

Responsible AI, Security, and Governance

Gerald C. Kane, Ph.D.

Executive MBA course, Terry College of Business, University of Georgia
Fall 2026

This public edition omits the in-class activity specifications, which are designed to be experienced live in session.

© 2026 Gerald C. Kane. Licensed under CC BY-NC 4.0. You may share and adapt it for noncommercial use with credit to the author; you may not present it as your own. Full terms on the last page.

Module role: Earlier modules establish the strategic landscape, executive AI literacy, use-case logic, and the technology-stack choices that follow. Module 4 translates those choices into accountable deployment. Module 3 owns the architecture: models, retrieval, context systems, evaluation, observability, identity, permissions, vendor choices, and technical dependencies. Module 4 asks how much authority that architecture should receive and what evidence is required before it affects real people, sensitive data, decisions, or operations. The boundary matters. This is not a responsible-AI philosophy course or a technical security class. The module is built for executives who need to know what questions to ask, what artifacts to require, and what decisions they cannot delegate away.

Module Learning Objective

By the end of the module, you will be able to build a practical AI governance plan for one AI use case in your organization. Every student will complete the Governance Card's Core fields: use case and owner, risk tier, authority statement, most important control or assurance gap, and executive owner of residual risk. At the Challenge depth — called the Full Card in this module — the same artifact extends to the complete control plan, assurance evidence, oversight and recourse, audit-log requirements, and incident-response design. For the Transform-path capstone, the tangible output is the complete one-page AI Governance Card.

If You Want to Know More

A useful way to prime the reading is to hold three questions about a real AI use case in your organization or industry in mind as you go: what is the AI system allowed to read, draft, recommend, decide, or do; what could it do, reveal, recommend, or decide that would create a board-level problem; and who would be accountable if that happened? The formal preparation sits in "Bring This to Class" at the end of Part 1.

The Pre-Class Read below carries the argument; the materials in the table carry the underlying frameworks and regulatory baseline. These references are optional — read what fits your time.

Source	Why it matters for this module
European Commission AI Act implementation FAQ and Regulation (EU) 2026/1744	Gives the current 2025–2028 regulatory baseline. AI literacy and the original prohibited-practice provisions already apply, and GPAI and governance provisions began applying on August 2, 2025. Regulation 2026/1744 was published on July 24, 2026 and entered into force on July 27. It moved Annex III high-risk obligations to December 2, 2027 and product-embedded (Annex I) high-risk obligations to August 2, 2028, while leaving Article 50 transparency on the original August 2, 2026 date. ¹
NIST AI Risk Management Framework 1.0 and AI RMF Playbook	Practical operating structure: Govern, Map, Measure, Manage. Use it as enterprise risk-management language, not as a slide of definitions. ^{2 3}
NIST Generative AI Profile (AI 600-1)	Extends the AI RMF for generative-AI risks and connects governance to evaluation, monitoring, content provenance, and lifecycle controls. ⁴

Source	Why it matters for this module
OWASP Top 10 for LLM Applications 2025	Concrete security vocabulary for LLM applications: prompt injection, sensitive information disclosure, improper output handling, excessive agency, vector and embedding weaknesses, misinformation, unbounded consumption. ⁵
OWASP Top 10 for Agentic Applications 2026	Optional reading where agentic AI is central to the use case. Moves beyond general LLM risk toward autonomous planning, tool use, and multi-step action. ⁶

The goal is not to become a compliance specialist; it is to bring enough vocabulary to participate in a serious governance conversation.

Part 1: Pre-Class Read

BOTTOM LINE. AI governance is not a brake on speed — it is the management system that lets an organization say yes to AI with eyes open, evidence in hand, and a way back when something fails. The discipline is to make authority and evidence move together: the more an AI system is allowed to know, recommend, decide, or do, the stronger the proof must be that it behaves acceptably for that level of authority. Executives can delegate operational governance work, but they cannot delegate the questions of risk classification, authority, evidence, and accountability. This module gives you the artifacts, the routine, and the questions to govern AI without slowing the work that matters.

TERMS. Selected governance and technical terms not already defined in Modules 1-3 appear in **bold** at their first substantive use in Part 1; each is defined in the Executive Glossary immediately before Part 2. Use the glossary as a lookup aid, not assigned vocabulary to memorize.

Opening Vignette: Aflac, Revisited — The Evidence Behind the Authority

Module 1 opened with Aflac as a study in strategic discipline. We return to the same case because it is also a clean public example of the question this module is built around: authority and evidence moving together.

Aflac built one of the earliest large-scale AI claims platforms in its industry, deployed it in 2022 and 2023, and has since extended automation across accident, hospital, cancer, dental, and vision lines. Shelia Anderson, then chief information officer, described the design choice on MIT Sloan Management Review's *Me, Myself, and AI* podcast: "We had about 46 of our claims that were fully automated, or what we call straight-through processing — primarily wellness claims... 49% of our manual claims result in payouts less than, for example, \$200" (*Shelia Anderson, CIO, Aflac, MIT Sloan Management Review podcast, 2023*).¹⁷ The company concentrated automation in repetitive, lower-value claims and preserved human judgment for higher-stakes, complex cases where empathy, edge-case discretion, and accountability matter.

The Aflac case shows what authority and evidence look like when they move together. Wellness claims received a high level of system authority because the population of cases was well-understood, the evidence base on those cases was deep, and the **residual risk** on a wrong call was bounded. Complex claims kept their human judges because the evidence required for autonomous action did not yet exist. The governance question is not "do we trust the AI?" It is "for which cases does our evidence justify which authority?"

Governance Is Not the Team That Says No

Perhaps the most important shift to recognize as governance enters its mature phase in 2026 is that the conventional framing has it backwards. AI governance is widely treated, inside organizations and in the press, as a brake — the function that delays pilots, demands paperwork, and discovers reasons not to launch. We tend to think of governance as a constraint applied from outside the work, after the fact, by people whose job is to find problems. The evidence from organizations that ship AI well suggests otherwise. AI governance, when it works, is not the team that says no. It is the management system that lets the organization say yes — knowing what it is approving, on what proof, and how it will recover if the approval turns out to be wrong.

The reframing matters because the alternative produces two equally weak answers. The first is trust by enthusiasm: the system seems useful, so the organization grants it more authority than the evidence supports. The second is paralysis by abstraction: governance becomes so general that no one can tell what would make a use case safe enough to try. Both come from treating governance as a value statement rather than as a working management system. The discipline that resolves both is to attach authority to evidence, tier by tier, and to hold the line on what authority a system has actually earned.

The Core Frame: Authority + Evidence

KEY LESSON. Authority and evidence together produce the trust decision. Authority asks what the AI is allowed to know, recommend, decide, or do. Evidence asks what proof the organization has that the system behaves acceptably for that level of authority. The trust decision is whether, for this use case, with these users, under these risks, the evidence justifies the authority.

Authority question	Evidence question
What data can the AI access?	What data lineage, access-control, privacy, and retention evidence exists?
What systems or tools can it use?	What permission tests and tool-call logs prove least privilege ?
Can it draft, recommend, decide, or act?	What evaluations, red-team results , approval gates, and monitoring thresholds exist?
Who can override it?	What human review, escalation, and recourse evidence exists?
What happens when it fails?	What incident playbook , audit trail, notification path, and remediation record exist?

The practical position is disciplined. Low-authority systems need lighter controls. High-authority systems need stronger evidence. Risk appetite, authority, and evidence move together.

Reflection prompt as you read. In your organization, who can answer these five questions today: what AI systems are in use, which touch sensitive data, which affect customers or employees, which can take action, and which have been tested since launch?

In our experience and in the public record on enterprise AI governance, few organizations can answer all five cleanly today. That gap is the opening for the first governance artifact.

From Answer Risk to Action Risk

The critical distinction is between answer risk and action risk. A wrong answer can mislead a user. The same answer connected to tools can send the wrong email, update the wrong record, approve the wrong exception, expose the wrong document, or trigger the wrong operational step. Model output and retrieved content should therefore be treated as untrusted until checked; tool use should be treated as privileged access.

Agentic AI makes this distinction operational. An agent can plan, execute multi-step tasks, use tools, and continue toward a goal. NIST now treats agentic vulnerabilities as a distinct family, and OWASP provides a companion catalogue of agent-specific failure modes.^{11 6} The governable object is no longer only what the model says. It is what the complete system can do with real credentials, permissions, data, and tools.

The governing principle is **least agency**: grant the smallest authority needed for the workflow, with permissions narrow enough that one failure has a bounded blast radius. A prompt can influence behavior; the harness must enforce the boundary.

We use the same five-level authority ladder introduced in Module 2. Module 4 adds the control burden required at each level.

Level	Course term	Permitted authority	Human control	Minimum controls (Module 4 contribution)
1	Assistant	Drafts, summarizes, or analyzes only.	Human reviews all output before use.	Source boundaries, user disclosure where needed, basic logging, data-use rules
2	Copilot	Recommends actions inside a human-led workflow.	Human steers, approves, and remains the decision-maker.	Human review, output validation, source citation, data-loss controls, approval gates
3	Workflow Agent	Completes bounded tasks in approved systems.	Human approves key gates, exceptions, and escalation points.	Least privilege, action allowlists , tool-call logs, rollback path , escalation rules
4	Supervised Digital Worker	Acts across systems within defined thresholds.	Human sets boundaries, monitors exceptions, and owns outcomes.	Pre-deployment evaluations, red teaming, monitoring, kill switch , incident playbook
5	Multi-Agent Process	Coordinates multiple agents across an end-to-end process segment.	Human leaders approve the operating model, risk tier, monitoring, and incident response.	Senior risk acceptance where material, independent assurance, strong auditability, mature incident response, operating model owner

This is an authority model, not a maturity ladder where higher is always better. Some workflows should remain at Assistant or Copilot because the cost, risk, or ambiguity of autonomy is not worth it. The executive question is which authority the evidence justifies, not how much autonomy the technology can demonstrate.

The April 2026 PocketOS incident shows the consequence of getting that decision wrong. A coding agent assigned a staging task found an over-permissioned infrastructure token in an unrelated file

and deleted the production database and backups in seconds, despite a written rule requiring approval for destructive actions.²⁴ The safeguard existed only inside the agent's instructions. No credential scope, permission boundary, or deterministic block stood in the execution path. A rule the agent holds is a suggestion; a control the harness enforces is a boundary.

The managerial test is simple: if an AI system can act, can your organization name the identity it uses, the tools it can reach, the actions it may take, the approval gates, the evidence produced, and the recovery path?

Keep capability separate from permission. A system may be technically able to send a message, change a record, or approve an exception while remaining authorized only to draft or recommend. Write the permitted verbs down, deny everything else by default, and require new evidence before expanding them. That discipline prevents a successful demo from becoming an accidental production mandate.

A Maturity Model for AI Governance

Use the maturity check to diagnose your organization's starting point. The point is not to label the organization but to identify the next governance constraint.

Stage	What it looks like	Leadership focus
Ad hoc	Employees and teams use AI in scattered ways, often without a common inventory or policy interpretation	Discover actual use and clarify obvious boundaries
Visible	The organization can name major AI systems, tools, pilots, and vendor uses	Build the inventory and assign owners
Controlled	Risk tiering, approved tools, data rules, and review gates exist for meaningful use cases	Match controls to authority and risk
Assured	High-risk or high-authority systems have test evidence, red-team results, monitoring, oversight, and incident paths	Require evidence before production and expansion
Adaptive	Governance learns from incidents, user behavior, model changes, vendor changes, and frontline feedback	Keep governance current as use and capability change

The demanding shift is from Controlled to Assured, because it requires evidence the organization may not yet produce.

The Governance Routine

The core routine below is the full playbook. Each step produces a named artifact that you can point at when a regulator, auditor, board member, or affected employee asks how the AI is governed.

Step	Executive question	Artifact
Inventory	What AI system is this, where is it used, and who owns it?	AI inventory record
Classify	What risk tier applies and why?	Risk-tier rationale
Define authority	What can the AI read, draft, recommend, decide, or do?	Authority statement
Assign controls	What controls match the risk and authority?	Control plan

Step	Executive question	Artifact
Require evidence	What proof is needed before production or scale?	Assurance packet
Design oversight	Where does human judgment matter?	Oversight pattern
Log and monitor	Can we reconstruct and learn from behavior?	Audit and monitoring plan
Prepare response	What happens when something fails?	Incident playbook

The routine is deliberately simple. The complexity comes from applying it to real use cases. A public-facing chatbot, an internal knowledge assistant, a claims triage tool, a legal drafting copilot, a hiring screen, and an autonomous customer-success agent should not face the same process. They should be governable through the same questions.

Why Visibility Comes First

The first governance problem is visibility. AI is often adopted faster than it is cataloged. It appears inside enterprise SaaS tools, vendor platforms, internal prototypes, analytics workflows, call-center systems, code assistants, HR tools, marketing automation, and employee use of public models.

An AI inventory is not paperwork for its own sake. It is the map that lets the organization make different decisions for different systems. A low-risk summarization assistant should not face the same process as a model that affects credit, employment, healthcare, safety, legal rights, pricing, fraud decisions, or regulated customer communications. The Centers for Disease Control and Prevention's annual contribution to the HHS AI Use Case Inventory — 103 use cases as of December 31, 2025, layered on top of the Public Health Data Strategy launched in 2023²⁶ — is one public example of the discipline at federal-agency scale. The format is not the point; the visibility is. Without an inventory, everything is either invisible or overgoverned.

Meaningful ownership has at least four parts. First, a business owner who owns the outcome and residual risk. Second, a technical owner who owns the system design, data flow, permissions, monitoring, and rollback path. Third, a risk, legal, compliance, or security owner who interprets obligations and control requirements. Fourth, an operating owner who understands how the workflow behaves in daily use. For larger organizations, this maps cleanly to a **three-lines model** — business or product team in the first line, risk and compliance in the second, internal audit or independent assurance in the third — but the model only works if the first line actually owns the risk. Governance cannot be something the business tosses over the wall to legal after the system is already built.

This is also where shadow AI deserves a careful frame. Shadow use is partly a policy problem, but it is more usefully treated as a discovery signal. Employees often use public tools because official tools are too slow, weak, inaccessible, or unclear. A mature governance system does not only punish shadow use. It asks what unmet need the shadow use is revealing and whether the organization can provide a safer path.

Tiering Risk Before Scaling Authority

Once the system is visible, the next question is how much governance it needs. Risk tiering should be simple enough to use early and strong enough to route serious use cases into deeper review. The five-tier scheme below is a starting point; the Field Reference contains the full screen.

Tier	Typical pattern	Governance response
Low	Internal productivity, read-only assistance, public or low-sensitivity data, no material decisions	Basic policy, approved tools, user training, logging where appropriate
Medium	Business workflow support, confidential data, customer or employee impact, recommendations with human review	Inventory, owner, documented use limits, testing, monitoring, human review, access controls
High	Regulated decisions, legal rights, safety, employment, credit, healthcare, pricing, fraud, material customer impact, significant agent authority	Formal review, assurance evidence, security testing, human oversight, auditability, incident playbook, legal and compliance involvement
Prohibited	Uses barred by law, policy, or organizational values	Do not deploy
Pause pending review	Facts are unclear, vendor obligations are unresolved, or authority and data exposure are not understood	Stop expansion until classification and controls are clear

At the high-authority end, governance artifacts become more substantial and visible. Lockheed Martin's ethics committee and high-assurance AI subsidiary are one public example; the Field Reference carries that case and the related public-sector cross-reference.¹⁶ The executive principle is enough for the required reading: high authority requires proportional evidence. High authority with thin evidence creates **responsibility debt**.

Responsibility debt accumulates when teams answer the tiering questions vaguely and move anyway. The common signs are easy to recognize: no one can say exactly what the AI system is allowed to do; the business owner is enthusiastic but has not accepted residual risk; the model was tested on examples that do not resemble the real workflow; human review exists in theory but the reviewer lacks time or context; logs are too thin to reconstruct what happened; vendor claims are treated as assurance evidence; incidents would be handled through improvisation.

How to Build One Governance System Across Many Regulators

Treat regulation as a design constraint for enterprise governance rather than as a legal lecture. The executive question is whether your organization can build one governance system that satisfies legal, risk, security, procurement, audit, and business needs without making every useful experiment move at the speed of the heaviest case.

Framework	What it answers	How to use it
EU AI Act	What legal obligations may apply, especially for prohibited, high-risk, limited-risk, and general-purpose AI systems	Classify risk and identify documentation, transparency, oversight, monitoring, and incident obligations
NIST AI RMF 1.0	How to structure AI risk management across Govern, Map, Measure, and Manage	Adopt as the enterprise operating model for risk ownership and continuous improvement

Framework	What it answers	How to use it
NIST Generative AI Profile (AI 600-1)	What risks are unique to or amplified by generative AI	Update testing, monitoring, content provenance, incident, and third-party review practices
OWASP LLM and Agentic Security	What can go wrong in LLM-powered applications and agents	Threat modeling, red teaming, least privilege, output validation, agent-permission design
ISO/IEC 42001:2023	What a certifiable AI Management System looks like	Adopt as the integration point with ISO/IEC 27001 (information security) and ISO/IEC 27701 (privacy); procurement counterparties increasingly ask whether you are certified ¹²

2026 sits on the active edge of EU AI Act implementation. 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. It preserved the Article 50 transparency date and moved the high-risk dates. The Field Reference's EU AI Act timing box carries the full schedule. The executive summary: AI literacy and the original prohibited-practice provisions already apply; Article 50 transparency obligations began applying on their original August 2, 2026 date; Annex III high-risk obligations begin December 2, 2027; and product-embedded Annex I high-risk obligations begin August 2, 2028.^{1 19} Recheck the governing law before making legal commitments.

Penalties scale to the violation. The practical implication is more important than memorizing the ceiling: someone must own regulatory monitoring and keep the governance playbook current.

U.S. corporate law adds a board-level reason to require these artifacts. The *Caremark* oversight line asks whether leaders built and used a reasonable system for monitoring mission-critical compliance risk. For AI, an inventory, named owners, escalation routes, and evidence of review help show that oversight was real rather than assumed.¹⁸ The case history sits in the Field Reference.

What Can Actually Go Wrong

The risk matrix below is deliberately short and executive. Its purpose is to help you name the risk, recognize it in the workflow, and ask for the right control. The Field Reference holds the full thirteen-row register.

By 2026, many of the AI failures appearing in incident registries — including the AI Incident Database and the OECD AI Incidents Monitor — are no longer best described as generic "hallucination." They are context failures or harness failures. A context failure means the system saw the wrong information, too much information, contradictory information, stale information, unauthorized information, or untrusted information that changed its behavior. A harness failure means the system lacked the surrounding controls that would have caught, contained, explained, or improved the failure: evals, tool limits, permission boundaries, output validation, logging, monitoring, rollback, or human review.

Risk family	Executive control question
Reliability and hallucination	What test set, source check, groundedness metric, and human review standard are required?
Bias and unfair treatment	What populations, outcomes, and proxy variables have been tested, and who reviews fairness evidence?

Risk family	Executive control question
Privacy and data leakage	What data classes are allowed, masked, logged, retained, or prohibited?
Prompt injection and instruction conflict	How are direct and indirect prompt injection tested and contained?
Context failure	What context-quality tests, source ownership, freshness rules, isolation patterns, and cache-invalidation rules exist?
Harness failure	What harness components make the workflow reliable beyond a good demo?
Excessive agency	Which permissions can be removed, separated, scoped, or put behind approval?
Vendor and supply-chain risk	What assurance packet, notice rights, audit support, and incident duties are in the contract?

Two cases sharpen the bias and privacy rows. *Mobley v. Workday* shows that using a hiring vendor does not dissolve accountability for discriminatory outcomes; the vendor may also face direct liability.¹⁰ *NYT v. OpenAI* shows that a legal hold can override ordinary deletion practices and that exposure can differ by service tier and contract.¹³ The executive takeaways are durable: test affected outcomes, preserve recourse, and do not promise deletion your vendor cannot deliver. The procedural and docket detail sits in the Field Reference.

Assurance: Evidence Before Scale

Assurance is the bridge between ambition and trust. It answers a practical question: what proof do we have that this AI system works well enough for this use case, with these users, under these risks, at this level of authority? You do not need to design every evaluation. You do need to require evidence before production, before expansion, and after meaningful changes to the model, data, workflow, vendor, or permissions.

Optomed's Aurora AEYE handheld fundus camera offers a clean illustration of what assurance evidence looks like at the point of release. The device received FDA clearance on April 30, 2024 as the first FDA-cleared handheld AI fundus camera for autonomous diabetic retinopathy detection in primary care; published diagnostic sensitivity ranges from 92% to 93% and specificity from 89% to 94%.¹⁵ David Frasc, Managing Director, Optomed USA, framed the deployment as bridging the diabetic retinopathy gap with on-the-spot detection in roughly sixty seconds. The relevant point for governance is that Aurora AEYE could not deploy with autonomous diagnostic authority without published, peer-reviewed performance evidence on the specific population, in the specific setting, against the specific decision the device makes. Authority granted, evidence produced. Where similar evidence does not exist, the authority should not be granted.

A short academic anchor sharpens the point. Liao and colleagues, writing on clinical AI governance, frame the assurance challenge as a tiered review problem: the higher the consequential impact of an AI-influenced decision, the more independent and evidence-rich the review must be.⁷ The Field Reference contains the full eleven-row Assurance Evidence Table; the Pre-Class Read carries the principle.

For high-risk systems, the **release gate** should be short but real. It should force your organization to name who is accepting what risk, on what evidence, with what conditions. The practical test is

whether the release gate would change a decision. If it never causes a team to narrow scope, add controls, delay rollout, or reject a use case, it is probably theater.

What Meaningful Oversight Looks Like

Different audiences need different things from a transparent AI system. End users need to know what the AI did, what sources it used, and when human review is required. Human reviewers need the basis for the output, the confidence limits, the source links, and the option to approve or override. Affected customers and employees need notice where required, a meaningful explanation, and a correction path. Auditors and regulators need documentation, logs, model or vendor version, data sources, controls, testing, monitoring, and incident history. Executive sponsors need clear residual risk, unresolved issues, and decision options.

Human oversight is often named and rarely designed. A checkbox that says "human in the loop" does not mean much if the human lacks time, context, authority, training, or incentives to intervene. Meaningful oversight has four properties. First, the human can understand the basis of the AI output. Second, the human can change the outcome. Third, the human can pause or escalate the workflow. Fourth, the organization does not punish the human for slowing the system down when risk requires it.

The Field Reference contains the full set of oversight patterns and the audit-log minimums for consequential systems. The principle for the Pre-Class Read is simpler: oversight is a design problem, not a label.

When Something Fails

If there is no incident path, your organization has not finished designing the system. It has only launched it. AI incidents do not all look like cyber breaches. They may include hallucinated customer communications, data leakage, discriminatory outcomes, unsafe recommendations, unauthorized tool use, prompt-injection success, runaway cost, model drift, harmful content, or a vendor model change that alters behavior. The Field Reference contains the seven-step Incident Response Playbook.

For relevant Annex III high-risk systems, the cited Article 73 windows become applicable on December 2, 2027: 15 days as the standard maximum, 2 days for widespread infringement or a qualifying severe incident, and 10 days when a death is involved.¹⁹ They are not a generally applicable current reporting obligation. Other laws, contracts, and sector-specific duties may apply now. Recheck the governing law and reporting path before legal reliance.

Bring This to Class

Pick one AI use case in your organization or your division. Every student should bring the Governance Card's Foundation/Core fields:

1. **Use case and owner.** Name the workflow, business purpose, and accountable owner.
2. **Risk tier.** Classify it under the five-tier scheme and name the evidence supporting that classification.

3. **Authority statement.** State what the AI may read, draft, recommend, decide, or do, and name the human approval point.
4. **Most important gap.** Identify the single control or piece of assurance evidence most missing today.
5. **Residual-risk owner.** Name the executive who accepts the remaining risk — not only the technical owner.

For the Challenge depth — the Full Card in this module — extend the same artifact with the complete control plan, assurance evidence, oversight and recourse, audit-log requirements, and incident-response design. This is deeper analysis of one card, not a separate graded pathway or a new assignment. Bring the gaps; the tabletop will run faster if you know which question your use case cannot yet answer.

A Final Word Before Class

We have walked through the artifacts of governance: an inventory, a risk tier, an authority statement, controls, assurance evidence, oversight patterns, audit logs, and an incident playbook. The artifacts matter, but they are the surface. Underneath sits a quieter question — whether your organization is willing to hold the line on what authority an AI system has earned and what evidence it has actually produced. That holding is the governance work, and it is what turns governance from a brake on speed into the organization's capacity to act on AI without losing what matters. The Governance Card you build for your capstone use case is the first place to practice it.

Executive Glossary

Use this alphabetized glossary as a lookup aid for Part 1. It includes only terms not already defined in Modules 1–3; definitions reflect how the terms are used here.

Term	Plain-language definition
Action allowlist	A predefined list of actions an AI system may take; everything else is blocked.
AI inventory	A maintained record of AI systems, owners, data, authority, risk, controls, and evidence.
Incident playbook	A prepared sequence for detecting, containing, documenting, escalating, and learning from an AI failure.
Least agency	Granting an AI system only the autonomy needed for the task.
Least privilege	Granting an identity only the data and system permissions needed for the task.
Red teaming / red-team result	Adversarial testing intended to expose weaknesses before they cause real harm, and the evidence it produces.
Release gate	A required decision point before production, expansion, or higher authority.
Residual risk	The risk remaining after controls are applied and accepted by an accountable owner.
Responsibility debt	Unresolved risk and accountability that accumulate when AI scales without adequate owners, controls, or evidence.
Rollback / kill switch	Controls that reverse a change or quickly stop a system from continuing.

Term	Plain-language definition
Three-lines model	Governance that separates operating ownership, risk and compliance oversight, and independent assurance.

Part 2: Field Reference

The Field Reference is the leave-behind. Use it during the course to scaffold board memos, vendor conversations, release-gate reviews, and capstone decisions; keep it after the course as an executive reference. Each component is self-contained — flip directly to a table, prompt list, or worksheet without reading surrounding prose.

Decision Tables

Minimum viable AI inventory

Field	What to capture	Why executives need it
Use case	Business process, users, decision or workflow supported	Clarifies purpose and scope
Owner	Business owner, technical owner, risk or compliance reviewer, vendor owner	Prevents accountability from dissolving across functions
System type	Assistant, Copilot, Workflow Agent, Supervised Digital Worker, Multi-Agent Process, embedded vendor AI	Shows authority and operating context
Data used	Public, internal, confidential, restricted, personal, regulated, proprietary, retrieved knowledge sources	Flags privacy, security, IP, and data-quality risk
Model or source	Internal model, vendor model, open-source model, API provider, version where known	Supports vendor diligence and model-change management
Actions allowed	Read, draft, recommend, update records, send messages, transact, trigger downstream systems	Determines approval gates and control strength
Risk tier	Low, medium, high, prohibited, or pause pending review	Routes the use case to the right governance path
Evidence	Testing, evaluations, red-team results, monitoring, oversight, logs, incident plan	Shows whether the system is ready to scale

Risk tiering screen

Question	If yes, increase scrutiny
Does the system affect employment, credit, healthcare, education, safety, legal rights, access to services, pricing, fraud, or regulated customer treatment?	Likely high risk; legal and compliance review required
Does it use personal, regulated, confidential, or competitively sensitive data?	Privacy, security, retention, and access controls required
Can it take action through tools, APIs, email, CRM, finance, HR, code, or operational systems?	Agentic controls required
Can external or retrieved content influence its instructions?	Prompt-injection and RAG-security testing required
Will customers, employees, regulators, auditors, or the board expect an explanation?	Documentation, auditability, and recourse required
Would a wrong answer create material financial, reputational, safety, or legal harm?	Assurance evidence and incident response required before scale

Executive AI Risk Register (full)

Risk	What it looks like in practice	Executive control question
Reliability and hallucination	The system gives a confident answer that is wrong, unsupported, stale, or misapplied	What test set, source check, groundedness metric, and human review standard are required?
Bias and unfair treatment	Outputs disadvantage a group, amplify historical bias, or create inconsistent outcomes	What populations, outcomes, and proxy variables have been tested, and who reviews fairness evidence?
Privacy and data leakage	Sensitive data enters an unapproved tool, appears in output, or is retained by a vendor	What data classes are allowed, masked, logged, retained, or prohibited?
Prompt injection and instruction conflict	User, document, webpage, email, ticket, or retrieved content causes the system to violate boundaries	How are direct and indirect prompt injection tested and contained?
Context failure	Context poisoning, distraction, confusion, clash, stale knowledge, stale cache, or missing source material makes the output look plausible but wrong	What context-quality tests, source ownership, freshness rules, isolation patterns, and cache-invalidation rules exist?
Harness failure	The prompt and context may be strong, but the system has weak tools, permissions, evals, validation, monitoring, or rollback	What harness components make the workflow reliable beyond a good demo?
Improper output handling	Model output is passed into code, workflow systems, emails, or records without validation	Where is output treated as untrusted data before execution?
Excessive agency	The AI has more tool access or autonomy than the use case requires	Which permissions can be removed, separated, scoped, or put behind approval?
RAG and embedding weakness	Retrieved material is wrong, stale, poisoned, over-permissioned, or missing needed context	How are retrieval sources governed, permissioned, refreshed, and tested?
Drift and performance decay	Performance changes as data, users, business rules, or model versions change	What monitoring thresholds trigger review, retraining, rollback, or retirement?
Vendor and supply-chain risk	The organization relies on vendor claims, model changes, subcontractors, or unclear data terms	What assurance packet, notice rights, audit support, and incident duties are in the contract?
Accountability and recourse	Affected people cannot understand, challenge, appeal, or correct an AI-influenced decision	Who explains the decision, who can change it, and what evidence is preserved?
Unbounded consumption	Usage creates runaway cost, denial of service, or resource exhaustion	What rate limits, cost limits, timeouts, throttles, and anomaly monitoring are in place?

Frameworks map

Framework	What it answers	Where it lands
EU AI Act	Legal obligations for prohibited, high-risk, limited-risk, and general-purpose AI systems	Classification, documentation, transparency, oversight, monitoring, incident
EU AI Act GPAI Code of Practice (final July 2025; in force August 2, 2025)	Operational expectations for general-purpose AI providers across transparency, copyright, and safety/security chapters	Vendor diligence and procurement attestations from GPAI providers ²¹
NIST AI RMF 1.0 (Govern, Map, Measure, Manage)	Enterprise AI risk-management operating structure	Risk-ownership backbone

Framework	What it answers	Where it lands
NIST Generative AI Profile (AI 600-1)	Risks unique to or amplified by generative AI	Testing, monitoring, content provenance, incident, third-party review
NIST AI 100-2 (March 2025 update)	Adversarial-AI taxonomy including agentic vulnerabilities (indirect prompt injection chained to tool use, agent memory poisoning, supply-chain attacks on agent tools)	Threat modeling and red teaming ¹¹
OWASP Top 10 for LLM Applications 2025	Application-layer LLM risks	Threat modeling, output validation, agent permissions ⁵
OWASP Top 10 for Agentic Applications 2026	Agentic-AI-specific failure modes	Agent design, tool-call governance, blast-radius bounding ⁶
ISO/IEC 42001:2023	Certifiable AI Management System standard	Procurement attestation; integration point with ISO/IEC 27001 (information security) and ISO/IEC 27701 (privacy) ¹²
ISO/IEC 23894:2023	AI risk management guidance	Companion to 42001
ISO/IEC 22989, 23053, 5338, 24027	Terminology, ML framework, lifecycle, bias	Cross-reference family
MITRE ATLAS	Adversarial-AI threat landscape (counterpart to MITRE ATT&CK)	Red-team finding mapping ²⁰
C2PA / Content Credentials	Content provenance and authentication	Article 50 transparency obligations; native signing on Samsung Galaxy S25 and Google Pixel 10 as of January 2026 ²²

EU AI Act timing box

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 27.¹ Treat the dates below as the current baseline and recheck the governing law before legal reliance.

Date	Baseline application	Executive implication
February 2, 2025	General provisions, AI literacy (Article 4), and prohibited AI practices (Article 5) began to apply	As amended, providers and deployers must take measures to support AI literacy for staff and others operating or using AI systems on their behalf; organizations also need a clear list of prohibited uses
August 2, 2025	Governance provisions and general-purpose AI model obligations began to apply; GPAI Code of Practice in force	Model providers face documentation, copyright-policy, training-summary, systemic-risk, incident-reporting, and cybersecurity obligations where applicable
August 2, 2026	Article 50 transparency rules (chatbots, deepfakes, AI-generated content on matters of public interest); Commission and AI Office enforcement powers over GPAI providers	Transparency work is due on the original date — the Omnibus did not move it. Machine-readable marking of AI-generated content has a grace period to December 2, 2026 for systems already on the market
December 2, 2026	New Article 5 prohibitions on AI-generated non-consensual intimate imagery and CSAM	Any organization shipping generative or image features into the EU needs a documented scope position before this date
December 2, 2027	Annex III high-risk rules, Article 27 Fundamental Rights Impact Assessments for designated deployers, and Article 73 serious incident reporting for relevant high-risk systems	Companies operating in or selling into the EU need risk classification, documentation, human oversight, monitoring, and incident processes; the cited Article 73 windows are not a generally applicable current duty

Date	Baseline application	Executive implication
August 2, 2028	Rules for high-risk AI embedded in regulated products (Annex I) — deferred from August 2, 2027	Product companies need longer-horizon conformity and lifecycle planning

Article 73 reporting clock

For relevant Annex III high-risk systems, these Article 73 windows become applicable on December 2, 2027. Product-embedded Annex I high-risk systems follow the August 2, 2028 timing shown above. These are not generally applicable current reporting obligations; other legal, contractual, or sector-specific duties may apply now. Recheck the governing law before legal reliance.¹⁹

Incident class	Reporting window
Standard serious incident	15 days
Widespread infringement or qualifying severe incident	2 days
Incident involving death	10 days

Agent authority ladder (full)

See Pre-Class Read for the full table; the ladder columns are Course Term, Permitted Authority, Human Control, and Minimum Controls.

Controls for agentic AI

Control	Executive question
Scope tools narrowly	What functions does the agent actually need for this workflow?
Use least privilege and least agency	Are read, draft, update, send, approve, and transact permissions separated?
Isolate risky context	Are untrusted emails, webpages, documents, tickets, and tool outputs separated from instruction-bearing context?
Govern context freshness	Who owns source updates, stale-file removal, knowledge-file replacement, and cache invalidation?
Apply action allowlists	What can the agent do, for whom, under what conditions, and within what thresholds?
Require human approval for high-impact actions	Which actions are irreversible, regulated, customer-facing, or financially material?
Log tool use	Can we reconstruct prompts, retrieved sources, model version, tool calls, approvals, overrides, and outputs?
Test for direct and indirect prompt injection	Can malicious content in emails, documents, webpages, tickets, images, or RAG sources change behavior?
Validate outputs before execution	Are model outputs treated as data to be checked, not commands to be trusted?
Build limits and kill switches	Can cost, rate, scope, and tool access be contained quickly?

Salesforce's published Trust Layer is one procurement reference for translating these questions into architecture: mask sensitive inputs, validate outputs before downstream action, and log the trip.⁸

Two vendor-commissioned Cloud Security Alliance surveys published in April 2026 reported AI-agent incidents, unknown agents, and permissions exceeding intended scope across respondent organizations.²⁵ Treat the percentages as directional, not as neutral prevalence estimates. The useful diligence question is whether your own inventory, permission tests, and incident logs can reveal the same failure modes.

Assurance evidence table

Evidence type	What good looks like	Executive question
Purpose and limits	Documented intended use, non-use cases, user groups, and known limitations	What should this system never be used for?
Impact assessment	A concise review of affected stakeholders, harms, benefits, legal duties, fairness concerns, data exposure, and operating context	Who could be affected, and how would we know if harm is occurring?
Test set and evaluations	Representative examples, edge cases, failure cases, and success criteria linked to business risk	What must be true before we scale?
Groundedness and source quality	Outputs are supported by approved sources and do not invent facts or citations	Can users see where the answer came from?
Context-quality review	Context is relevant, sufficient, isolated, economical, current, permission-aware, and traceable	Did the system receive the smallest reliable context packet for this task?
Harness review	System prompt, retrieval, tools, memory, output structure, evals, orchestration, and monitoring are documented and tested	Is the system reliable because of structure, not because the pilot went well?
Bias and fairness	Assessment against relevant groups, outcomes, and proxy variables	Who could be harmed or treated inconsistently?
Security testing	Prompt-injection, RAG-poisoning, permission, output-handling, and tool-use tests	Can malicious or messy inputs make the system violate boundaries?
Red teaming	Adversarial testing by people not invested in the build, with findings mapped to MITRE ATLAS technique IDs where applicable	Who tried to break it, and what changed afterward?
Monitoring	Accuracy, groundedness, drift, escalation, override rate, latency, cost, safety events, and incidents	How will we know it is getting worse?
Vendor assurance	Security posture, data-use commitments, model-change notice, incident duties, audit support, ISO/IEC 42001 status if certified	What are we relying on the vendor to prove?

Release gate for high-risk or high-authority AI

Owner	Confirmation required
Business owner	Intended use, expected value, affected stakeholders, residual risk
Technical owner	Architecture, data sources, model or vendor source, permissions, logging, monitoring, rollback
Security owner	Threat model, prompt-injection tests, output handling, access controls, agent permissions
Legal or compliance owner	Regulatory classification, documentation, disclosure, retention, incident obligations
Human oversight owner	Who reviews, approves, overrides, escalates, and explains decisions
Governance body	Approve, approve with controls, pause, or reject

Transparency audiences

Audience	What they need
End user	What the AI did, what sources or data it used where relevant, and when human review is required
Human reviewer	Basis for the output, confidence limits, exceptions, source links, and approval or override options
Affected customer or employee	Notice where required, meaningful explanation, correction path, appeal or recourse path
Auditor or regulator	Documentation, logs, model or vendor version, data sources, controls, testing, monitoring, and incident history
Executive sponsor	Clear residual risk, unresolved issues, and decision options

Human oversight patterns

Pattern	Best use	Failure to avoid
Human as approver	High-impact actions such as regulated communications, exceptions, pricing changes, or customer-facing commitments	Rubber-stamp approval with no time, context, or authority
Human as reviewer	Medium-risk outputs where sampling and escalation are sufficient	Reviewing only successes and missing systematic failures
Human as exception handler	High-volume workflows where AI handles routine cases and humans handle uncertainty	Letting the AI define its own uncertainty threshold without monitoring
Human as auditor	Periodic assurance of logs, outcomes, complaints, overrides, and drift	Audit trails that omit prompts, sources, model version, or tool actions
Human as appeal path	Decisions with material effect on customers, employees, or citizens	No recourse, no explanation, and no owner for correction

Audit log minimums

For consequential systems, logs should make the decision or action reconstructable. At minimum, preserve: user identity or role; timestamp; model, vendor, or version; prompt and relevant context; retrieved document identifiers; context assembly path including retrieval filters, memory used, and cache status where available; data sources accessed; tool calls and system actions; output; human approvals or overrides; policy or guardrail events; incident links and remediation notes.

Logs are not only for forensic review. They are a learning system. They show where users override the AI, where the AI struggles, where controls are creating friction, and where the risk profile is changing.

AI incident response playbook

Step	Question	Action
Detect	What signal suggests failure or misuse?	Capture alerts, complaints, abnormal tool use, hallucination reports, safety events, cost spikes, or external reports
Triage	What is the severity and scope?	Assess affected users, data exposure, financial impact, regulatory relevance, customer impact, and ongoing harm

Step	Question	Action
Contain	How do we stop additional harm?	Disable tool, reduce permissions, pause model, block source, revert workflow, or switch to human process
Preserve evidence	What will we need to reconstruct events?	Freeze logs, prompts, outputs, retrieved sources, model version, approvals, and system configuration
Notify	Who needs to know?	Route to business owner, security, legal, privacy, compliance, vendor, regulator, board, or affected stakeholders as applicable
Remediate	What must change before restart?	Update prompts, context sources, cache rules, tool access, evaluations, monitoring, user training, vendor terms, or governance controls
Learn	How does this change the playbook?	Update risk tiering, add tests, refresh controls, and report trends to the governance body

Six board questions for meaningful AI deployment

1. What AI system is being deployed, who owns it, and what authority will it have?
2. What data, systems, people, and decisions can it affect?
3. What risk tier applies, and what regulations, standards, or policies shape the controls?
4. What evidence shows it is accurate enough, secure enough, fair enough, monitored, and recoverable?
5. Where is human oversight meaningful, and where is it only symbolic?
6. What is the incident response path if the system fails, leaks data, discriminates, hallucinates, or takes the wrong action?

Cautionary Tales

The cases below illustrate the risk register in practice. Each entry pairs a public, dated incident with the section of the module it sharpens.

Air Canada chatbot — Moffatt v. Air Canada (BC Civil Resolution Tribunal, February 2024).

Discussed in detail in Module 1; cross-referenced here for the accountability-and-recourse implication. The tribunal rejected the airline's argument that the chatbot was a separate legal entity, holding that statements an AI agent makes to customers may bind the organization regardless of how the system was procured. *Lands in*: accountability and recourse; vendor and supply-chain risk.

Mata v. Avianca and the lawyer-fabricated-citations ecosystem. Discussed in detail in Module 2; cross-referenced here for the assurance-evidence implication. The 2023 *Mata v. Avianca* sanction is the canonical anchor; the pattern has expanded substantially through 2025 (*Mavy, Noland, Johnson, Heppner*). The governance lesson is that delegated AI work without a human critique step exposes the human's professional accountability — at the rate of roughly several cases per day in U.S. courts by year-end 2025. *Lands in*: reliability and hallucination; assurance evidence.

Arup Hong Kong deepfake (January 2024). A finance worker authorized fifteen wire transfers totaling roughly HKD 200 million (~\$25 million USD) after a video conference call in which the company's CFO and other senior executives appeared to authorize the transfer. Every "person" on the video call except the victim was an AI-generated deepfake. *Lands in*: prompt injection and

instruction conflict (broadly construed to include synthetic-media social engineering); incident response.

Samsung ChatGPT data leak (April 2023). Discussed in detail in Module 2; cross-referenced here for the privacy-and-data-leakage implication. The case is the canonical demonstration that prompt-level confidentiality instructions are not the same as data-architecture controls — relevant to the data-class-and-retention questions on the Risk Tiering Screen. *Lands in:* privacy and data leakage; shadow AI.

iTutor Group EEOC settlement (September 2023). An AI-driven hiring screening tool automatically rejected female applicants over 55 and male applicants over 60. The EEOC's first AI-discrimination settlement: \$365,000 plus injunctive relief. *Lands in:* bias and unfair treatment; accountability and recourse.

Sports Illustrated AI authors (November 2023). Articles attributed to AI-generated "authors" with fabricated bios and AI-generated headshots; brand credibility damaged when the practice surfaced. *Lands in:* improper output handling; transparency.

McDonald's IBM AI drive-thru (June 2024). Discussed in detail in Module 1; cross-referenced here for the release-gate implication. The system handled simple cases well but failed in long-tail edge cases that dominated actual customer interactions — a release-gate question of whether the eval set actually represented production conditions. *Lands in:* reliability and hallucination; long-tail performance; release gate.

LiteLLM supply chain (March 2026). Threat actors compromised the LiteLLM open-source library, present in approximately 36% of cloud environments and serving 97 million monthly downloads at the time of the incident. Thousands of organizations were affected. *Lands in:* vendor and supply-chain risk; assurance.

Replit production-database deletion (July 2025). During an explicit code freeze, a software-engineering agent deleted a live production database, fabricated roughly 4,000 user records that masked the deletion, and told the operator rollback was unavailable even though a recovery path existed.¹⁴ Replit CEO Amjad Masad acknowledged the incident publicly and announced automatic separation of development and production environments. The failure combined excessive tool access, no dev/prod permission boundary, and no human approval gate for destructive operations. *Lands in:* excessive agency; harness failure; release gate.

PocketOS production deletion (April 25, 2026). An AI coding agent (Cursor on Claude Opus 4.6) working a routine staging task found an over-permissioned API token in an unrelated file and deleted the company's production database and volume-level backups in nine seconds, despite explicit written rules against destructive commands. Discussed in detail in the Pre-Class Read; listed here because it is the cleanest single illustration of the controls column of the agent authority ladder — least privilege, permission boundaries, and deterministic blocks on destructive operations enforced outside the agent's own reasoning. *Lands in:* agent authority; least privilege; release gate.

Cursor "Sam" hallucinated policy (April 2025). A customer-service AI built by the developer-tools company Cursor fabricated a one-device-per-subscription rule while posing as a human ("Sam"). Customer cancellations and a public apology followed; the company committed to labeling

AI responses going forward. *Lands in*: reliability and hallucination; transparency; accountability and recourse.

Wolf River Electric v. Google (2025). Google's AI Overviews falsely stated that Wolf River Electric was a defendant in a state-AG lawsuit; Wolf River reported customer cancellations and filed a defamation suit. *Lands in*: reliability and hallucination; vendor and supply-chain risk.

Bombay Stock Exchange CEO deepfake (January 2026). BSE issued an urgent public warning after deepfake videos circulated promoting fraudulent stock tips attributed to BSE leadership. *Lands in*: prompt injection and instruction conflict (synthetic-media class); transparency.

Singapore deepfake CFO incident (March 2025). A Singapore company authorized a \$499,000 wire transfer after a video call the attackers proactively offered to verify; an evolution of the Arup Hong Kong pattern. *Lands in*: incident response.

FDA "Elsa" AI tool fabricating studies (CNN reporting, July 23, 2025). The U.S. Food and Drug Administration's internal generative-AI tool, Elsa, was reported to have fabricated study citations in regulatory work. The reporting is based primarily on secondary press coverage; we treat the case as illustrative rather than load-bearing. *Lands in*: reliability and hallucination; assurance evidence.

Caremark oversight lineage (Delaware, 1996-2025). *In re Caremark* established the board-oversight frame; *Marchand v. Barnhill* sharpened it for mission-critical risk; *In re McDonald's* extended oversight duties to officers; and the 2025 *Giuliano v. Grenfell-Gardner / Teligent* ruling allowed compliance-oversight claims against directors and officers to proceed.¹⁸ These cases do not create an AI-specific checklist. They explain why boards need a reasonable information-and-reporting system for material AI risk and evidence that leaders use it. *Lands in*: accountability; board oversight; governance routine.

NYT v. OpenAI litigation (S.D.N.Y., 2023-present). Magistrate Judge Wang's preservation order of May 13, 2025 — affirmed by Judge Stein on June 26, 2025 — required retention of ChatGPT conversation logs that otherwise would have been deleted. Enterprise agreements and approved zero-retention API configurations sat outside the order; consumer and standard API usage did not. The going-forward preservation obligation ended in September-October 2025, but previously preserved material remained in scope; a large de-identified log sample was ordered produced, and Judge Stein affirmed that production order on January 5, 2026.¹³ The implication for any deploying enterprise is that it cannot promise deletion its vendor cannot deliver, and that exposure turns on tier and contract rather than vendor brand alone. The associated AI-copyright docket (including *Bartz v. Anthropic's* \$1.5 billion preliminary class settlement on pirated training-data acquisition; *Kadrey v. Meta*; *Thomson Reuters v. Ross Intelligence* on training fair use) is moving fast; vendor contracts now routinely add training-data and indemnification clauses in response. *Lands in*: privacy and data leakage; vendor and supply-chain risk.

Mobley v. Workday (N.D. Cal., 23-cv-00770). Conditionally certified as an ADEA collective on May 16, 2025, with HiredScore extension on July 7, 2025; Workday's own filings reference 1.1 billion rejected applications. The court's acceptance of the EEOC's "agent" theory of vendor liability is now precedent for AI-procurement diligence on hiring tools. The court-approved opt-in deadline tied to the December 2, 2025 notice plan made *Mobley* the largest AI-bias collective ever certified.

¹⁰ *Lands in*: bias and unfair treatment; vendor and supply-chain risk; release gate.

Cobb County Sheriff's Office DEKA Sentry pilot (October 2024). The Cobb County Sheriff's Office in Marietta, Georgia, deployed three nearly six-foot-tall DEKA Sentry robots — equipped with 360-degree cameras, night vision, heat detection, and two-way audio — for a 90-day pilot of autonomous and live-operator-controlled perimeter and dormitory patrols inside the Adult Detention Center. The deployment is one named instance of a wider category of carceral and surveillance AI: speech-recognition systems flagging inmate calls in Alabama, Georgia, New York, and other states; recidivism scoring AI applied across more than 50,000 Virginia convictions; third-party vendors (LeoTech, Securus) that have faced multiple lawsuits over recordings of attorney-client-privileged calls. The Berkeley Technology Law Journal's "AI Wardens" essay names the governance question directly: AI deployed against people who cannot consent, refuse, or appeal raises authority-and-evidence questions distinct from consumer or workplace AI, and the existing framework for privilege, recidivism scoring, transcription accuracy, and Sixth Amendment access to counsel does not yet have a settled answer for AI's role.²³ *Lands in:* accountability and recourse; bias and unfair treatment; privacy and data leakage; transparency.

AI Insurance Landscape

By early 2026, the insurance market has begun to bifurcate around AI exposure. Verisk's January 1, 2026 ISO endorsements **CG 40 47** and **CG 40 48** allow commercial general liability carriers to exclude AI-related claims; AIG, W.R. Berkley, and Great American have introduced AI-specific exclusions in directors-and-officers, errors-and-omissions, and employment-practices liability lines. Affirmative AI products have entered the market in parallel: Munich Re's aiSure, Armilla (Lloyd's / Chaucer), Testudo (launched January 2026), and HSB AI Liability Insurance (launched March 2026). Gartner's April 2026 advisory recommends that general counsel assess AI insurance as a 2026 priority.

The implication for governance is structural. Underwriters now ask for governance documentation — inventory, risk-tiering rationale, assurance evidence, incident playbook — as a coverage prerequisite. The artifacts this module describes are no longer only compliance artifacts; they are the price of insurability.

Sectoral Matrix

For attendees in regulated sectors, the table below names the specific guidance to read alongside the general frameworks.

Sector	Specific guidance	What to read
Banking and financial services	Federal Reserve SR 11-7 (model risk management, 2011) and OCC Bulletin 2011-12; revised April 2026 as SR 26-02 / OCC Bulletin 2026-13, with a forthcoming joint AI/GenAI MRM RFI	Three-pillar model risk management adapted to AI; effective challenge, ongoing monitoring, documentation
Medical devices and clinical AI	FDA Predetermined Change Control Plans for AI-DSF (final guidance December 2024); January 2025 lifecycle-management draft; August 2025 FDA / Health Canada / MHRA five guiding principles	Pre-market change planning so model updates do not require new clearance

Sector	Specific guidance	What to read
Healthcare delivery and payers	HHS / OCR Section 1557 Final Rule (May 2024) on patient-care decision-support tools; CMS WISer Model (January 1, 2026 – December 31, 2031): AI/ML-assisted prior authorization for 17 selected Part B services in traditional Medicare across six states (NJ, OH, OK, TX, AZ, WA), with every non-affirmation reviewed by a licensed clinician	Discrimination prohibitions on AI-assisted clinical decisions; WISer as the federal template for AI-in-utilization-review design — machine screening, human denial authority, and vendor accountability metrics
Defense and national security	DoD AI Strategy memo (January 2026); DoD Directive 3000.09 (autonomous weapons); RAISIP; CDAO oversight; "any lawful use" contract language	Procurement and use-restriction architecture for high-assurance environments
Critical infrastructure / OT	CISA + NSA + international partners "Principles for the Secure Integration of AI in Operational Technology" (December 3, 2025); CI Fortify (2026)	Four-principle framework with AIBOM, AI red-teaming, vendor transparency, human-in-the-loop oversight
Insurance	NAIC AI Model Bulletin (adopted by roughly two dozen states as of spring 2026); NAIC AI Systems Evaluation Tool in a twelve-state pilot through September 2026 (CA, CO, CT, FL, IA, LA, MD, PA, RI, VT, VA, WI), with adoption anticipated at the Fall 2026 National Meeting; NAIC March 2026 Issue Brief opposing federal preemption of state insurance AI oversight	De facto governance baseline for insurance; the Evaluation Tool previews the structure of AI market-conduct examinations — usage inventory, governance risk assessment, high-risk model detail, and data inputs — and works as a self-assessment benchmark now
Federal procurement	GSA proposed "American AI" Schedule clause; OMB M-25-21	Model-provenance statement, risk assessment, government use rights

Vendor Data Terms and Connector Scoping: What the Contract Actually Says

The tier is the control

Enterprise AI risk conversations tend to focus on which vendor was selected. The more consequential variable is which tier the organization bought, because training defaults, retention, and legal protections differ more across tiers within one vendor than they do across vendors at the same tier.

	Consumer (free or personal paid)	Team / Business	Enterprise / API
Trains on inputs by default	Frequently yes; opt-out available	No	No
Retention	Varies; opt-in to training extends it substantially at some providers	Admin-configurable	Configurable; zero-retention available on approval
Processor relationship	None	Data processing agreement	DPA, plus BAA and residency options
Human review	Possible	Limited, abuse-related	Limited, abuse-related; contractually bounded
Suitable for	Public information	Internal business information	Confidential, regulated, client data

The governance implication is that an inventory listing "ChatGPT" or "Copilot" as an approved tool has recorded almost nothing. The inventory item is the tier and the contract, not the brand.

What a "we do not train on your data" commitment does not cover

The commitment is generally accurate and considerably narrower than executives read it to be. Four carve-outs recur across providers:

Carve-out	What it means operationally
Abuse-monitoring retention	Logs typically persist for a period regardless of training settings, accessible to authorized vendor staff
Safety review	Content flagged by safety classifiers can be reviewed by humans whether or not training was declined
Legal retention	Litigation holds and legal obligations override both user deletion and vendor retention policy
Subprocessors	Content traverses cloud infrastructure and contractors; the commitment is only as strong as its flow-down

The executive question is therefore not whether the vendor trains on your data. It is whether your organization retains control of the data, which is a contractual question with a different answer.

NYT v. OpenAI is the instructive case because it demonstrates the third carve-out operating at scale. A court ordered preservation of output logs that would otherwise have been deleted, overriding both user deletion settings and the vendor's own retention practice. Enterprise agreements and approved zero-retention API configurations sat outside the order; consumer and standard API usage did not. The going-forward preservation obligation was subsequently terminated, but material already preserved remained subject to the litigation, and a large de-identified log sample was ordered produced. Two lessons survive the procedural detail: an organization cannot promise deletion its vendor cannot deliver, and the tier determined the exposure.

Connectors: a distinct exposure category

Connecting an AI system to email, file storage, calendars, or CRM is not an extension of chat. It changes three things at once.

Permission inheritance. A connected assistant surfaces anything the authorizing user can technically access. Where file permissions have accumulated loosely over years, the assistant makes that accumulation immediately discoverable. This is a pre-existing access-governance problem that AI converts from latent to active.

Retention outside the chat boundary. Synced connectors typically build an index that is stored separately from conversation history and is therefore not governed by conversation-retention settings. Disconnecting removes access and begins a deletion window. Deleting a conversation does not touch it.

An instruction channel. Any connector that ingests externally authored content (email, web pages, form submissions, tickets) creates a path for content to influence behavior. The first documented zero-click exfiltration of this kind, patched in June 2025, worked because reading an inbox meant reading whatever the inbox contained. Later incidents extended the pattern to CRM lead forms and document repositories. The "Controls for agentic AI" table in this Field Reference carries the control set; the point here is that the exposure begins at authorization, not at agent deployment.

The 2025 compromise of a widely used sales-integration vendor, in which stolen authorization tokens exposed hundreds of downstream customer environments, is the reference case for the supply-chain version of this risk. Over-permissioned integrations concentrate exposure at the integration rather than the model.

Connector diligence questions

Question	Why it matters
Which tier authorized this connection, and under what contract?	Determines whether a processor relationship exists at all
Is the grant read-only, or does it include send, post, or modify?	Separates retrieval risk from action risk
Is authorization per-item, per-session, or persistent?	Persistent blanket grants are the ones that outlive their purpose
Where does connected content live, and under whose retention clock?	Synced indexes commonly sit outside conversation retention
Does this connector ingest externally authored content?	If yes, treat it as an instruction channel and scope accordingly
Who can revoke it, and how quickly?	Revocation speed is the containment control

Shadow AI as a tier problem

Module 4's Pre-Class Read frames shadow use as a discovery signal rather than only a policy violation. The vendor-terms material sharpens that frame. Most shadow AI is not employees choosing an unapproved vendor; it is employees using an approved vendor at the wrong tier, on a personal account, because the sanctioned deployment is slower, weaker, or unavailable to them. The exposure is contractual rather than technological, and it is invisible to tooling that inventories by brand.

This is why prohibition performs poorly as a control. An organization that bans a tool without providing a governed equivalent has converted a visible, contractually bounded risk into an invisible, unbounded one.

BOARD QUESTION. *"For every AI tool in our inventory, can we name the tier, the contract, and who authorized the connectors, or do we only know the brand?"*

Current as of August 2026. Vendor terms in this area changed materially during 2025 and 2026 and remain volatile. Specific retention windows, zero-retention eligibility, and connector permission models should be verified against current vendor documentation before contractual reliance.

High-Assurance AI: A Cross-Reference

At the high-authority end of the agent ladder — fire-control loops, signals-intelligence assurance, autonomous-systems oversight, disaster-response decision support — the governance artifacts become substantial and visible. Ethics committees, separately chartered subsidiaries, public commitments to explainability and robustness, interagency testing taskforces, named public points of accountability: these are not optional dressing on the deployment. They are the evidence that

earns the authority. The arrangement scales up because the consequences of error scale up. The opposite arrangement — high authority with thin evidence — is the responsibility-debt pattern the Pre-Class Read describes.

The cross-reference below names five anchors in the U.S. high-assurance AI ecosystem as of mid-2026. The pattern across the five is what executives in regulated and consequential deployments should ask their own programs to mirror, scaled to the authority being granted.

Anchor	What it is	Governance artifacts
Lockheed Martin / Astris AI	Prime-contractor side. The Lockheed AI Factory and the Lockheed Martin AI Center, externalized in December 2024 as Astris AI under Lockheed Martin Evolve, with Mike Baylor (Chief Digital and AI Officer) and Greg Forrest (Director of AI Foundations) as public faces. ¹⁶	Separately chartered subsidiary; Lockheed Martin AI Ethics Advisory Committee; public commitments to explainable, robust, trustworthy AI/ML aligned with high-assurance MLOps standards
NSA Artificial Intelligence Security Center (AISC)	Government-customer side. Launched September 2023 inside NSA's Cybersecurity Collaboration Center; mission to secure AI adoption across the National Security System and the defense industrial base; identifies foreign-intelligence threats to AI systems; develops best-practice and risk frameworks	Joint guidance with CISA, FBI, and Australia's ASD ACSC on integrating AI in operational technology and on AI data security best practices
DoD / DoW Chief Digital and AI Office (CDAO)	Central authority. Owns the DoD AI Strategy and the Responsible AI Strategy and Implementation Pathway (RAISIP); oversees Project Maven (computer vision for ISR), Replicator (autonomous systems), and Task Force Lima (generative AI). Executive Order 14347 (September 5, 2025) authorized "Department of War" as a secondary, non-statutory title; the legal name remains DoD until Congress acts	DoD Directive 3000.09 (autonomous weapons); DoD AI Ethical Principles (adopted 2020); CDAO oversight of high-volume AI/ML contracts with Palantir, Anduril, Microsoft, Google, AWS, Scale AI, Lockheed Martin, and others
NIST Testing Risks of AI for National Security (TRAINS) Taskforce	Interagency testing. Announced November 20, 2024 within the U.S. AI Safety Institute. Members include CDAO and NSA. The cross-agency testing function is the federal counterpart to the assurance-evidence discipline this module describes	Cross-agency evaluation, red-teaming, and risk-framework development for AI systems with national-security implications
Georgia Army National Guard / Project Theia	Civic-AI / disaster-response side. Centralizes video data and applies AI for situational awareness; Master Sgt. Kenneth Houck publicly attested that Theia reduced post-flight imagery analysis from twelve hours to seconds. Pairs with Project ARIA (March 2025) and the new MOS 49B AI/ML enlisted specialty as the workforce-side counterpart	Mission-specific authority limits; human-in-the-loop framing for disaster-response decisions; published interagency partnerships including DoD/DIU work with Google X's Bellwether for aerial imagery analysis after disasters

The pattern that connects all five is the structural principle this module keeps returning to. As authority scales, the evidence required to earn it scales too, and the public-facing artifacts that document the evidence scale alongside. The arrangement is what makes the authority defensible — to a regulator, to an inspector general, to a board, or to the public — and what gives the organization a way back when something fails. Executives whose programs operate at lower-authority tiers can still treat this as the model: scale the artifacts to the authority, not to the headlines.

AI Bill of Materials and Model Cards

An **AI Bill of Materials (AIBOM)** is the AI counterpart to a software bill of materials (SBOM). Standards are converging around CycloneDX 1.6 ML-BOM and SPDX 3.0 AI extensions, with

TAIBOM (Trusted AI BOM) and the JHU APL AIRS (AI Risk Scanning) framework as emerging variants. AIBOMs track training data lineage, model weights, dependencies, and provenance — increasingly demanded by procurement counterparties and federal customers.

Model cards and **system cards** are the published-by-the-vendor counterparts: short structured documents describing intended use, training data, performance limits, fairness assessments, and known failure modes. Mitchell et al. (2018) is the canonical reference; Anthropic's Claude system cards and the Hugging Face Model Card Guidebook are the most-cited operational examples. The Stanford Foundation Model Transparency Index, which dropped from an average score of 58 (2024) to 40 (2025), measures how much of this information frontier-model providers actually publish.⁹ The directional implication is that vendor-side evidence is becoming scarcer just as enterprise demand for it is growing.

Chief AI Officer and Algorithm Review Board Structural Design

By 2026, roughly 26% of enterprises (IBM Institute for Business Value, Q1 2025) and approximately 48% of FTSE 100 firms have a designated Chief AI Officer or equivalent. Truist's November 2025 appointment of a Chief AI & Data Officer is one recent named example. Hub-and-spoke structures show 36% higher reported ROI than fully decentralized AI organizations. Three structural choices are common: a standalone AI Center of Excellence with separate governance authority; AI capability embedded in existing CoEs and shared services; or a hybrid in which a small central AI team sets standards and existing functions own use cases. The hybrid is most consistent with the four-owner taxonomy this module uses.

An **Algorithm Review Board** (ARB) — sometimes called an AI Ethics Committee — sits above the four owners. The body's purpose is not to relitigate every use case but to set risk-tier criteria, approve high-tier deployments, and review incident learnings. Lockheed Martin's AI Ethics Advisory Committee is one public example.

Stanford AI Index 2026 Governance Numbers

A short set of board-discussion data points from the Stanford AI Index 2026:⁹

- Approximately 88% of organizations report AI use in at least one business function.
- 62% of organizations cite security and risk as the top blocker to scaling agentic AI.
- The share of organizations *without* a responsible-AI policy fell from 24% (2024) to 11% (2025).
- Only 31% of Americans report trusting government to regulate AI.
- The Foundation Model Transparency Index average score fell from 58 (2024) to 40 (2025), indicating that frontier-model providers are publishing less about training data, compute, and post-deployment impact even as deployment accelerates.

Internal Audit and AI

The Institute of Internal Auditors released an updated AI Auditing Framework in 2025 and has positioned internal audit as the third-line backbone for AI assurance. The practical implication for

deploying enterprises is that internal audit should be reading the same governance artifacts the second line produces — risk-tier rationale, assurance packet, audit logs, incident reports — and reporting independently to the audit committee on whether the artifacts hold up under scrutiny.

SEC Disclosure and AI Washing

Recent disclosure trends sharpen the executive question. In 2025, 380 of the S&P 500 added or expanded AI material-risk disclosures in annual filings; 36% of the Fortune 100 disclosed AI as a standalone 10-K risk factor (up from 14% the prior year); 38% of the S&P 500 disclosed reputational concerns related to AI. The SEC brought its first AI-washing actions in March 2024 against investment advisers Delphia and Global Predictions. Under Chair Atkins (2026), the SEC has withdrawn earlier proposed AI rules and signaled principles-based reliance on existing materiality framework (Reg S-K Items 101, 103, 106, 303), with the December 2025 Investor Advisory Committee recommending separate disclosures for internal-operations effects vs consumer-product effects. The June 3, 2026 deadline for the cyber disclosure rule's amended provisions is the live filing event of the term.

EU GPAI Code of Practice Chapter Map

The General-Purpose AI Code of Practice, finalized July 10, 2025 and in force August 2, 2025, organizes around three chapters.²¹

- **Transparency.** What providers must publish about model capabilities, training-data summaries, and intended-use documentation.
- **Copyright.** Training-data acquisition, opt-out compliance, and downstream-use commitments.
- **Safety and security.** Applies primarily to GPAI providers above the systemic-risk threshold (~5–15 providers); requires Safety and Security Frameworks, model reports, serious-incident reporting (Measure 12), independent external assessors (Measure 11), and non-retaliation protections (Measure 13).

For most deploying enterprises, the Code matters as a procurement reference: ask whether the vendor's GPAI provider has signed the Code and what the most recent model report says.

Workbook Prompt: Build a Governance Card

Use one canonical Governance Card in two passes. The Foundation/Core pass is the common starting point for every student. The Challenge/Full Card extends that same artifact; it is deeper analysis, not a separate graded pathway or a new assignment.

Depth	Field	Draft
Foundation/Core	Use case, business objective, and accountable owner	
Foundation/Core	Risk tier and rationale	
Foundation/Core	Authority statement: what the AI may read, draft, recommend, decide, or do; where human approval is required	
Foundation/Core	Most important missing control or assurance evidence	

Depth	Field	Draft
Foundation/Core	Executive owner of residual risk	
Challenge/Full Card	Data, systems, context sources, freshness ownership, and permissions	
Challenge/Full Card	Complete control plan tied to the top risks, including context and harness failure where relevant	
Challenge/Full Card	Assurance evidence required before production and before scale	
Challenge/Full Card	Oversight, override, escalation, explanation, and recourse	
Challenge/Full Card	Audit-log requirements	
Challenge/Full Card	Incident-response design, owner, containment path, and first actions	
Challenge/Full Card	Decision: approve, approve with controls, pause, or reject	

For a thirty-minute meeting opener, complete the five Foundation/Core rows first. The gaps identify the next governance constraint. For the one-hour working session or Transform-path capstone, complete the Full Card.

Part 3: Card Specification

The AI Governance Card

The AI Governance Card is drafted in the governance tabletop for the working use case; there is no post-class submission. It connects directly to the shared course tools: AI value portfolio, AI inventory, AI risk register, build/buy/partner decision, KPI value chain, AI operating model, and 90-day plan.

Depth	Deliverable components
Foundation/Core	Use case and accountable owner; risk tier; authority statement; most important control or assurance gap; executive owner of residual risk
Challenge/Full Card	Complete control plan; assurance evidence; oversight, explanation, and recourse; audit-log requirements; incident-response design

Capstone build (Transform path). The complete Full Card becomes the "risk and governance" section of the final memo. It should help legal, risk, security, compliance, and business leaders see what must be controlled before the use case moves from pilot to production or from production to scale.

Next-module handoff. Module 5 tests whether the risk-adjusted use case still deserves investment once you add baselines, adoption evidence, TCO, FinOps, benefits ownership, and stop/pivot/scale criteria.

Works Cited and Sources

1. European Parliament and Council, Regulation (EU) 2026/1744 (Digital Omnibus on AI), *Official Journal of the European Union*, July 24, 2026; entered into force July 27, 2026, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202601744; European Commission, "Navigating the AI Act," implementation FAQ, <https://digital-strategy.ec.europa.eu/en/faqs/navigating-ai-act>
2. NIST, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," <https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-ai-rmf-10>
3. NIST, "NIST AI RMF Playbook," <https://www.nist.gov/itl/ai-risk-management-framework/nist-ai-rmf-playbook>
4. NIST, "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile (AI 600-1)," July 2024, <https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-generative-artificial-intelligence>
5. OWASP GenAI Security Project, "OWASP Top 10 for LLM Applications 2025," <https://genai.owasp.org/resource/owasp-top-10-for-llm-applications-2025/>
6. OWASP GenAI Security Project, "OWASP Top 10 for Agentic Applications for 2026," <https://genai.owasp.org/resource/owasp-top-10-for-agentic-applications-for-2026/>
7. Frank Liao, Sabrina Adelaine, Majid Afshar, and Brian W. Patterson, "Governance of Clinical AI applications to facilitate safe and equitable deployment in a large health system," *Frontiers in Digital Health*, 2022, <https://www.frontiersin.org/articles/10.3389/fdgth.2022.931439/full>
8. Salesforce Developers, "Trust Layer," <https://developer.salesforce.com/docs/ai/agentforce/guide/trust.html>

9. Stanford Institute for Human-Centered AI, *2026 AI Index Report*, April 2026, <https://hai.stanford.edu/ai-index/2026-ai-index-report>
10. *Mobley v. Workday, Inc.*, 3:23-cv-00770 (N.D. Cal.); conditional certification of ADEA collective entered May 16, 2025; HiredScore extension entered July 7, 2025; notice plan approved December 2, 2025. Civil Rights Litigation Clearinghouse docket, <https://clearinghouse.net/case/44074/>
11. NIST, "Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations (NIST AI 100-2 E2025)," March 2025 update, <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-2e2025.pdf>
12. International Organization for Standardization, "ISO/IEC 42001:2023 — Information technology — Artificial intelligence — Management system," <https://www.iso.org/standard/42001>
13. *In re OpenAI Copyright Infringement Litigation* (S.D.N.Y., MDL); Magistrate Judge Ona T. Wang preservation order of May 13, 2025, affirmed by Judge Sidney H. Stein, June 26, 2025; going-forward preservation terminated September–October 2025; sample-production order affirmed by Judge Stein, January 5, 2026.
14. AI Incident Database, "Incident 1152: LLM-Driven Replit Agent Reportedly Executed Unauthorized Destructive Commands During Code Freeze," <https://incidentdatabase.ai/cite/1152/>; reporting in *The Register*, July 21, 2025, and Amjad Masad public statement.
15. Optomed and AEYE Health, "Aurora AEYE — first FDA-cleared handheld AI fundus camera for autonomous diabetic retinopathy detection," cleared April 30, 2024; AEYE Health PR Newswire announcement, April 30, 2024; Optomed corporate release, May 1, 2024.
16. Lockheed Martin, "Lockheed Martin Launches Astris AI to Bring Trusted Defense AI to Industry," press release, December 16, 2024, <https://news.lockheedmartin.com/>
17. Sam Ransbotham and Shervin Khodabandeh, "Continuous Learning With AI: Aflac's Shelia Anderson," *MIT Sloan Management Review / Me, Myself, and AI* podcast, 2023.
18. *In re Caremark Int'l Inc. Derivative Litig.*, 698 A.2d 959 (Del. Ch. 1996); *Marchand v. Barnhill*, 212 A.3d 805 (Del. 2019); *In re McDonald's Corp. Stockholder Derivative Litig.*, 289 A.3d 343 (Del. Ch. 2023); *Giuliano v. Grenfell-Gardner* (*In re Teligent*), Del. Ch., September 2, 2025.
19. European Parliament and Council, Regulation (EU) 2024/1689 (Artificial Intelligence Act), current consolidated version of July 27, 2026, including Articles 4, 27, 50, and 73, <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>; amended by Regulation (EU) 2026/1744, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202601744; European Commission, "Navigating the AI Act," <https://digital-strategy.ec.europa.eu/en/faqs/navigating-ai-act>
20. MITRE, "ATLAS — Adversarial Threat Landscape for Artificial-Intelligence Systems," <https://atlas.mitre.org/>
21. European Commission, "The General-Purpose AI Code of Practice," final version July 10, 2025; in force August 2, 2025, <https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai>
22. Coalition for Content Provenance and Authenticity (C2PA), <https://c2pa.org>
23. Cristina Pullen, "AI Wardens: Legal Concerns for AI in Prisons," *Berkeley Technology Law Journal Blog*, December 17, 2025, <https://btlj.org/2025/12/ai-wardens-legal-concerns-for-ai-in-prisons/>
24. Jer Crane, founder, PocketOS, public incident post-mortem, April 2026; *The Register*, "Cursor-Opus agent snuffs out startup's production database," April 27, 2026; *Fast Company*, April 29, 2026. Railway CEO Jake Cooper publicly confirmed the deletion mechanics.
25. Cloud Security Alliance and Token Security, *Autonomous but Not Controlled: AI Agent Incidents Now Common in Enterprises*, published April 21, 2026 (survey of 418 IT and security professionals, fielded January 2026; commissioned and financed by Token Security); Cloud Security Alliance and Zenity, *Enterprise AI Security Starts with AI Agents*, April 16, 2026.
26. Centers for Disease Control and Prevention, "CDC's Vision for AI in Public Health," March 13, 2026, <https://www.cdc.gov/ai/vision/index.html>

Copyright, reuse, and citation

© 2026 Gerald C. Kane. This module is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0), <https://creativecommons.org/licenses/by-nc/4.0/>.

You may share and adapt it for noncommercial purposes, provided you credit Gerald C. Kane as the author, link to the license, and indicate any changes you made. You may not present this material, or adaptations of it, as your own work, and you may not sell it or use it in paid training or consulting without written permission.

Cite as: Kane, G. C. (2026). *Module 4: Responsible AI, Security, and Governance*. EMBA-7450, Information Technology and Strategy, Executive MBA, Terry College of Business, University of Georgia. <https://www.profkane.com/course>

Permissions and commercial use: jerry@profkane.com