

Governing Artificial Intelligence

A practical framework for authority accountability and judgment



John Gnotek | Gnotek Ai | 9-24-26

Organizations are giving AI access to information, influence over decisions, and sometimes the ability to act. Many leaders have approved those capabilities without a clear account of what the system may do, what it may infer, how its work will be checked, or who can stop it. The central governance problem is the delegation of intelligence and authority under uncertainty.

This paper offers an overall, plain language framework for leaders, particularly those who are new to AI. It explains the different kinds of AI and their relationship to conventional software; distinguishes governance from compliance, risk management, ethics, and security; and gives seven questions for approving and supervising an AI use. It then translates those questions into ownership, testing, vendor oversight, monitoring, incident response, and a path for adoption. Its purpose is to support the use of established frameworks, *not* to claim certification or replace applicable law.

Central proposition. The authority granted to an AI system should *never* exceed the organization's demonstrated ability to understand its purpose, test its performance, oversee its use, intervene when it fails, and remain accountable for the consequences. Human accountability is an institutional obligation, not a ritual approval click.

The **Gnotek Ai** contribution is a **decision lens**: identify the intended use and context, bound what the system may know and do, assess consequences and uncertainty, demand evidence proportionate to the stakes, assign a real owner, and preserve intervention. These judgments come before choosing the detailed controls. [1–4]

1 The essential differences

Artificial intelligence and its common forms

AI is a broad category of machine-based systems that infer from inputs to produce outputs such as predictions, content, recommendations, or decisions. The boundary between AI and other software is not always sharp. The important governance question is what a system does in its actual setting, including the tools and people connected to it. [1]

Predictive AI estimates a future or unknown outcome from patterns in data, such as equipment failure or demand. Its outputs are estimates, not facts.

Classification and ranking systems sort cases or people into categories or ordered lists; a score can quietly influence a human decision even when the system has no formal decision authority.

Generative AI produces new text, images, audio, code, or other content in response to a prompt or context. Fluency and apparent coherence do not establish truth.

Agentic AI is a design pattern in which a model can plan or select steps and use connected tools; its authority depends on its permissions, boundaries, and review gates, not on the word “agent.”

AI-enabled automation embeds one or more of these capabilities in a workflow that can trigger actions repeatedly, with or without a person reviewing each case.

These categories overlap. A customer service workflow can classify a complaint, generate a reply, query an account, and issue a credit. Govern the complete workflow and its consequences, not merely the model that supplies a sentence.

Conventional software and AI

Conventional software generally follows specified rules written or configured by people. Given the same code, inputs, and environment, a deterministic rule ordinarily produces the same result. Its governance emphasizes requirements, access control, software testing, change management, cybersecurity, and audit logs. Conventional software can still be complex, unsafe, adaptive, or stochastic; none of those controls becomes obsolete with AI.

Many AI systems instead infer a mapping from training data, produce outputs that are sensitive to context, and may respond differently to wording, retrieved material, model versions, or distribution shifts. A model can produce a plausible answer without a sound basis. Its behavior in an open-ended domain cannot usually be specified exhaustively as rules or tested against every possible input. The system also changes in practical effect when users, data, connected tools, incentives, or vendor models change. NIST describes AI risk as emerging from both technical design and social context and notes that changes in training data and deployment conditions can affect trustworthiness in difficult-to-understand ways. [1]

Why ordinary software governance is insufficient. A passing security review or a successful installation does not establish that the AI's conclusions are appropriate for a particular population, that a reviewer can detect a wrong answer, or that a model update preserves behavior. AI adds questions about evidence for outputs, data provenance, representativeness, inference about people, misleading confidence, human reliance, ongoing evaluation, and the authority of connected tools. The organization should extend its existing IT, security, privacy, quality, and procurement controls with these AI-specific checks. It should not declare all AI inherently unpredictable or all conventional software perfectly predictable.

Consider a hiring tool. A rule that rejects an incomplete application can be specified and tested against the required fields. An AI summary may omit a relevant qualification, a ranking model may encode historical disadvantage, and a conversational agent may give applicants inconsistent explanations. Even if the interface and hosting pass IT review, the organization must test job relevance, disparate effects, review quality, appeal routes, and the limits on automated rejection.

Failure and misuse patterns a leader should recognize

An AI output can be *false but persuasive*: a generated citation, customer history, or analysis can look complete while resting on invented detail. It can be *selectively incomplete*: a summary that omits a caution may be more damaging than one with an obvious error. It can *infer beyond the supplied facts*: correlations can reveal sensitive personal or commercial attributes even if nobody directly entered them. It can *reproduce historical patterns* that are inappropriate for the new setting. It can *launder authority*: users may treat a confident recommendation as if it came from a policy, clinician, engineer, or executive. It can *expand through integration*: retrieval, connectors, and automation can expose data or take action at a scale no individual prompt suggests. It can also be *deliberately misused* to profile people, manipulate messages, forge evidence, or bypass normal approvals.

These are pathways to test for a particular use, not claims that every model will fail in every way. The source of harm can be a model defect, bad source data, poor interface design, excessive permission, organizational pressure, or deliberate exploitation. Governance should examine the whole arrangement, including incentives and the ability of people affected by it to challenge an outcome.

Governance and its neighboring disciplines

Governance determines the purpose, permitted authority, ownership, decision rights, oversight, and conditions for continuing or stopping a use.

Risk management identifies possible benefits and harms, assesses their likelihood and severity where feasible, selects responses, and monitors residual risk.

Compliance establishes and demonstrates adherence to applicable law, contracts, and binding organizational requirements.

Ethics asks what is justified in relation to people, fairness, dignity, and wider effects, including actions a rule might permit.

Security protects systems, data, and availability against compromise or misuse.

Privacy and data governance govern lawful, appropriate collection, access, use, inference, retention, sharing, and deletion of information.

These disciplines overlap and should inform one another. Governance is exercised by people with real decision rights. *A policy without an owner, review capacity, evidence, or a stop mechanism is an aspiration rather than a working system.*

2 The delegation problem

An AI system can **inform** by retrieving or summarizing information; **recommend** by proposing an option; **decide** by selecting an outcome according to delegated criteria; **act** by changing a record, sending a message, purchasing, or controlling equipment; and **automate** by repeatedly making and executing such choices within a workflow.

These are functional descriptions, not universal legal categories. One workflow may occupy several levels at once. “*Decide*” describes selecting an outcome; “*act*” describes executing it; “*automate*” describes ongoing execution without a separate human decision on every instance.

Moving rightward generally increases delegated authority, but the impact is not determined by that ladder alone. A recommendation used unquestioningly can have more practical power than a low-value reversible automatic action. Influence, scale,

access, and user behavior matter. A genuine human decision requires enough time, competence, information, and freedom to disagree. A person approving thousands of recommendations per day may provide no meaningful check.

Authority over **knowledge** deserves separate treatment. A system may be permitted to read a document, retain a conversation, combine databases, infer sensitive traits, construct a behavioral profile, or disclose a derived conclusion. These permissions can cause harm even if the system never sends an email or executes a purchase.

The inventory and approval record should therefore specify both action permissions and knowledge permissions: access, combination, inference, memory, reuse, sharing, and deletion.

The organization should distinguish information explicitly supplied from information inferred by combining sources. A derived profile can be sensitive even when each source document was individually innocuous. Memory can improve continuity while increasing exposure and making consent, correction, and deletion harder. Govern retention and inference as **permissions**, not merely by labeling incoming files confidential or public. Affected people may need to know that a consequential inference was used and have a practical route to correct its underlying facts.

The degree of oversight should rise with the potential consequence, uncertainty, scale, and difficulty of reversal. This is a reasoning relationship, not a numerical equation or a claim that those factors can be multiplied into one valid score.

High consequence calls for stronger evidence, narrower authority, more independent review, and a tested intervention path. Some uses may remain unacceptable even with controls. Legal prohibitions and rights-based limits cannot be traded against an attractive average accuracy.

3 The seven governance questions

1 Intent

What problem or decision is the system meant to improve, for whom, and by what observable outcome? Who might bear a cost? Why is AI preferable to a simpler process, better data, training, or conventional automation? State the allowed purpose and foreseeable misuse before procurement. A use-case record should name the owner, users, affected people, baseline process, expected benefit, and criteria for success or withdrawal. Compare the cost of adoption with the cost of delay, including unmanaged shadow use, while resisting an assumption that either adopting or abstaining is automatically prudent. This operationalizes Gnotek Ai's "judgment before automation" principle. [5]

2 Context

What information and domain knowledge does the system need, and what does it lack? Which population, language, geography, time period, process, and edge cases define the actual deployment? Who can recognize when local circumstances make an answer inappropriate? Record assumptions and dependencies. A model that performs well in a vendor demonstration may fail where local terminology, outdated records, or rare cases dominate. **Organizational situational awareness** includes preserving the difference between a signal, an inference, and a conclusion that someone has authorized for action.

3 Authority

May the system inform, recommend, decide, act, or automate a recurring workflow? Define exact permissions, transaction limits, affected records, allowed recipients, human approval gates, and prohibited actions. Define separately what it may access, combine, infer, retain, or disclose. Configure technical permissions to match the policy. A natural language instruction alone is a weak boundary for an agent with powerful tool access.

4 Consequence

What happens if the answer is wrong, late, selectively incomplete, persuasive but unfounded, or used outside its scope? Consider harm to individuals and groups, safety, legal rights, finances, operations, reputation, and downstream systems. Assess severity, exposure, scale, reversibility, and whether affected people can notice and correct an error. *Do not reduce every concern to a single score*: a rare severe harm can matter even when average performance is strong.

5 Evidence

What evidence supports use in this setting? Demand representative tests, observed failure modes, source provenance where relevant, independent checking for consequential claims, and clear limits. Distinguish facts in source material from a model's inference. Document the version and configuration tested, test population, thresholds, human review results, and residual uncertainty. A benchmark score, certification claim, or vendor demonstration is a starting point, not proof of fitness for the local workflow. NIST's Map, Measure, and Manage functions provide a deeper risk-management structure. [1, 2]

6 Accountability

Which named person owns the deployment and has budget, authority, and time to govern it? Who is responsible for approval, daily operation, data stewardship, independent challenge, incident response, and legal review? Who accepts residual risk at the appropriate level? A vendor may owe contractual duties, but the deploying organization remains accountable for its choice of use, integration, and supervision, subject to applicable law and allocation of responsibility. The OECD principle assigns accountability according to roles and context and calls for traceability through the lifecycle. [3]

7 Intervention

Who can question an output, correct a record, override a result, pause actions, revoke credentials, roll back a change, or retire the system? Can affected people seek explanation and meaningful review where warranted? Define the trigger, decision maker, time limit, and fallback process before

launch. A stop button that is inaccessible during an incident is not an effective control. OECD explicitly addresses override, repair, or safe decommissioning when harmful behavior appears. [4]

4 From questions to operating controls

Inventory and classification

Maintain an inventory of purchased, built, embedded, and employee-used AI, including features introduced by software updates. Record the business process, provider and sub-providers where known, model or service, owner, data categories, affected people, action and knowledge permissions, integrations, location, version, contract, and review date. Start with consequential uses and expand. The inventory should cover pilots and “shadow AI”; an unregistered use is not thereby permitted.

Classify **impact** and **authority** separately. Impact concerns what can happen when the system is used as designed or misused; authority concerns what it is allowed and practically able to do. Add uncertainty, exposure, reversibility, and legal status as distinct considerations. *A low-authority tool can still have high impact if it shapes a consequential decision.*

For a first pass, impact may be *low, significant, high, or critical*; action authority follows the five functions in Section 2 (inform, recommend, decide, act, automate); and knowledge permission may be *limited, sensitive, or broad*. Prohibited uses or inadequate evidence stop approval regardless of tier. The examples in Section 7 show how to combine these dimensions without pretending they yield a precise risk score.

Classification should describe concrete scenarios. “*Low impact*” might cover a reversible draft of an internal note with no confidential material. “*Significant*” might involve a customer-facing response that could create obligations. “*High*” might involve a decision about employment, credit, benefits, care, or physical operations. “*Critical*” might involve irreversible or large-scale harm. A seemingly

low-impact feature can become high-impact when scaled, integrated, or treated by staff as authoritative. *Reclassify it when the setting changes.*

Human oversight that works

Specify which outputs require review, what the reviewer must inspect, which source evidence is available, how much time is allotted, and how disagreement is recorded. Test reviewers with realistic erroneous outputs. Track override rates and missed errors without treating either number alone as proof of success. Give reviewers authority to challenge targets, deadlines, or a manager's expectation to "trust the AI." For high stakes decisions, ensure the review is independent enough to catch the system's characteristic failures.

Oversight depends on literacy. Staff must know what the tool is authorized to do, which errors are plausible, what evidence to inspect, what information must not be entered, how to challenge an output, and how to report an incident. Train with examples drawn from the actual workflow; a generic awareness course does not show whether a reviewer can catch a wrong recommendation under normal workload. Include workers affected by deployment in testing and feedback, especially where a system changes tasks, pace, discretion, or the basis for evaluating performance.

Staged authority and lifecycle

Start in a sandbox with synthetic or appropriately controlled data. Evaluate the system against representative cases, edge cases, and adversarial or unusual cases. Run a bounded pilot with a fallback process, limited users and permissions, named incident owner, and defined success and stop criteria. Expand only after evidence supports the specific additional authority.

Log the approval and reassess when the purpose, population, data, model, vendor, tools, or stakes materially change.

Monitoring should include quality in the actual use, harmful errors, fairness or accessibility where relevant, security and privacy incidents, user workarounds, over-reliance, complaints, vendor changes, and downstream effects. Sample outputs and decisions at a frequency proportionate to

risk. Preserve the records needed for audit while applying retention and privacy limits. An incident review should ask whether the use case, evidence, permissions, human workload, incentives, vendor dependency, or escalation path failed, not just whether a prompt needs editing.

Before deployment, choose metrics and triggers suited to the use: an increase in missed defects, a shift in input characteristics, a rise in unsupported claims, a spike in overrides or complaints, a model or connector update, or a change in who is affected. Specify who reviews a trigger and whether it calls for investigation, reduced authority, or immediate suspension. A universal percentage threshold would be false precision; establish thresholds from baseline performance and the consequence of each error. Run regression cases after material changes and periodically sample difficult cases even when aggregate averages appear stable.

Roles sized to the organization

An executive sponsor accepts the organization's strategic and residual risk. A use-case owner defines purpose and daily performance. Technical and data stewards manage integrations, access, logging, and retention. An independent reviewer or appropriate control function challenges high-impact uses. Procurement and legal specialists assess vendor terms and applicable obligations. In a small organization, one person can hold multiple roles, but approval and challenge should remain distinguishable and external expertise should be brought in for consequential uses. Record a named person and deputy for each critical action.

5 The vendor risk you inherit

Most organizations will use AI through a vendor product rather than train a model. The vendor may supply the interface, but the service can depend on a foundation model provider, cloud host, retrieval service, third-party data, subcontractor, plugin, and local integration. Each layer can change behavior, availability, security, rights, or evidentiary quality. “You inherit some portion of your vendor's AI risk” means that a failure in this chain can reach your customers, workers, decisions, data, and operations even when the component is outside your direct control.

It does not mean every supplier duty legally transfers to you or that you can inspect every proprietary model. Your share of responsibility depends on your role, jurisdiction, contract, configuration, and actual use. NIST expressly addresses third-party software and data, contingencies for high-risk suppliers, and regular monitoring. [2, 6]

The local-use problem. A vendor can provide general assurances about safety or accuracy without testing your workflow, terminology, populations, or tolerance for error. A model used to draft internal brainstorming notes poses a different risk from the same model used to propose a benefits denial. Your organization chooses the purpose, prompt and retrieval sources, thresholds, human review, permissions, affected people, and downstream action. Vendor assurances cannot answer those local questions.

The visibility problem. A supplier may withhold model weights, training records, test details, or subcontractor identities for commercial or security reasons. That opacity does not excuse a high-impact deployment. *Ask for the evidence you need to govern outcomes:* intended and excluded uses, evaluation methods and populations, known limitations, incident history or reporting practice, data handling, model and configuration identifiers, change notices, and ways to test locally. If adequate evidence cannot be obtained: ***narrow the use, add independent controls, choose another provider, or decline deployment.***

The data and inference problem. Determine whether prompts, uploaded files, retrieval content, feedback, logs, embeddings, and outputs are retained; used to train or improve services; exposed to subcontractors; transferred across borders; or available to administrators. Ask about deletion and backups, tenant isolation, breach notification, and generated or inferred sensitive information. Contractual promises and product settings must agree. A statement that “customer data is not used to train the model” does not resolve retention, human review, telemetry, or sub-processor access.

The change problem. A cloud product can change its underlying model, system instructions, safety filters, retrieval behavior, feature defaults, connectors, or terms without your team changing its own code. Define what counts as a material change; require appropriate notice and release information; test important workflows after changes; and preserve the ability to freeze, roll back, disable, or

switch service when feasible. A contractual right to notice matters only if someone receives it and knows which uses depend on the product.

The integration problem. Tool connections convert a bad answer into an executed mistake. Review the precise scopes of APIs and service accounts, transaction caps, separation of duties, approvals, logging, and emergency credential revocation. Treat retrieved documents and external content as untrusted input, including instructions embedded in them. Test for prompt injection, data exfiltration, incorrect tool selection, duplicated actions, and failure to stop on ambiguous cases. Limits should be enforced by systems outside the model where practical.

For agentic workflows, enforce authorization at the tool or API boundary. Use allowlisted operations, scoped credentials, parameter validation, transaction ceilings, sandbox environments, and human approval for consequential actions. Separate retrieved content from trusted instructions, and test whether poisoned documents or messages can redirect the agent. A model's instruction to "ignore malicious text" is useful guidance but cannot substitute for external access controls. The depth of these tests should reflect the connected system and the possible harm. [9]

The continuity problem. A vendor can suffer an outage, security incident, acquisition, price change, product retirement, or loss of a key sub-provider. Know which business processes depend on it. Establish export rights and usable formats, fallback manual procedures, response-time expectations, and a plan for customer and employee communication. Critical dependencies warrant rehearsed contingency processes; a promise that the supplier has disaster recovery is not your recovery plan.

The responsibility problem. Contracts should allocate duties for confidentiality, security, data protection, intellectual property claims, incident notice, cooperation, audit or assurance, subcontractors, and termination. Insurance, indemnity, and liability caps require legal review, especially when a small contractual cap is dwarfed by plausible downstream harm. A clause cannot restore a denied benefit, leaked trade secret, or unsafe action; operational controls remain necessary. Verify who is the provider and who is the deployer under applicable law rather than assuming the vendor handles all compliance.

A proportionate vendor review

For any AI supplier, record the product's AI features and purpose; the data and integrations it can reach; the intended outputs and possible actions; the known upstream dependencies; and the system owner. For significant uses, request product documentation, test evidence, security and privacy terms, sub-processor information, change and incident processes, and exit options.

Conduct local testing with realistic cases before approval. For high-impact or action-taking uses, add independent validation, contractual rights proportionate to the exposure, limited permissions, human review, supplier contingency, and periodic reassessment. If a vendor refuses an answer, record the resulting uncertainty and what decision it changes. Do not turn a questionnaire into a checkbox ritual.

For example, a hospital buys an AI assistant that drafts patient messages. The supplier controls the model and hosting; the hospital controls which records enter the prompt, whether the draft is sent automatically, the clinician review process, and whether a patient can correct an error. The hospital cannot outsource its clinical workflow judgment simply by purchasing a tool. The supplier must also meet its own obligations. This example illustrates allocation of control; specific legal duties depend on jurisdiction and use.

Questions to put to the supplier and to ourselves

Purpose and evidence. What exactly is the feature intended to do and what uses does the supplier exclude? Which populations, languages, and operating conditions were evaluated? What failure modes appeared, and what tests can the customer repeat with local cases? Is the supplier's reported accuracy a measure of the complete workflow or only one model task? What assumptions in the published evaluation differ from ours?

Data and rights. Where do inputs, outputs, embeddings, logs, and backups go? How long are each retained, and who can inspect them? Are customer materials used for training, fine-tuning, abuse monitoring, or service improvement? Which subcontractors can access them? What rights

does the supplier claim in inputs and outputs, and what evidence supports use of licensed third-party material? How are deletion, correction, and data export carried out in practice?

Security and operations. Which identities and API scopes does the product use? Can the customer set tenant controls, transaction limits, approvals, and audit logs? What happens when retrieved material attempts to redirect the model? How is an incident reported, investigated, and communicated? What is the agreed fallback if the service is unavailable or an upstream model is withdrawn?

Change and recourse. Can the supplier switch models or add an AI feature by default? Which changes trigger notice and retesting? Is an older version available long enough to investigate a regression? Can the customer inspect enough records to explain or contest an important outcome? What are the routes for dispute resolution, termination, and export? These answers should be matched against the actual contract, configuration, and observed product behavior.

Supplier diligence is not limited to the initial purchase. A procurement team can approve one product while an administrator later turns on a new feature with broader access. Assign an owner to read release notices, compare them to the approved use, and trigger reassessment. Record deviations and unresolved questions. For a critical workflow, verify the fallback by actually operating it. When evidence is missing, leadership should see that uncertainty in the approval record and decide explicitly whether to limit the deployment.

A shared responsibility example

Suppose a company connects a vendor assistant to its customer relationship system. The vendor asserts that its model does not train on customer prompts. That assertion answers one question. The company still needs to know whether logs are retained, whether a sub-processor can see them, and whether the connector reads every account or only the account relevant to the task. The company can restrict the connector's scope and require approval before an outward message. It can test the assistant against ambiguous accounts and malicious text in a retrieved note. The

supplier can document its infrastructure, model changes, and incident process. Both sets of controls are necessary; neither party can complete the other's work by contract language alone.

The decision becomes more consequential if the assistant is allowed to issue refunds or change contract terms. At that point, set enforceable amount limits, customer identity checks, exception routing, duplicate transaction protection, and a reconciliation process outside the model.

If the provider cannot support the required audit or rollback, a draft-only use may still be viable.

This is how governance changes the permitted use without treating the product as wholly safe or wholly unsafe.

6 Applying the framework to real decisions

A low-consequence drafting assistant

A communications team uses AI to propose internal meeting summaries. Intent is to reduce time spent drafting, while the meeting owner remains responsible for the final record. Context includes the agenda, participant notes, and decisions actually made. The assistant may read the approved notes and draft text but may not publish it, access unrelated personnel material, or retain it outside approved terms. Consequences include a misleading action item or disclosure of a private comment. Evidence can come from sampling drafts against source notes and measuring corrections. The meeting owner checks the final summary; users can correct it; publication can be paused. This can be a proportionate first use if data handling is acceptable. Calling the output “just a summary” would conceal its practical influence if colleagues later treat it as the authoritative record.

An employment recommendation

An HR team proposes AI ranking of candidates for interviews. Intent must state the job-relevant decision and why a ranking is needed. Context includes the role, required qualifications, application routes, accommodation needs, and the possibility that historical hiring data encode undesirable

preferences. The system's recommendation can influence who receives an opportunity even if a recruiter formally decides. Consequences include unfair exclusion, opacity, and legal exposure. Approval calls for evidence of job relevance, evaluation across relevant groups and edge cases, documented reviewer criteria, a route to correct applicant records, and legal review. The organization should prohibit automatic rejection unless a separate, sufficiently supported decision authorizes it. A vendor's generic accuracy claim cannot establish that the ranking is appropriate for this employer's applicant pool.

An operational agent

An operations team proposes an agent that monitors orders and buys replacement parts. Its intent is to avoid stock outs. Context includes inventory records, lead times, supplier reliability, contract limits, and emergency exceptions. The agent's ability to read demand forecasts is a knowledge permission; its ability to issue purchase orders is action authority. Consequences include duplicate or unnecessary spending, supplier disruption, and physical downtime if the wrong part arrives. A pilot might permit recommendations with human approval. Later evidence may justify execution under a fixed catalog, approved suppliers, a per-order cap, an aggregate daily cap, duplicate detection, and independent reconciliation. The stop procedure should revoke procurement credentials and revert to a manual queue. No model instruction should be relied upon as the sole spending limit.

A quality inspection system

A metal stamping shop proposes vision inspection to flag surface defects. The intended benefit is earlier detection and less rework; the baseline is the current inspection process, scrap, escapes, and operator time. The model may initially flag parts for a trained inspector, with no authority to release or reject production. Context includes part variants, lighting, camera placement, tool wear, finish, and the cost of false rejects and missed defects. Tests should include rare but consequential defects and changes across shifts, not just a convenient sample of good parts. Operators and quality staff should help define the cases the system misses and how alerts affect workload.

If the system later controls a reject gate or feeds a disposition in the QMS, its authority and consequences change. Quality records, traceability, calibration, change control, nonconformance handling, and customer requirements already governed by the shop's quality system remain in force. AI-specific evaluation should join those processes. *If AI touches a PLC, robot, machine stop, or safety function, treat the connection as an operational technology and physical safety matter:* maintain independent interlocks and appropriate engineering approval, test failure states, and preserve a manual recovery path. The AI governance framework guides the decision; it does not certify a quality or safety system.

These cases demonstrate why authority and impact need separate judgments. The drafting assistant might be low impact until it becomes the official source for a consequential decision. The HR model might have no direct execution permission while exerting strong influence. The purchasing agent can be more autonomous within narrow, reversible limits if controls and evidence warrant it. The inspection system crosses a consequential boundary when it moves from flagging parts for a person to controlling a reject gate or QMS disposition. ***The actual decision is about a situated system, not an abstract model.***

The decision record

For each material use, keep a short approval record that another responsible person can understand months later. It should state the owner and purpose; the people and processes affected; the system and vendor dependencies; action and knowledge permissions; foreseeable harms and benefits; applicable legal or policy constraints; test results and limitations; reviewer duties; monitoring and incident thresholds; fallback and stop authority; residual risks; approving decision maker; and date for reassessment. Link to test evidence, contract terms, and system configuration rather than copying them into an unwieldy memo. Record a decision to decline a use as carefully as an approval when the reason may matter later.

Approval is conditional. Name changes that void the decision: a new use, a different population, broader data access, a model replacement, new tools, a material incident, or evidence of harm.

A review date alone is insufficient when an important change occurs earlier. The owner should be able to say which version is operating and how the deployed permissions compare with those approved. If that cannot be established for a consequential system, pause the action pathway until it can.

7 A decision aid for a first review

The following patterns set *illustrative minimum starting controls*, not automatic approvals. Impact establishes a floor: a critical consequence needs leadership review and independent controls even if AI only informs. For combinations not shown, begin with the next more demanding pattern.

Sensitive or broad knowledge permissions call for a separate data and privacy review, restricted access and retention, and a check on what the system may infer or disclose, regardless of its action authority. Escalate for broad exposure, uncertainty, irreversibility, or legal restrictions. A prohibited use remains prohibited; strong controls do not make it permissible.

Low impact plus inform. *For an internal draft with limited data and an easy correction:* name the owner and allowed tool, apply data rules, require a responsible person to check before publication, and sample errors. Keep a way to report and correct mistakes.

Significant impact plus recommend. *For customer-facing or operational advice:* add local test cases, source or evidence checks, trained reviewers, defined escalation, version records, vendor review, and monitoring of complaints and overrides. Do not assume the reviewer can detect errors without testing that claim.

High impact plus recommend or decide. *For employment, benefits, care, quality disposition, or a similarly consequential use:* conduct a documented use-specific assessment and applicable legal review, test affected populations and rare failure cases, ensure meaningful independent review or a justified alternative, preserve contest-ability and traceability, restrict knowledge permissions, and set stop criteria. Decide-level authority requires explicit leadership approval and stronger evidence than recommendation alone.

Act or automate with external or physical effect. Begin with bounded pilots. Enforce scopes, limits, validation, and approval gates outside the model; log each action; reconcile outcomes; rehearse rollback and credential revocation; and reassess after changes. Critical or irreversible effects may require independent safety controls and may be unsuitable for delegation at all.

Seven question worksheet. Use one page per proposed use. Short answers are sufficient if linked evidence is available. Record uncertainty explicitly and identify who will resolve it.

Intent. What decision or work improves? What baseline and benefit target will show value? What is the cost of delay or unmanaged use?

Context. What data, people, operating conditions, and missing information matter? Which assumptions could fail?

Authority. Which of inform, recommend, decide, act, and automate apply? What can the system access, infer, retain, disclose, and change? Where are limits enforced?

Consequence. Who can be harmed, how severely, at what scale, and how readily can an error be detected and reversed?

Evidence. What local tests, source checks, vendor evidence, and reviewer trials support the proposed authority? What remains unknown?

Accountability. Who owns the use, approves residual risk, reviews outputs, monitors performance, and answers to affected people?

Intervention. What signals trigger escalation or suspension? Who can stop the system, revoke access, correct outcomes, and restore the prior process?

Decision. Approve within stated limits, pilot only, require more evidence, or decline. Record the approver, conditions, date, and triggers for reassessment.

8 Relationship to established frameworks

NIST AI RMF 1.0 structures risk work as **Govern**, **Map**, **Measure**, and **Manage**; its governance function covers roles and third-party risk across the lifecycle. It is voluntary. NIST states that the 1.0 core is being revised and has already supplemented it with a Generative AI Profile; use the current materials for implementation. [1, 2, 9] ISO/IEC 42001 specifies requirements for an organization-wide AI management system and continual improvement. [7] The OECD AI Principles articulate human-centered values, transparency, robustness, and accountability, including traceability and intervention. [3, 4]

The EU AI Act assigns obligations according to the system and the provider or deployer role. The 2026 Digital Omnibus extended the application of certain high-risk requirements to December 2, 2027 for Annex III uses and August 2, 2028 for Annex I product-embedded uses. Other provisions have different dates and duties, so a current use-specific legal check remains essential. [8] In the United States, existing employment law, including Title VII and the Americans with Disabilities Act, applies to covered employers' selection practices even when AI assists them. Privacy, consumer protection, safety, state, and sector requirements need their own use-specific review. [10]

The seven questions prepare a leader to use those frameworks well. **Intent** and **Context** support mapping; **Consequence** and **Evidence** support measurement and treatment; **Authority**, **Accountability**, and **Intervention** strengthen governance and management. This is a conceptual relationship; a separate mapping of specific controls and legal duties—necessary for compliance.

9 A practical starting sequence

First 30 days. Appoint an accountable sponsor and operational lead. Publish interim rules on approved tools, prohibited data and actions, and escalation. Inventory existing uses, including AI hidden in vendor features. Identify high-impact and action-taking uses for immediate review. Give staff examples of errors, sensitive data exposure, and the limits of generated answers.

Days 31–60. Apply the seven questions to priority uses. Document permissions, evidence, reviewers, supplier dependencies, and a stop path. Pilot one or two bounded uses that solve a real problem. Establish vendor review and incident intake. Suspend a use when the organization cannot identify its owner, purpose, permissions, or reasonable means of intervention.

By 90 days. Review pilot evidence against the original benefit and harm criteria. Formalize approval and reapproval triggers, monitoring, training, and record retention. Assign supplier change notices to owners. Report material unresolved risks to leadership and decide whether to narrow, improve, expand, or retire each use. A small organization can do this with short records and a monthly review; complexity should follow exposure, not fashion.

Governance capability can progress from uncontrolled use to awareness and inventory, governed experimentation, operational governance, and adaptive governance. This describes the ability to govern AI. **Gnotek Ai**'s separate adoption and Operational Intelligence models describe different dimensions of organizational capability; a company can be advanced in one and weak in another.

The test here is whether the institution can make and revisit a defensible decision about each consequential use.

Conclusion

AI governance begins when an organization recognizes that access to knowledge, influence over judgment, and authority to act are forms of delegation. A useful system may be powerful without being trustworthy for every purpose. Leaders should state the purpose, bound the permissions, examine the consequences, demand evidence in context, assign accountable people, and keep a real path to intervene. ***The more consequential and difficult to reverse the outcome, the stronger those conditions must be.*** That discipline allows organizations to benefit from AI while remaining answerable for what they choose to deploy.

References and source notes

- [1] NIST, *Artificial Intelligence Risk Management Framework AI RMF 1.0*, 2023, Executive Summary and Core. <https://airc.nist.gov/airmf-resources/airmf/0-ai-rmf-1-0/> and <https://airc.nist.gov/airmf-resources/airmf/5-sec-core/>
- [2] NIST, *AI RMF Core*, Govern 6 and Manage 3. <https://airc.nist.gov/airmf-resources/airmf/5-sec-core/> and <https://airc.nist.gov/airmf-resources/playbook/manage/>
- [3] OECD, *AI Principles*, Accountability, updated principles. <https://oecd.ai/en/ai-principles> and <https://oecd.ai/en/dashboards/ai-principles/P9>
- [4] OECD, *AI Principles*, Robustness, security and safety. <https://oecd.ai/en/dashboards/ai-principles/P8>
- [5] Gnotek Ai, *The Gnotek Ai AI Constitution*, *The Gnotek Ai Manifesto*, and *Judgment Before Automation*, 2026. https://cyberalley.com/GnotekAi/_documents/gnotek_ai_ai_constitution.pdf ; https://cyberalley.com/GnotekAi/_documents/the_gnotek_ai_manifest.pdf ; https://cyberalley.com/GnotekAi/_documents/judgement_before_automation.pdf
- [6] NIST, *AI RMF Playbook*, Govern and Manage guidance on third-party resources. <https://airc.nist.gov/airmf-resources/playbook/govern/> and <https://airc.nist.gov/airmf-resources/playbook/manage/>
- [7] ISO, *ISO IEC 42001 2023 AI management systems*. <https://www.iso.org/standard/42001>
- [8] European Union, *Regulation EU 2026 1744 Digital Omnibus on AI*, Official Journal, July 24, 2026; European Commission, *AI Omnibus enters into force*. <https://eur->

lex.europa.eu/eli/reg/2026/1744/oj/eng and <https://digital-strategy.ec.europa.eu/en/news/ai-omnibus-enters-force>

[9] NIST, *Artificial Intelligence Risk Management Framework Generative Artificial Intelligence Profile*, NIST AI 600-1, 2024; NIST, *AI Risk Management Framework* status page. <https://doi.org/10.6028/NIST.AI.600-1> and <https://www.nist.gov/itl/ai-risk-management-framework>

[10] U.S. Equal Employment Opportunity Commission, *Title VII of the Civil Rights Act of 1964 and Titles I and V of the Americans with Disabilities Act of 1990*, statutory texts. <https://www.eeoc.gov/statutes/title-vii-civil-rights-act-1964>; <https://www.eeoc.gov/statutes/titles-i-and-v-americans-disabilities-act-1990-ada>

This is an educational governance framework. It is not legal advice, a conformity assessment, or a substitute for a use-specific risk and regulatory review.