ARCPOINT CONSULTING · WHITE PAPER · AUGUST 2026

AI Governance as Organizational Design

Why the next era of AI governance is an operating model challenge

ArcPoint Consulting, LLC · Washington, DC · August 2026
Jermaine Leonard, AIGP · Principal

Executive summary

Jay Galbraith, whose Star Model is still the standard reference in organizational design, put structure in a single line. It "determines the location of decision-making power."[1]

AI systems now sit inside that location. They read unstructured information, rank alternatives, recommend actions, and in a growing number of processes carry those actions out. When something moves into the place where decisions get made, the organization has changed shape. Most companies book the change as a software purchase and govern it that way, with acceptable-use policies, model reviews, vendor questionnaires and risk tiers. All of those describe intended behavior. None of them distributes authority, and authority is what moved.

The scale is no longer in question. Stanford's 2026 AI Index reports that 88 percent of surveyed organizations used AI in at least one business function in 2025, up from 78 percent in 2024.[2] The same report counts 362 documented AI incidents in 2025 against 233 the year before, alongside a steady gap between the risks organizations name and the risks they do something about. Privacy was identified as a relevant risk by 63 percent and mitigated by 40. Cybersecurity was named by 72 percent and mitigated by 61.[3]

The safeguard most organizations reach for to close that gap has a measured failure rate. Ben Green surveyed 41 policies requiring human oversight of government algorithms and found that they "rest on an uninterrogated assumption: that people are able to effectively oversee algorithmic decision-making." His reading of the evidence is that they cannot, and that the policies then "provide a false sense of security" while letting institutions "shirk accountability for algorithmic harms."[4] Clinical medicine, which is two decades ahead of us in both deployment and study, has put a number on the same failure: a systematic review of computerized physician order entry found drug-safety alerts overridden "in 49% to 96% of cases."[5]

The governance instruments have meanwhile stopped asking for documents. NIST's AI Risk Management Framework requires that "mechanisms are in place and applied, and responsibilities are assigned and understood, to supersede, disengage, or deactivate AI systems that demonstrate performance or outcomes inconsistent with intended use."[6] The EU AI Act requires that overseers be able "to intervene in the operation of the high-risk AI system or interrupt the system through a 'stop' button or a similar procedure,"[7] and that deployers assign oversight to people "who have the necessary competence, training and authority, as well as the necessary support."[8] OMB M-25-21 tells federal agencies that where a high-impact use case fails its minimum practices, "the agency must safely discontinue use of the AI functionality."[9] Every one of those asks the organization to be able to do something, on demand, to a system that is currently running.

This paper takes the decision rather than the system as the unit of governance, treats reversibility as the boundary inside a decision where cost should step up, and sets out four conditions specific enough that you can check whether your human oversight is real. Meeting them is what ArcPoint's **AI Control Layer** — Visibility, Validation, Control and Auditability — exists to do.[10]

I. Structure is the location of decision-making power

Galbraith's model separates five design choices. Strategy "determines direction." Structure "determines the location of decision-making power." Processes concern "the flow of information." Rewards "provide motivation and incentives for desired behavior." People covers selection and development in alignment with the rest.[1]

Its use here is that it forces a question most AI programs skip. When a system starts to rank, score, prioritize or execute, which of the five just changed?

In almost every case, structure and processes. In almost every case, the organization records it as a technology purchase.

That pairing matters more than the others, and there is evidence behind the claim. Neilson, Martin and Powers analyzed survey data from more than 125,000 employees across roughly 1,000 organizations in more than 50 countries, and found decision rights and information flow to be the levers that determine whether strategy gets executed at all. "The single most common attribute of such companies," they wrote of strong performers, "is that their employees are clear about which decisions and actions they are responsible for." Structural moves and incentive changes worked best only after those two were settled.[11]

Applied to AI, that finding stings. An organization that drops AI into a process without settling who decides what has degraded the variable its own execution capacity depends on most, and has done so while believing it made a neutral technical change.

The traditional sequence in a credit, claims, hiring, procurement or service process ran from data collection to human analysis to human approval to execution. An AI-enabled version inserts machine analysis and recommendation ahead of the approval. A more automated version lets the system decide within thresholds and routes exceptions to people. An agentic version takes an objective and coordinates several systems before presenting a result or taking an authorized action. Four allocations of authority, one procurement label.

Accountability stays with people and institutions through all of it. Every instrument surveyed in this paper says so, some of them twice. But the system has become a participant in a decision structure built on the assumption that only people participate in it, and that is a design problem whoever remains legally responsible.

II. What the standards actually require

The first generation of enterprise AI governance was policy-led, and reasonably so. Acceptable use, data restrictions, prohibited applications, vendor review and a governance committee are cheap, fast and better than nothing.

Their limits show up in operation rather than in drafting. Policy describes what should happen; the operating model decides what does. A customer-service policy can require AI-generated communications to be accurate, appropriate and reviewed where necessary while leaving open every question that determines the outcome. Which interactions get reviewed. Whether the system can issue a refund. At what value a human has to approve. Whether an employee can override it. What happens when the model is uncertain. Who owns the result when an automated action harms a customer.

The regimes have already noticed. Read at the level of specific requirements rather than general orientation, they look less like policy frameworks than like control specifications.

NIST AI RMF 1.0 asks for mechanisms, and says so in those words: “Mechanisms are in place to inventory AI systems and are resourced according to organizational risk priorities” (GOVERN 1.6). “Policies and procedures are in place to define and differentiate roles and responsibilities for human-AI configurations and oversight of AI systems” (GOVERN 3.2). “Processes for human oversight are defined, assessed, and documented in accordance with organizational policies from the GOVERN function” (MAP 3.5). Then MANAGE 2.4, which carries more weight than the rest of the framework put together: “Mechanisms are in place and applied, and responsibilities are assigned and understood, to supersede, disengage, or deactivate AI systems that demonstrate performance or outcomes inconsistent with intended use.”[6] The framework is being revised under the July 2025 AI Action Plan. That does not disturb those subcategories, but it does mean any citation to it needs a date attached.[12]

ISO/IEC 42001:2023, published in December 2023, “specifies requirements for establishing, implementing, maintaining, and continually improving an Artificial Intelligence Management System (AIMS) within organizations,” intended for entities “providing or utilizing AI-based products or services.”[13] The load-bearing word is *system*. A management system in the ISO

sense means operating processes with defined ownership, audited on a cycle. What has grown up around the standard since is the part worth watching. ISO/IEC 42005:2025 covers AI system impact assessment; ISO/IEC 42006:2025 sets requirements for the bodies that audit and certify AI management systems.[14] There is now a standard for the auditors, which is where information security stood about a decade before customers began asking suppliers for certificates.

The **EU AI Act** requires high-risk systems to be built so that "they can be effectively overseen by natural persons during the period in which they are in use" (Article 14(1)), and to technically allow "the automatic recording of events (logs) over the lifetime of the system" (Article 12(1)).[15] Its oversight provisions get their own section below. Its timetable moved this summer: Regulation (EU) 2026/1744, the Digital Omnibus on AI, was published in the Official Journal on July 24, 2026 and entered into force on July 27, deferring the high-risk obligations in Sections 1, 2 and 3 of Chapter III to December 2, 2027 for Annex III systems and August 2, 2028 for Annex I. The deferral goes no further than that. The Article 5 prohibitions, the Chapter V general-purpose AI obligations and the Article 50 transparency requirements all kept their original dates, with Article 50 applying from August 2, 2026.[16]

OMB M-25-21 is the bluntest of the four. Federal agencies have to apply seven minimum risk management practices to high-impact AI — pre-deployment testing, an impact assessment, ongoing monitoring, adequate human training and assessment, human oversight with intervention and accountability, consistent remedies or appeals, and consultation with end users and the public — within 365 days, and "if a particular high-impact use case is not compliant with the minimum practices then the agency must safely discontinue use of the AI functionality." Relief runs through a named officer. An agency Chief AI Officer can waive a requirement only "after making a written determination, based upon a system-specific and context-specific risk assessment," and has to recertify every waiver annually.[9]

Different jurisdictions, different drafting traditions, one underlying demand. Each wants an inventory, a named human with authority, a testing cadence, an evidence trail and a way to stop the system. A document satisfies none of them.

III. Decision rights, bounded by reversibility

The useful question is not whether an AI system "makes decisions." It is what level of authority the organization handed it in a particular process. Four levels: inform, recommend, decide, act.

A system that *informs* supplies analysis while a person determines the outcome. One that *recommends* puts a default in front of that person, which is a weaker form of the same thing only in theory, since a default accepted almost every time is a decision the system is making with a human signature on it. One that *decides* settles the outcome inside stated parameters. One that *acts* carries that outcome into another system or business process.

Model capability does not separate these levels. An assistant that drafts a refund email and an agent that issues the refund can run on the same model. What separates them is whether the resulting action can be undone without a human, and if it cannot, who holds the authority to reverse it and how long that takes. Reversibility is where governance cost should step up, because it is where the cost of being wrong stops being recoverable.

Law drew a version of this line years ago. GDPR Article 22(1) gives a data subject "the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her," and where such decisions are permitted, Article 22(3) requires safeguards including "the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision."[17] The statutory trigger sits at the move from recommend to decide, plus consequence. An organization that cannot say which of its processes have crossed that line does not know whether Article 22 applies to it.

Which changes what an AI record has to contain. A system inventory — the artifact NIST GOVERN 1.6 asks for, and the one most programs build first — counts exposure. A decision inventory counts authority: for each consequential decision the organization makes, what feeds it, how far that influence runs, whether the action is reversible without a human, who is affected, and who can intervene.

The two lists diverge fast, and the divergence is the whole point. Twenty tools that draft text and summarize meetings fill twenty rows on a system inventory and none on a decision inventory. They are not harmless, since they remain a real data-protection and confidentiality question, but no amount of attention to them governs a decision. One system that sets a price, screens an applicant or releases a payment is a single row on the first list and the only row that matters on the second. A board that asks how many AI systems the organization runs has asked for the first list, but the number it wanted was on the second.

IV. Why human oversight usually isn't

Human oversight is the safeguard every instrument reaches for. The EU AI Act requires that overseers be enabled, as appropriate and proportionate, "to remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system (automation bias), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons"; "to decide, in any particular situation, not to use the high-risk AI system or to otherwise disregard, override or reverse the output"; and "to intervene in the operation of the high-risk AI system or interrupt the system through a 'stop' button or a similar procedure that allows the system to come to a halt in a safe state."[7]

Automation bias is named in the statute, in parentheses, as a thing the overseer must be equipped to resist. The drafters were right to put it there.

Green's study of those 41 oversight policies found two connected flaws. Evidence suggests "that people are unable to perform the desired oversight functions," and because of that, the policies "legitimize government uses of faulty and controversial algorithms without addressing the fundamental issues with these tools."[4] His proposed fix travels well outside government: replace human oversight with institutional oversight, in which an organization has to justify, with evidence, that its chosen form of human review will actually work before the system goes in.

The clinical numbers show what that failure looks like at scale. Van der Sijs and colleagues, reviewing overrides of drug-safety alerts in computerized physician order entry, found alerts overridden "in 49% to 96% of cases."[5] Trained professionals, high-stakes setting, dismissing a safety mechanism their own institution installed. Carelessness is not the explanation. A control that fires constantly and is usually wrong teaches the people underneath it to clear it without reading, and they do that because the alternative is not finishing their shift. Bainbridge saw the general shape of this in 1983: automation reassigns the human from operator to monitor, which is the harder job, and it erodes the very skill the rare intervention will need.[18]

So human oversight is a design, designs have failure modes, and writing "human in the loop" into a governance document implements nothing. Oversight becomes real where four things hold together, each of which can be checked in an afternoon.

Information. The reviewer can see what the system used and why it produced this output, at the moment of review, not in a model card written a year ago.

Competence and time. The reviewer has the expertise to disagree and enough time per case to use it. The AI Act makes this a deployer obligation in terms: oversight goes to natural persons

“who have the necessary competence, training and authority, as well as the necessary support.”[8]

Independence. The reviewer does not report to the team whose output is under review, and is not measured on the throughput the system exists to raise.

Authority, exercised. The reviewer can stop the process, and has. An intervention right nobody has ever used looks the same at audit as one that does not exist.

One measurement tests all four at once, and it is arithmetic rather than interpretive. Track the override rate. If reviewers agree with the system on essentially every case, either the system is extraordinary or the review is ceremonial, and the burden of proof sits with the first explanation. Watch what the rate does as case volume rises. Treat a collapse toward zero as a control failure rather than a productivity win, and put the number in the same management pack as the throughput it is competing against. In assessment work, this is the metric organizations are most surprised to find nobody owns.

V. Where accountability goes

Take the ordinary case. An AI system recommends rejecting a customer request. An employee accepts the recommendation. The model came from a vendor, the workflow was configured by an internal technology team, the risk function approved the use case, and the business unit owns the customer relationship.

The customer challenges the outcome, and every party can say truthfully that they did their part correctly.

Helen Nissenbaum named this in 1996, before any of the technology in question existed. Her four barriers to accountability in computerized societies are “1) the problem of many hands, 2) the problem of bugs, 3) blaming the computer, and 4) software ownership without liability.”[19] Thirty years on, a workflow spanning a foundation-model provider, an application vendor, an integrator, an internal platform team, a risk function and a business owner is close to a purpose-built machine for producing the first.

The second failure mode is worse, because organizations construct it deliberately while believing they are adding a safeguard. Madeleine Clare Elish calls it the moral crumple zone, describing “how responsibility for an action may be misattributed to a human actor who had limited control over the behavior of an automated or autonomous system.” In her words: “just as the crumple zone in a car is designed to absorb the force of impact in a crash, the human in a

highly complex and automated system may become simply a component — accidentally or intentionally — that bears the brunt of the moral and legal responsibilities when the overall system malfunctions."[20]

Set that beside Green's finding and the risk is hard to miss. An organization that inserts a human approval step it has not equipped has not added oversight; it has nominated somebody to absorb the impact. The reviewer clicking approve on their four-hundredth case of the day will carry the accountability. The design that made real review impossible will carry none of it.

The fix is to name authority rather than participation. Every consequential AI-enabled process needs a named business owner accountable for the outcome. Technology can own implementation, security the technical controls, legal and compliance the interpretation of obligations, risk the challenge mechanisms, and vendors whatever their contracts say. None of that removes the need for one internal owner of the business decision. The OECD AI Principles, updated in May 2024, tie accountability to that same bundle of roles, context, traceability and ongoing risk management across the lifecycle.[21]

And the truest test of whether accountability is real has nothing to do with who signs the approval. It is who can suspend the system. NIST puts it as responsibilities "assigned and understood" for mechanisms to "supersede, disengage, or deactivate."[6] M-25-21 puts the waiver in a named officer who has to recertify it every year.[9] Both are describing an authority with a person's name on it and a record of use.

VI. Exception governance, and how it rots

The strongest organizational opportunity AI creates is a redesign of where scarce human attention goes. Traditional escalation uses hierarchy: the frontline employee sends the unusual case up to a supervisor. AI-enabled processes can escalate on richer signals — confidence, consequence, transaction value, anomaly indicators, defined exception conditions.

That makes exception governance possible. Routine, reversible, low-impact cases proceed with light intervention, and cases with higher consequence, unusual features, uncertain outputs or conflicting signals route to people with the expertise and the authority to act on them.

Both extremes are easy to price. Requiring manual approval of every AI-supported action preserves control by discarding most of the value of the automation, and automating everything scales risk faster than productivity. The design question is where to sit between them, process by process, given reversibility.

What the literature adds is a warning the consulting version of this argument tends to leave out. Exception routing decays, and it decays in one direction. Clinical override is where that decay has already been measured. A threshold set to catch everything worth catching produces more exceptions than the reviewing population can absorb; the population adapts by clearing them faster; the control turns into a formality while its coverage statistics stay at 100 percent.[5] Nobody files an incident report about any of that. Thresholds drift toward whatever the reviewers have time for, and the drift stays invisible unless somebody watches for it.

Three numbers make it visible, and they belong in monthly management reporting rather than an annual audit: exceptions per reviewer per day, override rate, and time-to-decision on escalated cases. When volume climbs while override rate falls, the threshold has stopped functioning as a control and has become a queue with a governance label on it.

Managerial work changes here too, usually by discovery rather than design. Approval shrinks. A supervisor who used to sign individual transactions ends up setting the thresholds at which transactions route for signature, watching what arrives, and judging when the routing rule itself was wrong. That is a different job against different evidence, and most organizations go on measuring the old one for years. The workforce question raised by AI is not only which tasks get automated. It is what managing becomes when a manager’s product is a threshold rather than an approval.

VII. The AI Control Layer

Design produces requirements, and requirements need a mechanism. Ours is the AI Control Layer. This paper uses the definition set out in our companion white paper without varying it, because a framework that means different things in different documents is not a framework.[10]

The layer is defined by what it can stop. Policies, principles and committees describe intended behavior, intended values and intended attention; the layer is the part that acts. It also sits above the system it governs rather than inside it. A vendor’s built-in safety filter, a governance feature in an MLOps platform, or a risk register kept by the team whose work it evaluates may each be excellent, and none of them has a reporting line the operators do not control. That makes them settings.

Inside that boundary, four functions.

FUNCTION	OPERATING REQUIREMENT	THE REQUIREMENT IT SATISFIES
Visibility	A current inventory of AI systems *and* of the decisions they influence, including AI embedded in other software and tools adopted without approval, each with a named owner, the data touched, and whether the action is reversible.	NIST GOVERN 1.6 (inventory mechanisms); the predicate for knowing whether GDPR Art. 22 applies.
Validation	Testing before deployment and on a defined cadence, against criteria written in advance, producing an artifact someone outside the building team can read. A benchmark describes a model; a validation describes *your* deployment of it.	NIST MEASURE; M-25-21 pre-deployment testing and ongoing monitoring; ISO/IEC 42005 impact assessment.
Control	Named human accountability per consequential decision, an enforcement point where policy binds rather than advises, and intervention authority that has been exercised.	NIST MANAGE 2.4 (supersede, disengage, deactivate); EU AI Act Art. 14(4)(d)–(e) and Art. 26(2); M-25-21 discontinuation.
Auditability	Evidence produced as a by-product of operation, not assembled when someone asks. If the evidence is generated by the request for evidence, what exists is a reconstruction.	EU AI Act Art. 12(1) automatic logging and Art. 26(6) log retention; OECD traceability.

They fail in a chain. Policy without visibility cannot identify what is being governed. Visibility without validation shows deployment but not performance. Validation without control finds problems nobody is empowered to act on. Control without auditability may handle the incident and still leave the organization unable to demonstrate what happened, and an intervention nobody can reconstruct six months later tends to get reversed.

One clarification for anyone building against this. The four functions are not maturity stages. An organization either has each one or does not, and the weakest determines what the layer can do. A program with excellent monitoring and no enforcement authority will observe its own incident in high resolution.

VIII. Three processes

In financial operations, a system that flags anomalous payments for human investigation sits at recommend. Authorizing it to suspend transactions meeting defined criteria moves it to act, and suspension is reversible, which is what makes it a defensible first delegation. The governance follows from there. Visibility identifies where automated suspension is happening. Validation tests false positives and drift as fraud patterns change. Control names who may alter thresholds and who may release a held transaction. Auditability preserves what is needed to explain the hold to the customer, and later to the regulator.

Human resources is the harder case. Summarizing applications is inform. Ranking or screening candidates is decide, and the consequence is a person's access to employment. In the EU that is Annex III territory, with high-risk obligations now applying from December 2, 2027 rather than August 2, 2026. This is a deferral of the requirements, not of the exposure.[16] Oversight design is where these programs usually come apart, because a recruiter working through a ranked list under time pressure meets none of the four conditions in Section IV, and no contract moves the accountability to the software provider.

Supply chain runs the same ladder with money attached. Forecasting shortages is inform, recommending inventory changes is recommend, placing orders within parameters is act, and an order placed at scale is expensive to unwind. So the governance content here is financial thresholds, supplier restrictions, exception criteria, monitoring, and a named authority who can halt ordering — rather than a policy stating that AI will be used responsibly in procurement.

One principle runs under all three. Governance requirements should follow decision consequence and delegated authority, not the label attached to the technology.

IX. Five postures

Organizations sit at different points in this transition. The table below describes what an organization has settled about authority. It is not a maturity model for the four functions, which, as Section VII says, either exist or do not.

POSTURE	WHAT HAS BEEN SETTLED
1. Policy-led	Acceptable use, approved tools, prohibited activities, baseline compliance. Nothing has been settled about authority.
2. Risk-classified	Use cases are inventoried and differentiated by impact, regulatory exposure and risk. The organization knows what it has.
3. Decision-governed	Decision rights, delegated authority, oversight design, accountability and escalation are explicit per process. The organization knows who decides.
4. Control-integrated	Visibility, Validation, Control and Auditability are wired into workflows and management reporting rather than asserted in documents. The organization can act.
5. Adaptive	Autonomy is raised or lowered on evidence, as performance, risk and conditions change. The organization can learn.

Each posture builds on the one before it, and policy survives the whole climb. It just stops being the entirety of the program. What distinguishes the last posture is that autonomy becomes a variable under active management, where most organizations set it once, at procurement, and never revisit it in either direction.

X. The case for spending now

AI governance is usually sold defensively: prevent harm, protect data, meet obligations, manage third-party risk, preserve trust. Those are legitimate and they are also half the argument.

The best evidence for the other half is in the 2026 AI Index. Asked what obstructs the scaling of agentic AI, 62 percent of respondents named security and risk concerns, ahead of technical limitations at 38 percent and regulatory uncertainty at 38 percent.[3] The binding constraint on the most valuable current class of deployment turns out to be neither capability nor law. Organizations cannot supervise what they would be deploying. On that evidence the control layer is not a tax on adoption but the precondition for it.

The same capability answers a customer's security review, a partner's diligence questionnaire and an insurer's underwriting question, none of which appear on any regulator's calendar. The certification infrastructure is arriving on schedule too, since ISO/IEC 42006:2025 exists to accredit the bodies that will audit AI management systems.[14] Attestation follows accreditation, and customer requirements follow attestation, usually within a couple of procurement cycles.

There is a real counterargument and it deserves stating at full strength. The EU just moved a major deadline sixteen months to the right. Tooling gets cheaper and better every quarter. Capital spent early on governance is capital not spent on the thing being governed. We sell this work, so weigh our answer accordingly.

The answer is asymmetry. Building early carries an overhead you can price in advance and pay in installments. Building late costs the use of the system, on a date somebody else picks, which is what federal agencies discovered on April 3, 2026 when M-25-21's deadline arrived with discontinuation attached.[9] One error shows up as a line item. The other shows up as an outage.

And a concession. If nothing you run with AI touches a person's access to money, safety, employment, benefits, housing or legal standing, most of this paper is premature for you, and the proportionate response is an inventory, one named person who can switch things off, and a

date in the calendar to revisit. What does not scale down is that last item. There is no size of organization at which it is acceptable that nobody has the standing to say no.

XI. Conclusion

The first phase of AI governance addressed principles, policies, privacy, security, model risk and acceptable use. Those foundations hold. They describe intent, though, and what changed is the location of authority.

Structure determines where decisions get made. AI now participates in that structure, which makes this an organizational design question before it is a compliance question. Decision rights have to be explicit, and delegated authority has to be bounded by reversibility. Human oversight has to meet conditions you can test, because the default design fails and the failure is invisible from inside the workflow. Accountability needs a name rather than a committee, or the design will find somebody to absorb the impact. And someone has to be able to stop the system, on the record.

The AI Control Layer makes those requirements operational through Visibility, Validation, Control and Auditability. It replaces nothing — not NIST, not ISO, not the OECD principles, not any regulator's obligations. It answers a different question: how an organization turns governance expectations into the daily architecture through which people and intelligent systems work together.

Which is why the defining question for senior leaders is no longer "how should we control AI?" It is this. When AI holds part of the decision-making power, how should the organization be designed?

That is a structural question. It was always the harder one.

About ArcPoint

ArcPoint Consulting is a senior-led AI governance, risk, and control advisory practice in Washington, DC. We do not help organizations adopt AI. We help them govern it.

We build the AI Control Layer: the evidence, oversight, and documentation that make an AI deployment visible, validated, controlled, and auditable. It is produced by ArcPoint's Weftline

Framework, which assesses five dimensions together: Strategic Intent, Decision Authority, Operational Integration, Workforce Enablement, and Technical Infrastructure, with AI Influence Mapping running across all five. The framework builds on the NIST AI Risk Management Framework and adds structured ethical value elicitation drawn from IEEE 7000 along with a defined bias-testing protocol. Jurisdiction-specific requirements layer on top.

Every engagement is delivered and signed by the principal.

Jermaine Leonard, AIGP · ArcPoint Consulting, LLC · Washington, DC Certified Business Enterprise, District of Columbia, six designations · UEI JE2DLNX4BVQ3

This paper is analysis and general professional commentary. It is not a compliance assessment of any organization, and it is not legal advice.

Endnotes

1. Jay R. Galbraith, the Star Model™, Galbraith Management Consultants. Strategy "determines direction"; structure "determines the location of decision-making power"; processes concern "the flow of information"; rewards "provide motivation and incentives for desired behavior"; people covers selection and development in alignment with the other policies. https://jaygalbraith.com/services/star-model/ — read 2026-08-22.

2. Stanford Institute for Human-Centered Artificial Intelligence, *2026 AI Index Report*, Chapter 4 (Economy): "A large majority of respondents reported that their organization uses AI in at least one business function, up to 88% in 2025 from 78% in 2024." The underlying instrument is McKinsey's annual State of AI survey of self-selected respondents, not a census of organizations. **Disclosure:** the same chapter states generative-AI adoption two ways — a summary figure of 70% of organizations using generative AI in at least one business function, and a narrative figure of "79% of respondents reporting that their organizations regularly use generative AI in at least one business function, compared to 71% in 2024." Because the two measures could not be reconciled to a single figure reference, no generative-AI adoption percentage is asserted in the body of this paper. https://hai.stanford.edu/assets/files/ai_index_report_2026_chapter_4_economy.pdf — read 2026-08-22.

3. Stanford HAI, *2026 AI Index Report*, Chapter 3 (Responsible AI): 362 documented AI incidents in 2025 against 233 in 2024; privacy risk recognized by 63% and mitigated by 40%; cybersecurity recognized by 72% and mitigated by 61%; security and risk concerns named as the primary obstacle to scaling agentic AI by 62% of respondents, against 38% for technical

limitations and 38% for regulatory uncertainty. The same chapter reports that the share of organizations with no responsible-AI policies fell from 24% in 2024 to 11% in 2025, and that AI-specific governance roles grew 17% in 2025. https://hai.stanford.edu/assets/files/ai_index_report_2026_chapter_3_responsible_ai.pdf — read 2026-08-22.

4. Ben Green, "The Flaws of Policies Requiring Human Oversight of Government Algorithms," *Computer Law & Security Review* (2022). Surveys 41 policies; quotations from the published abstract. https://arxiv.org/abs/2109.05067 and https://www.sciencedirect.com/science/article/pii/S0267364922000292 — read 2026-08-22.

5. Heleen van der Sijs, Jos Aarts, Arnold Vulto and Marc Berg, "Overriding of Drug Safety Alerts in Computerized Physician Order Entry," *Journal of the American Medical Informatics Association* 13(2), 2006, 138–147: drug safety alerts are overridden "in 49% to 96% of cases." https://academic.oup.com/jamia/article-abstract/13/2/138/729701 — full text read 2026-08-22.

6. National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1, January 2023. GOVERN 1.6: "Mechanisms are in place to inventory AI systems and are resourced according to organizational risk priorities." GOVERN 3.2: "Policies and procedures are in place to define and differentiate roles and responsibilities for human-AI configurations and oversight of AI systems." MAP 3.5: "Processes for human oversight are defined, assessed, and documented in accordance with organizational policies from the GOVERN function." MANAGE 2.4: "Mechanisms are in place and applied, and responsibilities are assigned and understood, to supersede, disengage, or deactivate AI systems that demonstrate performance or outcomes inconsistent with intended use." https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf — read 2026-08-22.

7. Regulation (EU) 2024/1689 (AI Act), Article 14(4)(b), (d) and (e), consolidated text as of July 27, 2026. https://eur-lex.europa.eu/eli/reg/2024/1689/2026-07-27/eng/html — read 2026-08-22.

8. Regulation (EU) 2024/1689, Article 26(2). Same source as endnote 7.

9. Office of Management and Budget, Memorandum M-25-21, "Accelerating Federal Use of AI through Innovation, Governance, and Public Trust," April 3, 2025. Seven minimum risk management practices at §4(b); the 365-day documentation requirement and the discontinuation sentence at §4(a)(i); Chief AI Officer waiver authority at §4(a)(ii). https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-21-Accelerating-Federal-Use-of-AI-through-Innovation-Governance-and-Public-Trust.pdf — read 2026-08-22.

10. ArcPoint Consulting, *The AI Control Layer: Why every AI-enabled organization ends up building one, and what it costs to build it late*, August 2026. The definition of the layer, the stop test and the separation requirement are set out there and are used here without variation.

11. Gary L. Neilson, Karla L. Martin and Elizabeth Powers, "The Secrets to Successful Strategy Execution," *Harvard Business Review*, June 2008. Based on data from "more than 125,000 employees of some 1,000 organizations in more than 50 countries." https://hbr.org/2008/06/the-secrets-to-successful-strategy-execution — read 2026-08-22.

12. NIST states that "the AI RMF 1.0 is being revised as part of the White House AI Action Plan"; the directive appears in *America's AI Action Plan*, July 2025. No target date or comment period has been published as of this writing. https://www.nist.gov/itl/ai-risk-management-framework — read 2026-08-22.

13. ISO/IEC 42001:2023, *Information technology — Artificial intelligence — Management system*, published December 18, 2023. Quotations are from ISO's published description of the standard. Note that ISO does not publish the clause structure or the Annex A control list on its free pages; this paper therefore asserts neither. https://www.iso.org/standard/81230.html — read 2026-08-22.

14. ISO/IEC 42005:2025, *AI system impact assessment*, published May 28, 2025 (https://www.iso.org/standard/44545.html); ISO/IEC 42006:2025, *Requirements for bodies providing audit and certification of artificial intelligence management systems*, published July 7, 2025 (https://www.iso.org/standard/42006) — both read 2026-08-22.

15. Regulation (EU) 2024/1689, Articles 14(1) and 12(1). Same source as endnote 7.

16. Regulation (EU) 2026/1744 of the European Parliament and of the Council of 8 July 2026 amending Regulations (EU) 2024/1689, (EU) 2018/1139 and (EU) 2023/1230 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI). Published in the Official Journal July 24, 2026; entered into force July 27, 2026. Recital (40) confines the deferral to "Sections 1, 2 and 3 of Chapter III," setting December 2, 2027 for systems classified as high-risk under Article 6(2) and Annex III and August 2, 2028 for those under Article 6(1) and Annex I. Recital (38) adds a four-month transitional period for the Article 50(2) marking obligation for systems already on the market before August 2, 2026. https://eur-lex.europa.eu/eli/reg/2026/1744/oj/eng/html — read 2026-08-22.

17. Regulation (EU) 2016/679 (GDPR), Article 22(1) and 22(3). https://eur-lex.europa.eu/eli/reg/2016/679/2016-05-04/eng/html — read 2026-08-22.

18. Lisanne Bainbridge, "Ironies of Automation," *Automatica* 19(6), November 1983, 775–779, doi: 10.1016/0005-1098(83)90046-8. Characterization of the argument; not quoted.

19. Helen Nissenbaum, "Accountability in a Computerized Society," *Science and Engineering Ethics* 2(1), 1996, 25–42. Quotation from the published abstract. https://link.springer.com/article/10.1007/BF02639315 — read 2026-08-22.

20. Madeleine Clare Elish, "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction," *Engaging Science, Technology, and Society* 5, 2019, 40–60. https://estsjournal.org/index.php/ests/article/view/260 — full text read 2026-08-22.

21. Organisation for Economic Co-operation and Development, *OECD AI Principles* (OECD/LEGAL/0449), adopted 2019, updated May 2024. https://oecd.ai/en/ai-principles — read 2026-08-22.

Sources and verification

`/verify` pass run 2026-08-22. Every citation was checked against the source named in the endnote on that date. Status key: **P** = primary source read directly · **S** = secondary or publisher metadata · **D** = discrepancy disclosed in the text.

#	CLAIM	SOURCE	STATUS	RESULT
1	Star Model five design policies; "structure determines the location of decision-making power"	jaygalbraith.com	P	Confirmed
2	AI use at 88% of surveyed organizations in 2025, up from 78% in 2024; measure is "at least one business function"	AI Index 2026, Ch. 4 PDF	P / D	Confirmed. Generative-AI figure deliberately not asserted — the chapter states 70% and 79% under different measures and the figure reference could not be pinned
3	362 incidents vs 233; privacy 63/40; cyber 72/61; agentic obstacle 62% vs 38%/38%; no-policy share 24%→11%; governance roles +17%	AI Index 2026, Ch. 3 PDF	P	Confirmed
4	Green: 41 policies surveyed; people "unable to perform the desired oversight functions"; "false sense of security"	arXiv 2109.05067 abstract; CLSR 2022	P	Confirmed verbatim from the published abstract
5	Drug safety alerts overridden "in 49% to 96% of cases"	van der Sijs et al., JAMIA 13(2):138–147	P	Confirmed from the article text

#	CLAIM	SOURCE	STATUS	RESULT
6	NIST GOVERN 1.6, GOVERN 3.2, MAP 3.5, MANAGE 2.4 wording	NIST AI 100-1 PDF	P	Confirmed verbatim. Note: the AIRC Playbook renders MANAGE 2.4 without its internal commas; the AI 100-1 PDF wording is used
7	AI Act Art. 14(1), 14(4)(b), (d), (e)	EUR-Lex consolidated text, July 27, 2026 version	P	Confirmed verbatim
8	AI Act Art. 26(2) competence, training, authority, support	EUR-Lex consolidated text	P	Confirmed verbatim
9	AI Act Art. 12(1) automatic logging	EUR-Lex consolidated text	P	Confirmed verbatim
10	AI Act Art. 26(6) log retention	European Commission AI Act Service Desk	S	Confirmed on the Commission's official reproduction; the EUR-Lex consolidated text truncated before this provision. Cited in the table only, not quoted in the body
11	Digital Omnibus: Reg. (EU) 2026/1744, OJ July 24, 2026, in force July 27, 2026, Annex III to Dec 2, 2027, Annex I to Aug 2, 2028, deferral confined to Ch. III Sections 1–3	EUR-Lex, Regulation (EU) 2026/1744, recital 40; EUR-Lex metadata	P	Confirmed. The amended text of Article 113 itself was not retrievable verbatim; the recital is quoted instead
12	GDPR Art. 22(1) and 22(3)	EUR-Lex	P	Confirmed verbatim
13	M-25-21: seven minimum practices, 365 days, "the agency must safely discontinue use of the AI functionality," CAIO waiver	whitehouse.gov PDF	P	Confirmed verbatim. Seven is correct; the waiver authority is the agency CAIO, not OMB
14	NIST AI RMF under revision under the AI Action Plan	nist.gov; America's AI Action Plan PDF	P	Confirmed. No target date published
15	ISO/IEC 42001:2023 title, publication date, AIMS description	iso.org	P	Confirmed. Clause structure and Annex A control list are **not** published on ISO's free pages and are therefore not asserted anywhere in this paper
16	ISO/IEC 42005:2025 and 42006:2025 titles and publication dates	iso.org	P	Confirmed
17	Nissenbaum's four barriers	Springer, Science and Engineering Ethics 2(1)	P	Confirmed verbatim from the published abstract
18	Elish moral crumple zone definition	ESTS 5:40–60 full text	P	Confirmed verbatim

#	CLAIM	SOURCE	STATUS	RESULT
19	Neilson/Martin/Powers survey base and decision-rights finding	hbr.org	P	Confirmed verbatim
20	Bainbridge 1983 citation	Automatica 19(6):775–779, DOI	S	Citation confirmed through publisher metadata; the argument is characterized, not quoted, because the full text was not read
21	OECD AI Principles updated May 2024	oecd.ai	P	Confirmed

Two citations were considered and dropped rather than hedged. Skitka, Mosier and Burdick's automation-bias experiments (*International Journal of Human-Computer Studies*, 1999) would have supported Section IV directly, but every full-text route returned 403 or robots-disallowed and the findings could not be verified to the standard used elsewhere here; the AI Act's own automation-bias provision and the clinical override evidence carry the point instead. Jensen and Meckling's work on co-locating decision rights with knowledge was not verifiable within this pass and is omitted.