a named owner answers for the judgment actually exercisable at the checkpoint, while the executives, deployers, and providers who selected the system, staffed the review, and set the volume retain responsibility for those choices.

The record itself is protected from revision after the fact, since "Closed checkpoint records are immutable," and corrections are appended as new entries that leave the original in place (Puglisi, 2026a).

Human authority also carries a boundary of its own, because "no human governor may direct an AI-assisted outcome that injures a human being, allows harm through inaction, or harms humanity" (Puglisi, 2026a). A traceable record makes a bad exercise of authority visible and reviewable.

Where volume makes full review impossible, the tiers in Section 7 decide which tasks receive a named checkpoint and which receive scheduled sampling, so that human attention goes to consequential decisions. Green's institutional alternative remains an important alternative in the literature, and organizations in the public sector should weigh it alongside the approach set out here.

10. Audit Trail Becomes Evidence and Decision Traceability

Organizations often treat logging as evidence of accountability, and the two are not the same. A system may preserve timestamps, prompts, model outputs, user IDs, and technical events while leaving management unable to explain why an important decision

was made. An audit trail records activity, and transformation requires something stronger: **Evidence and Decision Traceability.**

NIST distinguishes transparency, explainability, and interpretability, and it links documentation with human review and accountability. Its framework treats transparency as spanning design decisions and training data through deployment, post-deployment, and end-user decisions, including how, when, and by whom those decisions were made (NIST, 2023).

For consequential AI-assisted work, the organization should be able to reconstruct enough of the chain to answer a short set of questions. What information mattered, what did the AI contribute, and what was checked? What did the human decide, who held the authority, and what happened afterward? Under the checkpoint structure in Section 9, those answers come from both layers, since the process record shows what was directed, challenged, and verified and the decision record shows who accepted, modified, or rejected the output.

Not every low-risk interaction requires an evidentiary dossier, and controls should remain proportionate. The greater the consequence, however, the less acceptable it becomes for the organization's answer to be that it knows the system was used and can't reconstruct the decision.

11. Escalation Paths Become Defined Escalation Triggers and Authority

Most organizations already have escalation paths, and the trouble is that employees may not know when to use them. AI creates new forms of uncertainty. A model may contradict a known source, two systems may reach materially different conclusions, or a generated recommendation may exceed the reviewer's expertise. Sensitive data may have been exposed, a system may behave differently after an update, and a user may suspect bias without knowing whether the concern rises to a formal incident.

A directory of people to contact doesn't resolve those situations. The stronger requirement is **Defined Escalation Triggers and Authority.** The organization should identify the conditions that stop normal work and require another level of review, and it should make clear who receives the escalation and what authority that person or function holds.

Escalation belongs to the process layer described in Section 9. It routes the work to someone with greater authority, and the receiving decision-maker then closes the

checkpoint through a separate decision. Escalation is therefore part of workflow design and something more than an emergency number. If an employee is expected to exercise judgment about AI, management also has to define what happens when that judgment reaches its limit.

12. Decision Rights Become Explicit Human Decision Authority

AI blurs the language around decisions. A system recommends, a platform selects, and an agent executes, while a person reviews and a manager approves. Those verbs can hide more than they reveal, and the central governance question remains who actually holds authority.

The requirement should become **Explicit Human Decision Authority.** Organizations should distinguish between generating information, recommending an action, initiating an action, approving it, stopping it, reversing it, and accepting responsibility for the result. Approving an action lets the process move forward, while accepting an outcome at a checkpoint assigns ownership of it. A sound design keeps those two authorities distinct.

NIST calls for roles and responsibilities to be documented and clear throughout the organization, and it states that executive leadership takes responsibility for decisions about risks associated with AI development and deployment (NIST, 2023). This matters because accountability tends to become ambiguous at exactly the point where automation becomes useful. If everyone can point to somebody else, nobody owns the decision, and AI transformation requires authority to remain identifiable even when execution is distributed between people and machines.

13. Vendor Management Becomes Third-Party AI Accountability

Much enterprise AI is somebody else's technology, and that does not make the resulting business decisions somebody else's responsibility. Organizations depend on model providers, cloud platforms, application vendors, data suppliers, implementation partners, and increasingly complex chains of AI-enabled services.

Traditional vendor management remains necessary, and AI adds questions about changing models, whether customer data is retained or used for training, performance shifts, system updates, transparency, monitoring, dependency, portability, and

continuity. The stronger requirement is **Third-Party AI Accountability.** NIST's governance function explicitly addresses the full product lifecycle, including legal and other issues concerning the use of third-party software, hardware systems, and data (NIST, 2023).

The business principle matters more than the procurement mechanism. An organization can outsource technology, but it cannot outsource responsibility for how that technology is used, so vendor assurances need to connect to the organization's own controls. The question goes beyond whether the vendor is trustworthy to whether the organization can still govern the work when the technology underneath it belongs to somebody else.

14. Capability Uplift Becomes AI Governance Competence

Much of the first generation of enterprise AI training centers on prompting, which makes sense because employees need to know how to interact with the systems they use. Generating a better output, however, is only one part of competent AI use. People also need to know when an output requires verification, when uncertainty matters, which sources can support a claim, when a task exceeds their expertise, which information is inappropriate to provide, when a system should be challenged, and when an issue must be escalated.

The stronger requirement is **AI Governance Competence**, defined as the ability to direct, challenge, verify, and own work done with AI. Those four verbs map onto the checkpoint in Section 9. Directing, challenging, and verifying are the work of the process layer, and ownership belongs to the decision layer. That competence is what makes the modify outcome real rather than ceremonial. A reviewer who cannot direct, challenge, or verify is left only to accept or reject whatever arrives.

NIST calls for personnel and partners to receive AI risk management training that enables them to perform their duties, and it separately calls for defined roles and responsibilities for human-AI configurations and oversight of AI systems (NIST, 2023). This is the difference between teaching employees to use AI and preparing them to work responsibly with it, and the organization needs both.

15. Value Realization Becomes Evidence-Based Value Realization

AI creates an unusually favorable environment for soft claims. Employees report saving time, managers observe faster output, vendors publish productivity estimates, and

leadership sees more work being produced. None of those observations is meaningless, and none should automatically count as realized business value.

The stronger requirement is **Evidence-Based Value Realization.** Every meaningful AI use case should begin with an expected outcome and eventually be compared against evidence. If the claim is speed, the organization measures time, and if the claim is quality, it defines quality before measuring it. A claim of cost reduction calls for the actual cost, a claim of added capacity calls for evidence of what that capacity produced, and a claim of better decisions calls for an agreed standard of what improvement would look like.

Each claim also needs three supports that turn a measurement into evidence. The first is a baseline taken before AI entered the workflow, and the second is an attribution rule that separates the effect of AI from the effect of the process redesign that came with it. The third is a stop rule that retires a use case when it fails its own test. Time saved counts as value only when quality, rework, and risk hold steady or improve over the same period.

Few AI interactions need to become laboratory experiments, but every meaningful use case requires management to distinguish enthusiasm from measurement. Transformation becomes credible when the organization can show where AI produces value, where it does not, and where an apparent gain in one dimension creates a cost somewhere else.

16. Continuous Iteration Becomes Continuous Review, Challenge, and Improvement

AI transformation never reaches a permanent finished state. Models and vendors change, data changes, and employees grow more experienced while workarounds emerge and new integrations appear. Business objectives shift, some risks once considered theoretical become observable, and some controls prove more restrictive than the risk requires. A governance structure that works on the day of deployment can therefore become ineffective without anybody formally changing it.

The requirement is **Continuous Review, Challenge, and Improvement.** NIST calls for ongoing monitoring and periodic review of the risk management process and its outcomes, with roles defined and the frequency of review determined in advance (NIST, 2023). ISO/IEC 42001 similarly structures AI management around continual improvement (ISO & IEC, 2023).

Continuous improvement should apply to the human system around the model as well as to model performance. Management should ask whether checkpoints are catching meaningful problems, whether people override the system when they should, and whether escalations occur too often or never. It should also ask whether controls remain proportionate to risk, whether claimed benefits still materialize, and whether a vendor change has altered the assumptions under which the system was approved. Transformation requires an organization that can revise its relationship with AI as quickly as the technology itself changes.

17. The Fifteen Requirements Are Connected

These fifteen requirements should not be treated as fifteen independent boxes on a compliance checklist, because together they form a management system. Poor data degrades output quality, and weak use-case selection sends investment toward the wrong problems. Bad workflow design makes human oversight superficial, superficial oversight weakens accountability, and poor traceability makes oversight difficult to prove. Unclear decision authority makes escalation ineffective, weak competence makes every other control less reliable, and missing measurement lets unsuccessful deployments continue because nobody has established what success means.

The system can also be tested in operation, and a short set of measures shows an executive whether it works. Those measures are the reconstruction success rate on a sample of consequential decisions, the approval rate and decision reversal rate at checkpoints, and the escalation cycle time. They also include the change in outcomes against the pre-AI baseline and the rate of control failures after a vendor or model change. Two further measures test whether checkpoints work rather than merely operate. The first is the rate at which sampled checkpoint decisions agree with independent expert review. The second is the share of adverse events whose root-cause analysis separates individual judgment failure from failures of system design, staffing, or assignment. Under CBG, management must be able to detect passive acceptance, and it sets the specific thresholds itself (Puglisi, 2026a), choosing targets for these measures according to the risk involved.

This is why AI transformation cannot be delegated entirely to IT. Technology teams matter enormously, but the transformation crosses operations, risk, finance, legal, human resources, procurement, security, data management, executive leadership, and the people performing the work. AI is the technology entering the organization, and transformation is the organizational response to it.

18. A Better Executive Question

The executive question is often framed as how quickly the organization can get AI into the business. Speed matters, and a more useful question follows immediately: what has to be true for the organization to trust the way AI is being used once it gets there?

That question moves leadership away from licenses and toward conditions. Leadership should be able to say whether the purpose is defined, whether an accountable person is named, and whether that person can actually intervene. It should also know whether consequential decisions can be reconstructed, whether people recognize when a system's output should not be trusted, whether problems can escalate, whether AI is improving the business, and whether the organization can adapt when the technology changes. When the answers are unclear, what the organization faces is often a management problem that AI has exposed, more than an AI problem in its own right.

Conclusion: Transformation Begins After the Purchase

The first phase of enterprise AI is easy to recognize: organizations experiment, select vendors, buy licenses, provide access, and train employees. The next phase is harder because it cannot be purchased in the same way. Organizations have to decide where AI belongs, redesign work around it, define human authority, establish risk boundaries, preserve evidence, assign accountability, build competence, measure results, and keep adjusting the system as both technology and behavior change. That is the work of transformation.

The phrase "human in the loop" illustrates the larger problem, because a comforting label is not enough. Responsible AI automation cannot say who answers when the work fails, and a human who only observes cannot stop it, so for consequential work neither can serve as the sole control. Human presence has to become human authority over the process and human ownership of the decision, which Checkpoint-Based Governance is designed to supply, and that authority has to be supported by evidence.

The same standard applies across the entire enterprise. Data ownership has to become data accountability, usage has to become outcomes, and model governance has to become governance of the AI-enabled process. Audit logs have to become decision traceability, training has to become competence, vendor management has to preserve organizational responsibility, and continuous improvement has to include the controls themselves.

Artificial intelligence can make an organization faster without making it better governed. It can increase output without increasing value, and it can place a human in a process without giving that human meaningful control. The technology will not resolve those contradictions for management, so management has to resolve them. AI products are purchased, AI transformation is built, and the real work begins after the license.

Appendix: Working Glossary

AI-Enabled Process: A business workflow in which an AI system materially contributes to analysis, content, recommendations, decisions, execution, or other work.

AI Governance Competence: The ability to direct, challenge, verify, and own work done with AI, while understanding one's authority and responsibilities, including when to escalate.

AI Transformation: Organizational change in which AI materially changes how work is performed, governed, and measured, or how it creates value.

Automation Bias: A tendency to rely excessively on automated output or recommendations, including circumstances in which human reviewers fail to sufficiently challenge automated results.

Checkpoint: A defined point in an AI-enabled workflow where a competent person holds authority on two layers: the process, where work is directed, challenged, verified, approved to continue, stopped, or escalated, and the decision, where the output is accepted, modified, or rejected.

Checkpoint-Based Governance (CBG): "CBG is AI Governance. It provides human oversight and accountability for AI-assisted work" (Puglisi, 2026a).

Decision Authority: Explicit authority to accept, modify, or reject an AI-assisted output. Reject ends the process, and modify returns the work to the process layer.

Decision Traceability: The ability to reconstruct the material evidence, system contributions, human reviews, authorities, and actions that produced a consequential decision.

Escalation Trigger: A defined condition that requires normal AI-enabled work to stop or move to another level of human authority.

Consequential Work: Work whose outputs carry irreversible consequences, involve high-stakes decisions, or fall within regulated domains, together with any work whose stakes have not yet been classified.

Human Accountability: Assignment of responsibility for consequential AI-assisted actions or decisions to an identifiable human or organizational authority.

Human in the Loop: An umbrella term for human participation in an automated process. As used in this paper, observation without any requirement that the human holds authority over the outcome or bears accountability for it.

Human Oversight: The human ability to understand, monitor, evaluate, challenge, intervene in, override, or stop an AI-enabled process.

Process Approval: Authorization for an AI-assisted interaction to continue to its next step. Process approval moves the work forward and does not accept the output.

Provenance: Information about the origin, history, custody, or material lineage of data or other evidence used within an AI-enabled process.

Responsible AI: As used in this paper, automation in which machines validate machine output or AI agents run the pipeline, with no human checkpoint and no named person answering for an individual output. Acceptable for reversible, lower-stakes work and unacceptable as the sole control on consequential work.

Risk Boundary: A defined limit on where or how AI may be used, including conditions requiring additional controls, approval, escalation, or prohibition.

Scope-Appropriate Assignment: The CBG requirement that checkpoint assignments match the governor's experience, under which placing a governor at decisions outside that scope is a governance design failure, not a human authority failure (Puglisi, 2026a).

Third-Party AI Accountability: Organizational responsibility for controlling and governing AI-enabled work even when models, infrastructure, data, or other technical components come from outside parties.

Value Realization: Measurable conversion of AI use into organizational outcomes, as distinct from increased system usage or reported productivity.

References

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*(6), 775-779.

Cloud Security Alliance. (2026, August 1). *EU AI Act's high-risk deadline: Deferred, not cancelled*. https://labs.cloudsecurityalliance.org/research/csa-research-note-eu-ai-act-high-risk-deadline-omnibus-20260/

Crootof, R., Kaminski, M. E., & Price, W. N., II. (2023). Humans in the loop. *Vanderbilt Law Review, 76*(2), 429-510.

Elish, M. C. (2019). Moral crumple zones: Cautionary tales in human-robot interaction. *Engaging Science, Technology, and Society, 5*, 40-60. https://doi.org/10.17351/ests2019.260

Fink, M. (2025). *Human oversight under Article 14 of the EU AI Act*. SSRN. https://papers.ssrn.com/abstract=5147196

Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association, 19*(1), 121-127.

Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review, 45*, 105681. https://doi.org/10.1016/j.clsr.2022.105681

Green, B., & Chen, Y. (2019). Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 90-99). Association for Computing Machinery. https://doi.org/10.1145/3287560.3287563

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). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1

National Institute of Standards and Technology. (n.d.). *AI risk management framework*. Retrieved September 18, 2026, from https://www.nist.gov/itl/ai-risk-management-framework

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. https://doi.org/10.1109/3468.844354

Puglisi, B. C. (2026a). *Checkpoint-Based Governance: A constitutional framework for human-AI collaboration*. basilpuglisi.com. https://basilpuglisi.com/cbg/

- Asimov, I. (1942). Runaround. Astounding Science Fiction.
- Asimov, I. (1985). Robots and Empire. Doubleday.
- Puglisi, B. C. (2025). Governing AI: When Capability Exceeds Control. basilpuglisi.com.
- Puglisi, B. C. (2026). HAIA-RECCLIN Multi-AI Framework Updated for 2026. basilpuglisi.com.
- Puglisi, B. C. (2026). HAIA-CAIPR: Cross AI Platform Review, Specification v1.1. basilpuglisi.com.
- Puglisi, B. C. (2026). HAIA-RECCLIN CBG Audit Log, Case Study 002. basilpuglisi.com.
- Puglisi, B. C. (2026). Case Study 006: The Discovery of CAIPR, v7. basilpuglisi.com.
- Puglisi, B. C. (2026). GOPEL v0.6.1. github.com/basilpuglisi/HAIA.

Puglisi, B. C. (2026b, January 1). *Ethical AI, Responsible AI, and AI Governance are not the same thing*. basilpuglisi.com. https://basilpuglisi.com/what-we-failed-to-define-is-how-we-fail/

- Bai, Y., et al. (2022). Constitutional AI: Harmlessness from AI feedback. arXiv. https://arxiv.org/abs/2212.08073
- Center for AI Safety. (2023, May 30). Statement on AI risk. https://safe.ai/work/press-release-ai-risk
- CGTN. (2024, December 28). 30 years left? AI 'Godfather' warns the technology may end humanity. https://newseu.cgtn.com/news/2024-12-28/AI-Godfather-warns-rapid-development-can-cause-human-extinction-1zHEIq63EDC/index.html
- Dietterich, T. G. (2000). Ensemble methods in machine learning. In Multiple classifier systems (pp. 1-15). Springer. https://web.engr.oregonstate.edu/~tgd/publications/mcs-ensembles.pdf
- European Commission. (2019). Ethics guidelines for trustworthy AI. High-Level Expert Group on Artificial Intelligence. https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf
- European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council (Artificial Intelligence Act), Article 14: Human oversight. Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng
- Harper, D. (n.d.). Govern. In Online Etymology Dictionary. https://www.etymonline.com/word/govern
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? Behavioral and Brain Sciences, 33(2-3), 61-83. https://www2.psych.ubc.ca/~henrich/pdfs/WeirdPeople.pdf
- IBM. (2025). AI governance. IBM Think. https://www.ibm.com/think/topics/ai-governance

- International Organization for Standardization. (2021). ISO 37000:2021 Governance of organizations. https://www.iso.org/standard/65036.html
- International Organization for Standardization. (2023). ISO/IEC 42001:2023 AI management systems. https://www.iso.org/standard/42001
- Irving, G., Christiano, P., & Amodei, D. (2018). AI safety via debate. arXiv. https://arxiv.org/abs/1805.00899
- Tao, Y., Viberg, O., Baker, R. S., & Kizilcec, R. F. (2024). Cultural bias and cultural alignment of large language models. PNAS Nexus, 3(9), pgae346. https://pmc.ncbi.nlm.nih.gov/articles/PMC11407280/
- Merriam-Webster. (n.d.). Govern. In Merriam-Webster.com dictionary. https://www.merriam-webster.com/dictionary/govern
- Merriam-Webster. (n.d.). Governance. In Merriam-Webster.com dictionary. https://www.merriam-webster.com/dictionary/governance
- Microsoft. (2022). Microsoft Responsible AI Standard v2: General requirements. https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/Microsoft-Responsible-AI-Standard-General-Requirements.pdf
- MIT Sloan School of Management. (2023, May 23). Why neural net pioneer Geoffrey Hinton is sounding the alarm on AI. https://mitsloan.mit.edu/ideas-made-to-matter/why-neural-net-pioneer-geoffrey-hinton-sounding-alarm-ai
- National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). U.S. Department of Commerce. https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf
- National Institute of Standards and Technology. (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (NIST AI 600-1). U.S. Department of Commerce. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf
- OECD. (2019). Recommendation of the Council on Artificial Intelligence. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
- OpenAI. (2023, July 5). Introducing Superalignment. https://openai.com/index/introducing-superalignment/
- Oxford English Dictionary. (n.d.). Govern, v. In OED Online. Oxford University Press. https://www.oed.com/dictionary/govern_v
- PBS. (2023, May 9). Geoffrey Hinton warns of the "existential threat" of AI. Amanpour and Company. https://www.pbs.org/video/godfather-of-ai-warns-of-the-existential-threat-of-ai-lj1i1c/
- UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000380455

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). (2024). *Official Journal of the European Union*, L, 2024/1689, 12.7.2024.

Romeo, G., & Conti, D. (2026). Exploring automation bias in human-AI collaboration: A review and implications for explainable AI. *AI & Society, 41*(1), 259-278. https://doi.org/10.1007/s00146-025-02422-7

White & Case LLP. (2026, August 4). *EU AI Omnibus enters into force, amending the AI Act*. https://www.whitecase.com/insight-alert/eu-ai-omnibus-enters-force-amending-ai-act

Disclaimer

I am not a lawyer, and this article does not provide legal advice. This is thought research and governance analysis based on public sources, cited materials, and human-AI review. It is intended to help executives, practitioners, insurers, and governance teams think more clearly about AI risk, liability exposure, and documentation practices. Readers should not rely on this article as a legal opinion, compliance determination, or substitute for qualified counsel. Any organization facing a legal, regulatory, contractual, or insurance question should consult its own attorney, broker, or professional adviser before acting.

The Other AI: Audio Briefings on Augmented Intelligence and AI Governance

Spotify | Apple Podcasts | Amazon Music | YouTube Playlist

#AIassisted using the HAIA Ecosystem | CC BY-NC-SA 4.0
Free for personal, educational, and noncommercial research use with attribution. Commercial exploitation, paid productization, and enterprise commercialization require separate permission and licensing.

Basil C. Puglisi, MPA
A Human-AI Collaboration

Contact: me@basilpuglisi.com | basilpuglisi.com