

Executive AI Literacy

Prompting, Context, and Judgment

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: Executive literacy module. Module 1 established the 2026 AI strategy landscape and the distinction between AI activity and AI value creation. Module 2 turns from the strategic frame to the executive's own bench — what the leader needs to know to specify, evaluate, and delegate AI-assisted work without weakening judgment, privacy, accountability, or trust. Subsequent modules go deeper on the technology stack and sourcing, governance and security, economics and scaling, capability acceleration, and the capstone action plan.

Module Learning Objective

By the end of the module, you will be able to convert a vague executive request into a well-scoped AI work brief, specify the context the system needs and the data it must not receive, ask for an output format that is inspectable rather than impressive, critique the result for unsupported claims and hidden assumptions, diagnose which layer of the system produced a disappointing output and fold the fix back into the standing instructions, and decide what level of authority the system should be granted in a recurring workflow.

If You Want to Know More

The Pre-Class Read below carries the argument and is the only assigned reading. The materials below are reference — read what fits your time and interest:

- Stanford HAI, *2026 AI Index Report* — adoption, capability, responsible-AI, and public-trust indicators.
 - McKinsey, *The State of AI in 2025: Agents, Innovation, and Transformation*.
 - One current model system card — Anthropic's Claude Opus 4.8 system card (May 2026) or OpenAI's GPT-5.5 system card (April 2026). A 20-minute skim is enough to see what the leading labs disclose and what they don't.
 - OWASP, *Top 10 for LLM Applications 2025*.
 - NIST, *AI Risk Management Framework* (AI RMF 1.0) and the *Generative AI Profile* (NIST AI 600-1, July 2024).
-

Part 1: Pre-Class Read

BOTTOM LINE. Executive AI literacy in 2026 is not the skill of writing better prompts. It is the discipline of delegating cognitive work — specifying what the AI should do, what context it needs, what it must not see, what the output should look like, and where a human re-enters the loop. The work happens at three layers: the prompt (what to do), the context (what the model needs to know), and the harness (the tools, permissions, evaluations, and approval gates that make the work dependable). Leaders who do this well treat AI the way they treat capable analysts — clear assignment, prepared briefing, defined review. Those who do not tend to find that the technology amplifies the weaknesses of their delegation rather than its strengths.

Opening Vignette: Lockheed Martin and the Architecture of Authority

Lockheed Martin began deploying generative AI across its enterprise in the early 2020s and now runs one of the most extensive named programs in the defense industrial base. By 2025, roughly 75,000 employees had access to Navigator, the company's secure internal chat system, and the company's Genesis platform had been used to build more than 16,000 internal AI agents since the start of that year — analytical assistants, document workflows, engineering aids, knowledge-base retrievers, and supervised digital workers operating inside narrow boundaries.¹ On December 16, 2024, the company launched Astris AI, a subsidiary that externalizes the AI Factory and Genesis platform to other defense contractors and high-assurance commercial customers under the Lockheed Martin Evolve umbrella.²

Mike Baylor, Chief Digital and AI Officer at Lockheed Martin, and Greg Forrest, Director of AI Foundations, are the program's public faces. The strategic detail that matters here is not the count of agents. It is the deliberate placement of each agent on a spectrum of authority — some allowed only to draft, some to recommend, some to act inside narrow approved systems, very few permitted to act broadly without human approval — and the policy infrastructure that surrounds the placement. Lockheed Martin has not built one AI system. It has built thousands of them, each at a defined level of authority, each with a defined human approval gate, each with audit and oversight attached. Lockheed had to formalize that placement at scale because the consequences of getting authority wrong in a high-assurance environment are severe. Most organizations face the same question on smaller stakes, and Module 2 is about the discipline that lets you answer it.

The Most Common Executive AI Failure

Perhaps the most consequential shift to recognize as 2026 begins is that the most common executive AI failure is not a poorly worded prompt. It is an unclear work order. Most readers approach AI literacy expecting the central question to be one of phrasing — which words to use, which keywords to include, which clever framing to try. The data and the cases increasingly point elsewhere. AI capability is converging across leading providers; phrasing has stopped being the

limiting factor for most senior tasks. What separates work that produces useful executive output from work that produces plausible but unreliable output is the discipline of the delegation itself: whether the leader has specified the decision the work supports, the audience that will use it, the context the system needs, the data it should not receive, the output that would be inspectable, and the criteria by which a human will judge quality. The shift is not from short prompts to long prompts. It is from informal asks to disciplined delegation. That shift is why this module spends very little time on prompt-writing technique and most of its time on the management discipline that surrounds the prompt.

Why Executive AI Literacy Matters in 2026

AI is now in routine use across most large enterprises. Stanford and McKinsey both report adoption above 85%, with generative AI used in at least one business function in roughly 70% of surveyed organizations (the full statement of these figures sits in Module 1).³⁴ McKinsey's 2025 survey adds that 23% of organizations are scaling agentic AI in at least one function.⁴

Two named CEO statements from 2025 frame the executive question at opposite ends. Tobi Lütke, CEO of Shopify, told employees in an internal memo that became public in April 2025 that "AI usage is now a baseline expectation" at the company, and that teams must justify why AI cannot do work before headcount is approved (*Tobi Lütke, CEO, Shopify, internal memo published April 2025*).

⁵ Aaron Levie, CEO of Box, told a TechCrunch Disrupt audience in October 2025 that the operational reality is more complicated, observing that "we've already seen great examples of either data being leaked because of agents, or an agent going and... maybe blowing up your database or doing something in production you didn't expect" (*Aaron Levie, CEO, Box, TechCrunch Disrupt 2025*).⁶ Both statements are accurate. The gap between them is the executive-literacy gap.

The fiduciary stakes have hardened alongside the adoption. Regulators have brought AI-washing enforcement actions against private and public companies since 2024, AI risk disclosure has moved from exception to norm in S&P 500 annual reports, and securities litigation over AI claims has followed the disclosures. The Field Reference carries the enforcement record, and Module 4 governs the response. The point that belongs here is narrower and more personal: the executive who cannot specify, evaluate, or constrain an AI workflow has become a fiduciary risk to their own company.

McKinsey's 2025 *State of AI* survey reports that only 17% of organizations say AI governance is overseen by their board, and 28% say it is overseen by the CEO; McKinsey's broader finding is that most organizations report little or no enterprise-level financial impact from AI yet, even at scale.⁴ Access is widespread. Disciplined use is not. That gap is what this module is built around.

From Prompt to Work Order

We can illustrate the executive shift with a simple contrast. Here is a request shaped like a typical executive prompt:

Analyze our Q3 customer churn and tell us what is going on and what we should do.

And here is the same request shaped as an AI work brief:

We are a mid-market B2B SaaS company. Q3 logo churn increased from 4.0% to 6.0%, above our 4.5% target. Use only the sanitized case data provided below. Act as a customer-retention strategist advising the Chief Customer Officer. Produce a one-page executive memo with five sections: headline finding, top three churn drivers, evidence table, recommended Q4 actions, and assumptions or missing data. Separate facts from inferences. Do not claim causality unless the data supports it. Flag any recommendation that would require finance, legal, product, or customer-success approval before action.

The point is not that the second version is longer. The point is that it gives the system a better work order. It tells the model what decision the work supports, what facts are available, what boundaries matter, what output is useful, and where human approval is required. The model has more to go on, and — equally important — the leader has more to inspect when the answer comes back.

The first prompt asks the model to invent the structure. The second asks the model to fill in a structure the leader has already designed. The second is also reviewable: every section of the requested memo gives the leader a place to check the work. The first produces an answer that is hard to disagree with point by point, because there are no points.

The rest of the module is the discipline that lets you write the second version reliably for a recurring workflow.

The Three-Layer Model

AI work happens at three layers. We use this model throughout the course because it carves the executive's responsibility cleanly and travels well between technical teams and the boardroom.

The **prompt layer** is what the leader writes — the role, the task, the audience, the output format, the criteria, the explicit constraints. It is the assignment.

The **context layer** is what the model is given to work with — the source documents, retrieved knowledge, definitions, examples, history, sanitized data, and information the model should not receive. It is the briefing packet.

The **harness layer** is the operational scaffolding around the work — the tools the system can call, the data it can reach, the permissions and audit trails, the evaluations that grade the output, the logging, the human approval points, and the rules about what the system may or may not do after producing a result. It is the workplace.

The layers nest. A prompt sits inside a context, and a context sits inside a harness, and each is a different unit of improvement: the one-off request, the recurring task, the production workflow. Most consumer-facing chat use happens at the prompt layer. Most enterprise-grade AI lives or dies at the harness layer, and the reason is worth stating plainly: prompts and context can only *influence* what a system does; a harness can *enforce* it.

Layer	Executive question	Employee analogy
Prompt	What should the system do?	Give a clear assignment.

Layer	Executive question	Employee analogy
Context	What does the system need to know?	Prepare the briefing packet.
Harness	How is the work made dependable?	Design the workplace, the tools, the supervision, the safety checks, and the reporting.

Module 2 spends most of its time at the prompt and context layers, because those are where executive responsibility is most direct. Module 3 takes up the harness — the stack and the sourcing choices that sit underneath. Module 4 governs the risk. Module 5 prices the cost of running the work at scale. Module 7 asks who keeps the prompts, projects, knowledge files, and evaluation routines healthy over time.

When the Output Is Wrong, Which Layer Failed?

The three layers earn their keep at the moment the work comes back wrong. The instinct in that moment is to rewrite the prompt, because the prompt is the part the leader touched. It is frequently the wrong repair. Disappointing AI output has five distinct causes, and each one has a different fix.

The failure may sit in the **model**. The task fails with a clear prompt and good context, it fails again on a second attempt, and it fails the same way across providers. The task is outside what today's systems do reliably. Redesign it, keep it human, and retest next quarter.

It may sit in the **prompt**. The output is generic, or it answers a question adjacent to the one that was asked. The test is whether a capable colleague reading the same request would have known what was wanted. If they would not have, the model was never the problem.

It may sit in the **context**. The output is fluent, confident, and wrong about your business, your terms, or your documents. The system was not given what it needed to know, or it was given so much that the relevant material was buried.

It may sit in the **data**. The source material is stale, inconsistent, or missing, and the system is faithfully reporting a problem that predates it. No prompt repairs bad data.

Or it may sit in the **workflow**. The output is fine. Nothing changed. The pilot works and the process ignores it. This is the one failure on the list that is not a technical failure at all, and it is the most common one at scale.

Naming the layer is most of the repair. A leader who answers every disappointing output by rewording the request will spend a year rewording requests. The Field Reference carries the diagnostic as a table you can run against a specific failure.

The AI Work Brief

The AI Work Brief is the executive-grade replacement for the prompt template. It is the management equivalent of the briefing packet a leader prepares for a strong analyst — specific enough to guide the work, constrained enough to prevent overreach, and structured enough that someone else could pick it up, run it, and review the result.

Element	Executive question	Example
Decision	What decision or work product is this supporting?	Prepare for a board discussion on Q3 churn.
Audience	Who will use the output?	Chief Customer Officer and executive team.
Role	What perspective should the system adopt?	Customer-retention strategist for B2B SaaS.
Context	What background does the model need?	Churn rose from 4.0% to 6.0%; Q3 target was 4.5%.
Data boundary	What information is allowed; what is prohibited?	Use only the sanitized case packet; do not infer customer names.
Task	What should the system do?	Identify likely churn drivers and recommend Q4 actions.
Output	What format should the result take?	One-page memo plus evidence table.
Evidence	What must be sourced, cited, or labeled as assumption?	Separate facts, inferences, and missing data.
Constraints	What should the system avoid?	Do not recommend discounting without margin analysis.
Review	How will a human judge quality?	Recommendations must be specific, feasible, and risk-aware.

One field in the brief deserves a correction, because it carries the most widely repeated piece of prompting advice in circulation. Nearly every executive has been told to open a prompt by assigning the system a role — *act as an expert strategist*, *act as a skeptical CFO* — the first move in the everyday Role → Task → Format shorthand taught in class. The research suggests the technique does not do what most people believe it does. The most systematic test to date, covering 162 personas across four model families and 2,410 factual questions, found that adding a role did not improve accuracy over adding none, and that attempts to pick the best role for a given question performed no better than random selection.⁴⁷ What a role reliably changes is the output's register: its vocabulary, its emphasis, what it foregrounds and what it challenges. A skeptical-CFO role produces work that reads like a skeptical CFO, which is genuinely useful when the output must land with one. It is not more likely to be right. The technique feels like it works because a shift in register reads as a shift in competence, and the distinction between what actually changed and what only appears to have changed is worth carrying well beyond prompting. It is the same discipline the Review field enforces: judge the work, not the voice it arrives in.

If a leader cannot fill in this template for a recurring AI workflow, the workflow is not yet ready to be repeatable. The discipline of completing the brief is what surfaces the questions that should have been answered before the AI was asked to start.

The AI Work Brief also shows how the three layers connect. The prompt layer sits in role, task, output, and audience. The context layer sits in business background, source packet, data boundary, and definitions. The harness layer sits in review criteria, approval gates, source-checking discipline, and the rules about what the system may or may not do after producing the output.

The work brief is also where AI literacy becomes visibly a management skill. Good leaders already know how to delegate work to people. They clarify the assignment, describe the audience, share relevant context, identify constraints, and define what good work looks like. AI raises the same management questions — but with less common sense, less accountability, and a much greater ability to sound confident while being wrong. The Field Reference's AI Delegation Card extends the

Work Brief from a single piece of work to a recurring workflow — the brief specifies one assignment; the card specifies the harness, approval gates, and re-entry conditions that make the workflow repeatable.

Let the System Interview You

The Work Brief has one hidden assumption: that the leader already knows which context matters. Frequently they do not, and the template becomes a blank form rather than a tool. There is a reliable shortcut, and it inverts the usual failure. Open with the interview:

I need help with [task]. Before you draft anything, ask me the questions you need answered to produce the best possible output. Ask them one batch at a time.

A capable model comes back with the questions a good chief of staff would ask. Planning a leadership offsite: what decisions must be made by the end of day two, who is in the room, what did last year's offsite get wrong, what topics are off the table, how much unstructured time does this group actually tolerate. Ten minutes of answering produces a context packet that no first-draft prompt would have contained.

The pattern works because each party supplies what it has. The system does not know your situation. It has seen the shape of ten thousand situations like yours, and it knows which details distinguish them. Let it ask. The answers become the Work Brief, and the brief you produce this way will be better than the one you would have written from a blank page.

Talk Instead of Type

The Work Brief and the interview pattern share a failure point: the leader who will not type that much. The instinct is reasonable. Typing rations detail — every additional sentence of context has a cost, so the brief gets compressed to what feels essential, and the system starts work missing most of what a human delegate would have absorbed in the first five minutes of a conversation. Speech removes the rationing. Spoken text entry runs roughly three times faster than typing, with a lower error rate, in controlled comparison.⁴⁶ Two minutes of talking through a situation — who the output is for, what good looks like, what not to touch, what went wrong last time — produces a context packet most executives would never sit down and type.

The objection that stops leaders here is that a dictated brief feels undisciplined: hesitations, repetitions, sentences that change direction midway. But tidiness at input is not what the system needs. Volume of specific context is. Imposing structure on unstructured input is precisely the work these systems do well; what they cannot do is recover the context you never gave them. The polish you would have spent ten minutes adding at the keyboard is polish the model would have discarded anyway. Say it messy. Then ask the system to organize what you said into a structured brief, confirm the brief is what you meant, and set it to work.

In the vocabulary of the Three-Layer Model, voice is the cheapest instrument the context layer has. It does not change what the system can do; it changes how much of your situation the system knows before it starts — and every major tool now takes voice input, on the phone as well as the desktop, which means the brief can be dictated in the car after the meeting that prompted it, while the context is still fresh.

Context Engineering and Privacy Boundaries

Context engineering is the executive skill of deciding what the system needs to know, what it does not need to know, and what it must not receive. That makes it both a productivity question and a risk question.

A useful context packet usually includes the business situation, the decision to be made, relevant constraints, source material or sanitized data, definitions of key terms, audience expectations, known uncertainties, and information the system should not assume. The point is not to dump every related document into the tool. Too much context can distract the model, increase cost, expose sensitive information, and make errors harder to find. A 2025 study from Chroma Research tested 18 frontier models — including GPT-4.1, Claude Opus 4, Gemini 2.5, and Qwen3 — and found measurable, non-uniform output degradation as input length grew, well before the model's nominal context-window limit; the so-called "lost-in-the-middle" effect produced sharp drops in retrieval accuracy on facts buried in long inputs.¹² More context is not better context. Smaller, sharper packets tend to produce better answers.

Georgia-Pacific's ChatGP — a generative AI tool built on Amazon Web Services and deployed across the company's more than 140 facilities — illustrates context engineering at industrial scale. The system surfaces operator and maintenance knowledge from documents, digital files, and the tribal knowledge that lived in long-tenured employees rather than in any document store. The work product is a curated knowledge corpus disciplined to the maintenance and operations decisions the model is asked to support (*Roshan Shah, Vice President of Applied AI, Georgia-Pacific, AWS case study, 2024*).⁴⁵ The discipline of deciding what knowledge belongs in the corpus and what does not is what the executive owns. The technology is the easier part.

A compact context-quality test:

Context-quality test	Executive question
Relevant	Does this information directly help the task?
Sufficient	Is anything a capable human reviewer would need missing?
Isolated	Have we excluded distracting, contradictory, stale, or restricted material?
Economical	Can we cut tokens without lowering quality?
Provenance-aware	Can the reviewer tell where key facts came from?

Good context is not more information. It is the smallest high-signal packet that makes the task plausibly solvable and reviewable.

Data boundaries are where context engineering becomes a control question. Most senior leaders now operate inside organizations where employees have routine access to AI tools that route data outside the company's perimeter. Microsoft's 2024 Work Trend Index reported that 78% of employees who use AI at work bring their own tools to the job rather than wait for sanctioned ones.¹³ IBM's 2025 *Cost of a Data Breach Report* found that breaches involving shadow AI cost organizations approximately \$670,000 more on average than breaches without it.¹⁴ An analysis of approximately 22 million enterprise prompts published by Harmonic Security in 2025 found that legal documents — including merger workpapers, settlement materials, and litigation strategy —

were the largest single category of sensitive data employees pasted into consumer AI tools, accounting for roughly 35% of the exposed material.¹⁵ The leak that ends up on the front page is rarely the result of a sophisticated attacker. It is the result of an executive or analyst pasting the wrong file into the wrong tool.

A short data-tier discipline:

Data type	Class-exercise rule	Executive rule
Public	Usually acceptable.	Verify source quality, copyright constraints, and freshness.
Internal	Use only if sanitized or approved.	Check company policy, tool terms, and access controls.
Confidential	Do not use in class tools.	Use only in approved enterprise systems with appropriate controls.
Restricted	Do not use.	Requires formal approval, technical controls, and often legal, compliance, or security review.

Restricted information typically includes personally identifiable information, protected health information, payment data, unreleased financials, trade secrets, privileged legal material, employee records, regulated customer data, credentials, and source code from sensitive systems.

The practitioner lesson is blunt: privacy is not solved by writing "keep this confidential" in a prompt. The real controls are the tool, the contract, the data architecture, the permissions model, the retention settings, and the organization's policy. The most cited reminder of this is Samsung Electronics's 2023 incident, in which engineers pasted internal source code and meeting transcripts into a consumer AI tool to summarize them; the company subsequently restricted employee use of public generative-AI services on company devices.¹⁶ A prompt can express an instruction. It cannot create a legal or technical boundary by itself.

Thread Hygiene and the Twentieth Turn

Context engineering has a daily operating counterpart that almost nobody is taught. The model's working memory is finite. Everything in the conversation competes for room in it, and a conversation that has run long enough accumulates false starts, abandoned framings, superseded drafts, and corrections that the system is still weighing against the instruction you gave it forty minutes ago. The Chroma Research finding cited above applies to the conversation, not only to the attached file: output quality degrades measurably as input length grows, and it degrades well before the tool reports any limit.¹² Practitioners call the failure mode context rot. You will recognize it as the long conversation that has become strangely worse, and as the moment you find yourself explaining to a system something you know you already told it.

The remedies are simple enough to state in a paragraph, and they are much of the difference between an executive who gets steady results and one who blames the tool.

One thread per task. When the task changes, open a new conversation. Stretching an old one is free, and it is not cheap.

One project per job. Most tools now offer a persistent workspace that carries standing instructions and a curated set of files into every conversation started inside it. Give each recurring

job its own. The isolation is the feature, not a limitation of the tool. A workspace built for board materials should not be reaching into the file you loaded for a personnel matter.

Ask the system. Before starting a new task, ask whether it belongs in this thread or a fresh one. Models answer that question well, and a good one will offer to write the handoff.

Carry the work forward deliberately. A half-page briefing memo — decisions made, constraints established, what remains open — pasted into a fresh thread will outperform the tired thread it replaces.

Distill the corpus before you load it. The most common context failure is loading everything on the theory that more is safer. An executive who loads a decade of archives and asks the system to work from them will get an answer that reads well and was assembled from the first pages, the last pages, and the system's own guess at the middle. The disciplined move is to ask the system to *analyze* the archive and produce a small set of derived artifacts: a summary of the house style, the recurring positions, a handful of exemplars of good work, a list of things never to do. Keep only those artifacts in the project. Hundreds of pages become half a dozen files, and the system stops guessing.

These practices are working calibrations drawn from practitioner use rather than findings from controlled study. The degradation they respond to is measured. The remedies are craft, and you should expect to adjust them to your own tool.

The vocabulary is worth learning because it makes failure diagnosable. Practitioners name five operations on context: **write** it down outside the conversation so it survives, **select** the right material at the right moment, **compress** long material so the signal survives the shortening, **isolate** separate jobs in separate workspaces, and **cache** stable material so repeated work is not paid for twice. Module 3 shows what these five look like as enterprise architecture. At your own desk they are a diagnostic: when the work degrades, the question stops being "the AI got worse" and becomes "which of the five failed?"

Structured Outputs and Executive Usability

Executives should learn to ask for outputs that can move directly into work. A long, well-written paragraph may be impressive, but it is often the least useful form of an answer. The better question is: what shape would make the output easier to inspect, compare, challenge, and reuse?

For executive work, useful formats include a one-page decision memo, a board briefing outline, a comparison table, a risk register, an assumptions log, a stakeholder map, a KPI value chain, a build/buy/partner option screen, a 90-day action plan, a meeting-prep brief, or a red-team critique. These formats do more than make the answer look tidy. They force the system to separate claims, evidence, tradeoffs, assumptions, and recommendations into places a reviewer can examine.

Structured outputs also make weak reasoning easier to spot. If every recommendation has to sit next to its evidence, its uncertainty, its approval gate, and its owner, the row that has no evidence becomes obvious. The reviewer does not have to read the answer back to themselves to find the gap.

A short example. A weak request:

Give me your thoughts on whether we should adopt this AI tool.

A practitioner version:

Produce a decision table comparing four options: adopt now, pilot first, wait, and reject. Use these criteria: strategic value, implementation effort, data sensitivity, vendor lock-in, workflow disruption, compliance risk, and expected ROI evidence. Rate each option on each criterion as high, medium, or low. End with a recommendation, two disconfirming questions, and the human approval gate required before moving forward.

The stronger request does not guarantee a correct answer. It does make the answer more inspectable.

A note on reliability that the Field Reference treats more fully: structured-output and JSON modes are not a guarantee. Production benchmarks find that prompt-only structured outputs fail to parse 5% to 20% of the time at scale, and even strict-schema enforcement does not prevent semantically wrong fields inside structurally valid responses.¹⁷ Asking for structure is good discipline; trusting structure as a stand-in for correctness is not.

AI Critique and Source Checking

A leader should treat AI output as a draft that has to earn its way into executive work. The review question is not "Does this sound good?" The review question is "What would have to be true for us to rely on this?"

A practical critique looks for the difference between facts, interpretations, assumptions, and recommendations. It asks what evidence supports the claims, which claims are unsourced or stale, what the model assumed, what data would change the recommendation, which stakeholders are missing, and what a skeptical CFO, COO, general counsel, CHRO, CIO, customer, or regulator would challenge.

A short critique prompt that surfaces the structure:

Review your prior answer. Create a table with four columns: claim, evidence provided, source or data basis, and verification needed. Mark each claim as fact, inference, recommendation, or assumption. Do not add new claims. Flag anything that should not be used in an executive memo without independent verification.

For public factual claims, you should verify against reliable sources. For internal claims, you should verify against internal systems of record. For strategic recommendations, the review is not only source-checking; it is assumption-testing. A recommendation can have accurate facts and still be strategically weak if it ignores constraints, incentives, stakeholders, second-order effects, or political reality.

The empirical case for an explicit critique routine is hardening. Researchers at Microsoft Research and Carnegie Mellon University surveyed 319 knowledge workers in early 2025 and found that higher self-reported confidence in an AI's output was associated with *less* effort spent on critical thinking; researchers in the MIT Media Lab's *Your Brain on ChatGPT* study (Kos'myna and colleagues, June 2025) used EEG to compare 54 participants writing essays with and without an AI assistant over four months, and reported lower neural connectivity, weaker recall of one's own arguments, and lower self-reported essay ownership in the AI-assisted group.^{18 19} Both studies have methodological limits — survey self-report on the Microsoft-CMU side, small sample and a non-peer-reviewed preprint on the MIT side — and should be cited as suggestive rather than conclusive. They both point in the same direction. Routine, frictionless AI use tends to displace the cognitive effort that produces good judgment unless the user actively reintroduces it through structured critique.

The broader literature points the same way. A 2025 systematic review in *AI & Society* examined 35 studies and found that automation bias persists across explainability strategies, and that more detailed explanations sometimes *increase* over-reliance rather than reduce it.²⁰ The European Union's AI Act addresses the finding directly. Article 14(4)(b) requires deployers of high-risk AI systems to remain "aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system (so-called 'automation bias')." ²¹ It is the only place a regulator has codified the calibration gap into binding obligation.

The practical implication is that the critique routine is not optional polish on top of AI output. It is the workflow step that prevents AI work from quietly displacing the judgment that justifies the executive's role.

The Tuning Loop

Nobody writes a good work brief on the first pass. Nobody configures a good project on the first pass either. The organizations getting reliable results from AI converge on the same loop, whatever they call it internally: build, test, critique, verify, fold the fixes back into the instructions, and run it again.

Build the first version with the Work Brief. Test it on real work, ideally work where you already know what good looks like, because you cannot grade an output on a question you cannot grade yourself. Critique the result against your own criteria, and have the system critique it as well. Verify the load-bearing facts.

Then comes the step almost everyone skips, and it is the one that compounds. Fold what you learned back into the written instructions, so that the next run starts where this one ended. An output that arrived with a fabricated statistic earns a standing line requiring a source for every figure. An output that overstepped into legal advice earns a standing constraint. Each fix is made permanent in the specification rather than applied by hand to a single result, and the same mistake cannot recur. Without that step, every improvement you discover dies in the conversation where you found it, and you will pay for the same lesson repeatedly.

Two rules belong inside the loop.

Write the traceability rule into the instructions. For any work involving facts you will rely on: require a source for every figure and every quoted claim, allow quotation marks only around exact wording, and instruct the system to flag what it could not retrieve rather than filling the gap. This is the standing-instruction version of the critique routine above. It costs a paragraph and it removes the most dangerous failure mode from the output.

The system doing the work does not grade its own homework. Self-critique is a useful aid to review, and it is not verification. When the stakes justify it, the check comes from a separate pass, a different model, or a named human. The review that counts is the one the drafter did not perform.

One discipline turns the loop from a personal habit into an organizational asset: one owned master, not a hundred forks. When a brief or a project configuration proves out, it gets an owner, a home, and a version history, and the team works from the master rather than from last month's copy pasted out of an old thread. Module 7 takes up what happens when fifty people are running fifty loops and nobody owns the artifacts. At the level of your own desk, the discipline is nothing more exotic than editing a document after every run.

The habit this teaches is specification: writing down what "done" means before the work starts, and improving the written version rather than the individual output. These practices, like the thread-hygiene routine above, are working calibrations drawn from practitioner use rather than findings from controlled study.

AI Delegation and Authority Levels

Executive AI literacy connects to agentic AI through one core question: what authority should the system have? The course uses one authority ladder throughout Modules 2, 3, 4, 6, 7, and the capstone, so you can separate three questions cleanly: what the system may do, what evidence is required, and where human accountability remains.

Level	Course term	Permitted authority	Human control
1	Assistant	Drafts, summarizes, or analyzes only.	Human reviews all output before use.
2	Copilot	Recommends actions inside a human-led workflow.	Human steers, approves, and remains the decision-maker.
3	Workflow Agent	Completes bounded tasks in approved systems.	Human approves key gates, exceptions, and escalation points.
4	Supervised Digital Worker	Acts across systems within defined thresholds.	Human sets boundaries, monitors exceptions, and owns outcomes.
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.

Two patterns matter at the executive level.

First, the levels are not a maturity ladder where higher is better. Higher is more autonomous, and therefore more consequential when it fails. McKinsey's 2025 survey finds that most organizations report little or no enterprise-level EBIT impact from AI despite widespread deployment,⁴ and Gartner's 2025 forecast projected that more than 40% of agentic AI projects would be cancelled by 2027 due to escalating costs, unclear value, or inadequate risk controls.²² The reliability problem

grows with autonomy because errors compound across steps. A model that hallucinates one fact in a memo is a quality issue. The same model embedded in a workflow that pulls customer records, takes actions, and updates downstream systems is a different category of risk altogether. An organization that has not earned the right to operate at level 3 should not be at level 4.

Second, the human re-entry point is not a vague aspiration. At levels 3, 4, and 5, the executive question is what specific conditions trigger a human review. The standard set in production deployments includes:

- **Low confidence** — the system signals uncertainty, or its outputs vary across runs.
- **Sensitive data** — the request touches restricted information categories.
- **Irreversible action** — the action cannot be cleanly undone.
- **Cross-system action** — the action propagates to systems outside the approved scope.
- **Threshold breach** — the action exceeds defined dollar, volume, or scope limits.

These conditions are the operational version of the Article 14 obligation quoted earlier: deployers of high-risk systems must ensure that a person can intervene effectively and override the output.²¹

The named cases make the levels concrete. Truist's banker-facing assistant sits at the Copilot tier in a regulated banking context: the system surfaces information and suggests actions, and the banker remains the decision-maker on any client-facing commitment.²³ Aflac's claims platform, covered in Module 1, sits at Supervised Digital Worker inside a bounded category, processing low-complexity claims under a threshold and routing higher-stakes cases to humans by design (*Shelia Anderson, then CIO, Aflac, MIT Sloan Management Review podcast, 2023*). Lockheed Martin's Genesis platform, covered in the opening, runs at every level at once, with formal policy attached to each transition. The accountability shift between levels is the executive lesson. The technology is downstream of it.

McKinsey's 2025 *State of AI* survey reports that only 17% of organizations say AI governance is overseen by their board, and 28% say it is overseen by the CEO.⁴ At the Workflow Agent tier and above, the question of who owns the outcome when the system fails is not a technical question. It is a governance question with a fiduciary edge. The Authority Ladder is a way of forcing that question to the surface before the deployment, not after the incident.

AI Literacy as a Management Skill

A useful way to read everything in the module so far: AI literacy is the application of management discipline to a new category of worker. Josh Blacksmith, Senior Director of Global Integrated Marketing and Experiences at Kimberly-Clark, framed the company's approach to *Consumer Goods Technology* in October 2025 in a single line that captures the right register: "AI helps us start from the same place... we talk about being business-led, tech-enabled" (*Josh Blacksmith, Senior Director of Global Integrated Marketing and Experiences, Kimberly-Clark, Consumer Goods Technology, 2025*).²⁴ Business-led, tech-enabled is not a slogan. It is a delegation discipline. The leader sets the work, the system enables it, the organization absorbs the result.

The leaders who do this well are not the most enthusiastic adopters. They are the ones who treat AI as an analyst worth briefing properly, an output worth checking honestly, and an authority worth granting cautiously.

Executive AI literacy in 2026 is the discipline of delegating cognitive work the way you have always known good delegation should be done. The systems are new. The discipline is not. Bring it to class with you, and bring it back to your work after.

Bring This to Class

Come prepared with one specific recurring executive workflow from your own organization or a chosen company — a workflow you would like to make more dependable through AI assistance. You will use it in the in-class lab and in the AI Delegation Card you draft in session. Useful candidates include a recurring board-briefing prep, a vendor-evaluation routine, a customer-issue triage process, an M&A diligence question-generation step, a recurring market-monitoring task, or a post-meeting synthesis routine. Avoid one-off projects. The value of the module is in workflows that repeat.

Try This Before Class

If you have access to an enterprise AI tool you trust with internal information, run a small experiment in advance. Take the workflow you will bring to class. Write the request first as a one-line prompt. Then write it as an AI Work Brief using the template in the body of this module, and let the system interview you before you finalize it. Run both versions in the same tool, and bring the two outputs to class. We will compare what the discipline produced against what the casual ask produced, and we will diagnose the weaker output against the five failure layers.

Part 2: Field Reference

The Field Reference is the leave-behind. Each component below is self-contained, so you can flip directly to the table or panel you need without re-reading surrounding prose.

Decision Tables

Three-Layer Model — Executive Questions

Layer	Executive question	Employee analogy
Prompt	What should the system do?	Give a clear assignment.
Context	What does the system need to know?	Prepare the briefing packet.
Harness	How is the work made dependable?	Design the workplace, the tools, the supervision, the safety checks, and the reporting.

AI Work Brief Template

Element	Executive question
Decision	What decision or work product is this supporting?
Audience	Who will use the output?
Role	What perspective should the system adopt?
Context	What background does the model need?
Data boundary	What information is allowed; what is prohibited?
Task	What should the system do?
Output	What format should the result take?
Evidence	What must be sourced, cited, or labeled as assumption?
Constraints	What should the system avoid?
Review	How will a human judge quality?

Context Quality Test

Test	Executive question
Relevant	Does this information directly help the task?
Sufficient	Is anything a capable human reviewer would need missing?
Isolated	Have we excluded distracting, contradictory, stale, or restricted material?
Economical	Can we cut tokens without lowering quality?
Provenance-aware	Can the reviewer tell where key facts came from?

The Frontier Diagnostic

Run this against a specific disappointing output before rewriting anything. The point of the table is to stop the reflex of rewording the prompt when the prompt was not the problem.

Failure	Symptom	First fix
Model	Fails even with good context and a clear prompt; fails the same way across providers and across attempts.	Try the strongest available model with extended reasoning. If it still fails, the task sits outside today's frontier. Redesign it, keep it human, and retest next quarter.
Prompt	Output is generic or answers an adjacent question; a capable colleague reading the same request could not have told what was wanted either.	Rewrite as an AI Work Brief. State the decision the work supports and what good looks like.
Context	Output is fluent but wrong about your business, your terms, or your documents.	Give the system what it needed to know, and check what it actually received. Smaller and sharper beats more.
Data	The source material is stale, inconsistent, or missing; the system is faithfully reporting a data problem.	Fix upstream. No prompt repairs bad data. Module 3 owns the remedy.
Workflow	The output is fine and nothing changed operationally; the pilot works and the process ignores it.	Redesign who does what, and when. This is a leadership failure, not a technical one. Modules 5 and 7 own the remedy.

The Five Context Operations

The working vocabulary that turns "the AI got worse" into a diagnosable claim. Module 3 develops these as enterprise architecture; the executive uses them as a diagnostic at their own desk.

Operation	What it does	Executive question
Write	Persist information outside the conversation so it survives the thread.	Is anything load-bearing living only in a chat window?
Select	Retrieve the right material at the right moment.	Did the system actually receive what it needed for <i>this</i> task?
Compress	Condense long material so the signal survives the shortening.	Have we distilled the archive, or dumped it?
Isolate	Keep separate jobs in separate workspaces.	Is this project reaching into material from an unrelated job?
Cache	Reuse stable material so repeated work is not paid for twice.	Are we re-establishing the same context every week?

Thread Hygiene Checklist

Working calibrations drawn from practitioner use, not findings from controlled study. Adjust them to your tool.

Practice	The move
One thread per task	When the task changes, open a new conversation.
One project per job	Each recurring job gets its own persistent workspace, with its own standing instructions and its own curated files.
Ask the system	Before a new task, ask whether it belongs in this thread or a fresh one.

Practice	The move
Carry work forward deliberately	Move a half-page briefing memo, not the whole transcript.
Distill before you load	Ask the system to analyze a large corpus and produce derived artifacts. Keep the artifacts in the project; keep the corpus out.
Keep the deliverables out	Finished outputs do not belong in project knowledge. Retrieval stays clean when the project holds the current canon, not every artifact ever produced.

The handoff prompt, for the moment a thread has run long:

This conversation has run long. Produce a portable briefing memo I can paste into a fresh thread: the decisions we have made, the constraints we established, the current state of the work, and what remains open. Half a page. Do not summarize the conversation; summarize the state.

Data Sensitivity Tiers

Data type	Class-exercise rule	Executive rule
Public	Usually acceptable.	Verify source quality, copyright constraints, and freshness.
Internal	Use only if sanitized or approved.	Check company policy, tool terms, and access controls.
Confidential	Do not use in class tools.	Use only in approved enterprise systems with appropriate controls.
Restricted	Do not use.	Requires formal approval, technical controls, and often legal, compliance, or security review.

Authority Ladder

Level	Course term	Permitted authority	Human control
1	Assistant	Drafts, summarizes, or analyzes only.	Human reviews all output before use.
2	Copilot	Recommends actions inside a human-led workflow.	Human steers, approves, and remains the decision-maker.
3	Workflow Agent	Completes bounded tasks in approved systems.	Human approves key gates, exceptions, and escalation points.
4	Supervised Digital Worker	Acts across systems within defined thresholds.	Human sets boundaries, monitors exceptions, and owns outcomes.
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.

AI Delegation Card Template

Field	Description
Workflow name	
Business decision or output	
Intended user or audience	

Field	Description
Authority level	Assistant / Copilot / Workflow Agent / Supervised Digital Worker / Multi-Agent Process
Allowed data	
Prohibited data	
Required context packet	
Context-quality check	Relevant / sufficient / isolated / economical / provenance-aware
Required output format	
Source or evidence requirement	
Harness needs if repeated	Project, knowledge file, tool access, eval checklist, logging, monitoring
Human approval gate	
Re-entry conditions	Low confidence / sensitive data / irreversible action / cross-system action / threshold breach
Key risks	
Success metric	
Link to course framework	AI value portfolio / risk register / KPI value chain / build-buy-partner / operating model / 90-day plan

The card should not read like a generic prompt template. It should read like a small operating agreement for how AI will participate in a recurring piece of work.

Do / Don't Quick Reference

Prompt layer

Do	Don't
Specify the decision the work supports.	Open with "tell me about..." or "give me your thoughts on...".
Name the audience for the output.	Assume the model will guess the right register.
Use role to set the register and audience fit of the output.	Expect a role assignment to make the output more accurate.
Specify the output format.	Accept long prose when a table would be inspectable.
Define the criteria a human will use to judge quality.	Ask for "your best answer" without saying what good looks like.

Context layer

Do	Don't
Provide the smallest packet that makes the task solvable.	Paste every related document into the tool.
Define key terms inline.	Assume the model knows your organization's glossary.
State what the model should not assume.	Leave the model to guess at boundaries.
Mark provenance — where each fact came from.	Mix internal sources, public sources, and assumptions without flags.

Do	Don't
Sanitize sensitive material before it goes into the tool.	Rely on a "keep this confidential" prompt instruction.
Distill a large corpus into derived artifacts before loading it.	Load the archive and hope.
Start a new thread when the task changes.	Run one conversation for a week.

Harness layer

Do	Don't
Define the human approval gate before deployment.	Assume "we'll review the output" is the same as a defined gate.
Specify re-entry conditions for any agentic workflow.	Promote a chatbot to a workflow agent without defining when humans return.
Log the inputs, outputs, and tool calls.	Run unsupervised AI workflows without an audit trail.
Run pre-deployment evaluations on representative tasks.	Treat a successful demo as evidence of production reliability.
Build the override path into the workflow.	Wait for the first failure to discover the override is missing.
Fold every fix back into the standing instructions.	Correct the output by hand and move on.
Diagnose which layer failed before rewriting anything.	Reword the prompt reflexively when the context was the problem.

Cautionary Tales

Samsung's ChatGPT Data Leak (April-May 2023)

Samsung Electronics employees pasted internal source code, meeting transcripts, and confidential semiconductor data into ChatGPT to summarize and debug them. Within weeks of discovering the disclosure, Samsung restricted employee use of consumer generative-AI services on company devices and networks.¹⁶ The case is the canonical example of why prompt-level confidentiality instructions are not the same as data-architecture controls.

Schwartz v. Avianca and the Lawyer-Fabricated-Citations Ecosystem (2023-2025)

In May 2023, a New York attorney filed a federal court brief in *Schwartz v. Avianca* containing six fabricated case citations produced by ChatGPT. The judge sanctioned the attorney \$5,000. By late 2025 the pattern had expanded substantially. *Mavy v. Commissioner of Soc. Sec. Admin.* (D. Ariz., August 2025) sanctioned an attorney whose brief contained twelve fabricated cases; *Noland v. Land of the Free* (Cal. Ct. App., September 2025) imposed a \$10,000 fine and faulted opposing counsel for failing to detect the fabrications; *Johnson v. Dunn* (N.D. Ala., July 2025) disqualified two attorneys at a major law firm; in late 2025, a federal *prosecutor* in the Eastern District of North Carolina filed an AI-generated brief with fabricated quotes that a *pro se* plaintiff caught. Damien Charlotin's tracking database recorded the rate moving from roughly two cases per week before spring 2025 to several per day by year-end.²⁵ ²⁶ The lesson is not that AI makes mistakes. It is that delegated AI work that is not critiqued by a human will eventually expose the human's accountability.

Moffatt v. Air Canada (February 2024)

Air Canada's customer-service chatbot told a passenger he could apply for a bereavement-fare refund retroactively. The airline refused to honor the chatbot's statement, arguing that the chatbot was a "separate legal entity." A British Columbia Civil Resolution Tribunal rejected the defense and held the airline liable for negligent misrepresentation by its AI agent.²⁷ The case is the foundational reminder that a leader is responsible for what their AI tells customers, regardless of how the system was procured or how it is described.

EchoLeak / CVE-2025-32711 (June 2025)

In June 2025, security researchers at Aim Labs disclosed the first publicly documented zero-click prompt-injection attack on a production AI agent: a single crafted email could cause Microsoft 365 Copilot to exfiltrate sensitive documents on the user's next benign query. The vulnerability — assigned CVE-2025-32711 with a CVSS score of 9.3 — bypassed Microsoft's existing prompt-injection classifiers, link redaction, and content-security policy. Microsoft patched the issue server-side, but the case stands as the canonical demonstration that an executive whose Copilot reads "this morning's emails" is operating in exactly the threat model the OWASP LLM Top 10 places at number one.²⁸

Klarna's AI Customer-Service Reversal (May 2025)

Discussed in detail in Module 5; cross-referenced here for the Authority Ladder implication. Klarna placed AI at an authority level its organization had not yet earned — the customer-service AI was operating effectively as a Supervised Digital Worker at scale before the eval discipline, the customer-experience monitoring, and the recovery path that would have caught the quality decline were in place. The reversal landed within months of the September 2025 IPO. Klarna has since rebuilt the function as a hybrid: by late 2025 the AI assistant still handled roughly two-thirds of customer chats, with human agents guaranteed on request — a recalibration of authority level, not a retreat from AI.²⁹

Deloitte Australia's AU\$440K Refund (October 2025)

In October 2025, Deloitte Australia issued a partial refund to the Australian Department of Employment and Workplace Relations for a 237-page report containing fabricated academic references, a fabricated Federal Court quote with a misspelled judge's name, and citations to non-existent case law. The errors were surfaced by University of Sydney researcher Chris Rudge. The revised report disclosed that Azure OpenAI GPT-4o had been used in its production.³⁰ The case is a Big-Four professional-services firm delivering an AI-hallucinated executive deliverable to a government client — a near-perfect cautionary tale for the gap between AI capability and AI work that is reviewable.

Workbook Prompt: Build an AI Delegation Card for One Workflow

Take one recurring executive workflow from your organization. Walk through the AI Delegation Card template in the Decision Tables section above, completing each field in writing. The exercise is the discipline.

When you finish, read the card three times:

1. As if you were the person who would have to operate the workflow tomorrow, not the person who designed it. Are the boundaries clear enough that they could not be misunderstood?
2. As if you were the company's general counsel. Are the data boundaries, approval gates, and re-entry conditions sufficient to defend if the workflow produced a bad outcome?
3. As if you were the customer or counterparty on the other end of the workflow. Is the role the AI is playing one you would want disclosed?

If the card fails any of those three reads, the workflow is not yet ready to be a repeatable AI deployment. Re-engineer it before deploying.

Prompt Examples Library

Paired examples — weak prompt followed by practitioner work brief — across executive use cases. Use these as templates.

Board-briefing preparation

Weak. Help me prepare for the board meeting next week.

Practitioner. Prepare a board pre-read brief for our Q3 board meeting on November 14. The audience is our nine-person board, including three independent directors with operational backgrounds and two with financial backgrounds. Use only the financial pack and operating review document I provide. Produce a two-page memo with: a one-paragraph headline of the quarter, three numbered findings with evidence pulled from the documents, two strategic risks the board should be tracking, and two questions the CEO recommends the board ask of management. Separate facts from interpretations. Do not speculate beyond the documents. Flag any claim that should be verified with the CFO before the memo goes out.

Vendor selection

Weak. Should we go with Vendor A or Vendor B?

Practitioner. Produce a vendor comparison table evaluating Vendor A and Vendor B against the following criteria: product fit to our use case, security posture, data residency, contract flexibility, total cost over three years, vendor financial stability, and integration effort with our existing stack. Rate each vendor on each criterion as high, medium, or low based on the materials I provide. End with a recommendation, two disconfirming questions for each vendor, and a list of references I should call before signing. Do not include criteria for which I have not provided source material.

Customer-issue triage

Weak. What should we do about the unhappy customer?

Practitioner. I am the chief customer officer. A named enterprise account has escalated a complaint about onboarding support. Use the case-history transcript I provide. Produce a one-page response plan with: the likely root cause, three questions I should answer internally before responding, the recommended response action, the executive who should own the customer conversation, and the

internal stakeholders who should be informed. Do not infer details that are not in the transcript. Do not propose remediation that would require finance, legal, or product approval without flagging.

M&A due-diligence question generation

Weak. What questions should we ask in diligence?

Practitioner. We are evaluating an acquisition of a privately held mid-market company in the same industry. Use the public materials and the management presentation I provide. Produce a structured diligence question list across financial, operational, commercial, technology, talent, and legal categories. For each question, mark it as a confirmatory question, a verification question, or a red-flag-investigation question. Identify the three highest-priority questions across all categories and the rationale for each. Do not propose questions outside what the materials suggest is in scope.

Performance-review preparation

Weak. Help me write a performance review for one of my direct reports.

Practitioner. I am preparing the year-end review for one of my direct reports. Use only the materials I provide: the year's monthly check-in notes, the role description, and the two formal feedback documents. Produce a draft review with three named strengths with evidence from the materials, three named development areas with evidence, two recommended priorities for the next year, and a calibration note flagging anything that should be verified or softened. Do not infer judgments not supported by the materials. Do not recommend compensation actions; that decision is mine.

Post-meeting synthesis

Weak. Summarize the meeting.

Practitioner. Use the meeting transcript I provide. Produce a structured meeting brief with: the decisions made, the decisions deferred and the reason, the action items with owners and dates, the open questions that need follow-up, and a list of anything that was discussed but where I cannot tell from the transcript whether agreement was reached. Mark each section explicitly. Do not invent decisions or action items not in the transcript.

The interview, for any task where you are not sure what context matters

Weak. Help me with the leadership offsite agenda.

Practitioner. I need help designing the agenda for a two-day leadership offsite. Before you draft anything, ask me the questions you need answered to produce the best possible output. Ask them one batch at a time. When you have enough, restate my answers as a structured brief and confirm it with me before drafting.

Self-critique, as a standing instruction

Weak. Is this any good?

Practitioner. Before you show me the draft, review it against the following: every figure carries a source; quotation marks appear only around exact wording; anything you could not retrieve is flagged rather than filled in; no claim is stated with more confidence than its evidence supports.

Then sweep the prose for the marks of unedited AI writing: floods of em-dashes, the "not X, but Y" reversal used repeatedly, one-sentence paragraphs for false drama, bolded fragments scattered through the text. Show me the draft and a short list of what you changed.

Use-Case Gallery

Three named-company cases to illustrate the module's frameworks at executive scale. These are starting points; many other cases would do similar work.

Aflac — Supervised Digital Worker, Bounded by Design

Discussed in detail in Module 1; cross-referenced here for the Authority Ladder implication. Aflac placed AI at the Supervised Digital Worker tier for a specific, bounded category — repetitive low-value wellness claims under approximately \$200 — and explicitly preserved human judgment for higher-stakes complex cases. The case is the cleanest published example of the Ladder operating with disciplined boundaries.³¹

Lockheed Martin — All Five Levels, Each with Policy

Lockheed Martin's Navigator chat system, the Genesis agent-building platform, and the Astris AI subsidiary together represent one of the most extensive named AI deployments in the defense industrial base. By 2025, approximately 75,000 Lockheed Martin employees had access to Navigator, and more than 16,000 internal AI agents had been built on the Genesis platform since the start of that year, spanning analytical assistants, document workflows, supervised digital workers, and multi-agent processes.¹ The Astris AI subsidiary, launched December 16, 2024, externalizes the platform to the broader defense industrial base.²

The case illustrates the Authority Ladder operating at every level simultaneously, with formal governance structure attached to each transition. Mike Baylor, Chief Digital and AI Officer at Lockheed Martin, and Greg Forrest, Director of AI Foundations, are the program's public faces. The strategic point is that scaling AI is not a single deployment decision; it is a portfolio of decisions, each at a defined level of authority, each with a defined human approval gate.

CDC — Public-Sector AI in High-Assurance Conditions

The U.S. Centers for Disease Control and Prevention has been formalizing AI governance and deployment under the agency's *AI Strategy and Vision*, including the AI Accelerator Program (AIX) for piloting and operationalizing AI tools.³² *Nextgov* and *Federal Computer Week* reported in December 2025 that CDC's early bets on AI were paying off in disease-surveillance applications.

The case illustrates context engineering and data boundaries operating under regulated public-health conditions. The work-brief discipline is not a productivity nicety in this setting. It is the formalization of which AI work touches protected health information, which AI work crosses jurisdictional and federation boundaries with state, tribal, local, and territorial partners, and which AI work requires explicit approval gates before action. The discipline that an enterprise can choose to adopt, an agency at this scale must adopt.

Frameworks and Standards Reference

NIST AI Risk Management Framework (AI RMF 1.0) and the Generative AI Profile (NIST AI 600-1)

The NIST AI RMF 1.0, published January 2023, organizes AI risk management around four functions — Govern, Map, Measure, and Manage.³³ The Generative AI Profile (NIST AI 600-1, July 2024) layers a generative-AI-specific overlay on top, identifying twelve risk categories (including confabulation, dangerous or violent content, data privacy, environmental impacts, harmful bias, human-AI configuration, information integrity, information security, intellectual property, obscene or abusive content, value chain, and homogenization) and roughly 200 suggested actions.³⁴ For executives, the RMF and 600-1 together are the U.S. baseline vocabulary for AI risk; familiarity with the names of the four functions and the twelve generative-AI risk categories is enough to participate productively in the governance conversation.

OWASP Top 10 for LLM Applications (2025)

The OWASP project's 2025 list of the top 10 LLM application risks: LLM01 Prompt Injection (still the leading category by a substantial margin), LLM02 Sensitive Information Disclosure, LLM03 Supply Chain, LLM04 Data and Model Poisoning, LLM05 Improper Output Handling, LLM06 Excessive Agency, LLM07 System Prompt Leakage (new in 2025), LLM08 Vector and Embedding Weaknesses (new in 2025), LLM09 Misinformation, and LLM10 Unbounded Consumption.³⁵ LLM01, LLM02, LLM05, LLM06, and LLM09 bear directly on the executive's authority-ladder and data-boundary decisions.

ISO/IEC 42001:2023

ISO/IEC 42001 is the first international AI management-system standard, published in late 2023. It is certifiable, and several major enterprises — including Microsoft — have completed certification. By 2026, ISO/IEC 42001 certification has become a procurement-due-diligence checkpoint in regulated industries and large-enterprise vendor evaluations.³⁶

EU AI Act, Articles 13 and 14

Article 13 of EU Regulation 2024/1689 requires providers of high-risk AI systems to disclose human-oversight measures and the system's limitations, including the conditions under which the system is intended to be used. Article 14 requires deployers to ensure that natural persons can effectively oversee the operation of high-risk systems, and Article 14(4)(b) specifically obliges them to remain "aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system (so-called 'automation bias')." ²¹ The Article 14(4)(b) clause is the only place a regulator has codified the calibration-gap problem covered earlier in this module into binding obligation.

The enforcement timeline matters for planning, and it is no longer uniform. Prohibited-practice and AI-literacy obligations have applied since February 2, 2025, and obligations for general-purpose AI models applied from August 2, 2025, with a grace period to August 2027 for models already on the market. The high-risk dates then moved. The Digital Omnibus on AI — politically agreed May 7, 2026, endorsed by the European Parliament on June 16, 2026, and approved by the Council on June

29, 2026 — deferred stand-alone high-risk obligations (Annex III) from August 2, 2026 to December 2, 2027, and product-embedded high-risk obligations (Annex I) from August 2, 2027 to August 2, 2028. Article 50 transparency obligations remain on the original August 2, 2026 date. An organization mapping its obligations should confirm which schedule applies to which system rather than assume a single deadline governs the portfolio.²¹

Stanford HAI Foundation Model Transparency Index (FMTI)

The third edition of the Stanford HAI Foundation Model Transparency Index, published in December 2025, assessed leading foundation-model developers across one hundred indicators of transparency. The mean score across assessed developers fell from 58/100 in 2024 to 40/100 in 2025; some incumbents declined sharply (Meta from 60 to 31; Mistral from 55 to 18); new entrants from China and other regions (DeepSeek, Alibaba, xAI) were assessed for the first time.³⁷ A note of caution: the score decline reflects both an actual reduction in disclosure and a tightening of the FMTI methodology between editions; the trend is real, but its magnitude is partly a function of measurement change.

Coalition for Content Provenance and Authenticity (C2PA)

C2PA is an industry consortium developing cryptographic content-provenance standards. Content Credentials version 2.2 was released in May 2025; major camera manufacturers (Leica, Sony, Samsung, Nikon) have shipped C2PA-compatible products; CISA endorsed the standard in January 2025.³⁸ For executives, C2PA matters when AI is involved in producing external-facing content — marketing, communications, board materials — where provenance verification will increasingly be expected.

Model Cards and System Cards

Model cards (Mitchell and colleagues, 2019) and system cards (the format adopted by OpenAI, Anthropic, and Google for their major model releases) are the practical disclosure documents that accompany frontier model releases. Anthropic's Claude Opus 4.8 system card (May 2026) and OpenAI's GPT-5.5 system card (April 2026) are among the most detailed examples available. Reading one current system card is the single most efficient way for an executive to develop calibrated intuition for what these systems can and cannot do reliably.

Reference Panels

Short panels for executive look-up on adjacency topics that did not fit the Pre-Class Read spine.

Few-Shot Prompting

The technique with the highest ratio of value to effort for executive work, and the one most often skipped. Give the system two or three examples of what good output looks like before asking for the new one. Three prior recommendation letters and a CV. Last year's performance review and this year's notes. Three board memos you wrote and this quarter's facts. It is especially powerful for voice and format, where a house style took years to build and describing it is slower and less accurate than showing it.

The standing caveat applies. The output still gets read, and it still gets edited. The time saved is in the drafting, not in the judgment.

What to Feed the Model

File format is an economic decision, not a technical detail. A Word document or a PDF costs the system a conversion step before it can read anything, and conversion is lossy on tables, footnotes, and layout. Plain text, markdown, and HTML are read directly and cheaply. For a document the system will read once, the format rarely matters. For standing project knowledge that gets read on every conversation, convert once and keep the converted version. The executives who complain that the tool "keeps missing things in the deck" are frequently paying a conversion tax they do not know about.

AI Tells and AI Slop

Readers have learned to recognize unedited AI prose, and the marks now have a name. Floods of em-dashes. The reversal construction deployed several times a page. Filler vocabulary. One-sentence paragraphs for false drama. Bolded fragments scattered through prose. Emoji-studded checklists in documents that did not ask for them. None is wrong in isolation. In accumulation they read as a signature, and audiences discount accordingly.

The economics are what make this a leadership topic rather than a style note. A document that visibly took no human effort tells the reader they were not worth any, and it spends the sender's credibility to save the sender's time. Internally, the cost compounds: recipients learn to skim or ignore output from teams known to ship unedited generation, which quietly cancels the productivity gain the tools were supposed to deliver. Markets are efficient at repricing content produced at volume without ownership.

The fix is a discipline rather than a ban, and it has three parts. Edit like an owner: the standard for anything that leaves your hands is that you would defend every sentence as your own. Disclose the workflow where it matters. And systematize the sweep, either as part of your editing pass or as a standing self-critique instruction, so the system checks its own draft before you do. The Prompt Examples Library carries that instruction.

The Fiduciary Record

The enforcement and disclosure history behind the Pre-Class Read's claim that AI literacy has become a fiduciary matter. The Securities and Exchange Commission brought its first AI-washing actions against asset managers Delphia and Global Predictions in March 2024,⁷ its first public-company AI-washing action against Presto Automation in January 2025,⁸ and in April 2025 the SEC and the Department of Justice filed parallel actions against Albert Saniger, founder and former chief executive of Nate, Inc., whose shopping app was allegedly powered by offshore contract workers rather than by AI.⁹ The SEC's Cyber and Emerging Technologies Unit, created in February 2025, has named AI a continuing examination priority. The Conference Board reports that the share of S&P 500 companies disclosing at least one material AI risk in their annual reports rose from 12% in 2023 to 83% by the end of 2025; EY's separate review of the Fortune 100 found more than 85% addressing AI in their risk factors and 36% carrying a standalone AI risk factor in their 10-K, more than double the prior-year share.¹⁰ Securities class actions targeting AI misrepresentations

roughly doubled from 2023 to 2024 and held that pace through 2025, with Cornerstone Research counting 16 AI-related filings for the year.¹¹ Module 4 governs the response.

Structured-Output Reliability

JSON-mode and structured-output features fail to parse 5% to 20% of the time at scale when used without strict-schema enforcement, and parse failures still run 2% to 5% even with vendor-specific JSON modes. Strict-schema enforcement (such as function-calling or tool-use schemas) reduces structural failure but does not prevent semantically wrong fields inside structurally valid responses.
17

LLM-as-Judge Limitations

Using an LLM to evaluate another LLM's output (the "LLM-as-judge" pattern) has documented biases — position bias (preference for the first or second answer), verbosity bias (preference for longer answers), self-enhancement bias (preference for outputs from the same model family), and weak alignment with human judgment outside the largest models.³⁹ An executive evaluating a vendor pitch that includes "we use an LLM judge to score our agent's quality" should ask which judge, against what reference set, with what disclosed disagreement rate against human reviewers.

Board-Level AI Oversight

McKinsey's 2025 *State of AI* survey reports that only 17% of organizations say AI governance is overseen by their board, and 28% say it is overseen by the CEO.⁴ At the Workflow Agent tier and above, the question of who owns the outcome when the system fails becomes a fiduciary question. Board oversight is not the same as a board AI committee; many of the organizations with the strongest governance posture have integrated AI into existing risk, audit, and compensation committee work rather than creating a parallel structure.

Model Routing as Cost Discipline

Production AI deployments increasingly route requests across multiple models — economy-tier models for routine tasks (roughly \$0.25 to \$4 per million tokens) and frontier models for complex reasoning (roughly \$15 to \$75 per million tokens). Documented cost reductions from disciplined routing are typically in the 30% to 70% range.⁴⁰ The executive question is whether your organization has a routing policy at all, and whether it is reviewed as part of operating-cost discipline.

Ambient AI in Meetings

Tools that transcribe, summarize, or recall meetings — Otter, Zoom AI Companion, Microsoft Copilot Chat, Microsoft Recall — generate persistent records of conversations that may be subject to discovery, regulatory examination, or future litigation. A January 2025 Microsoft Copilot Chat vulnerability processed sensitivity-labeled emails despite data-loss-prevention rules; Microsoft's Recall product was repeatedly delayed through 2024 and 2025 over privacy concerns.⁴¹ The executive lesson is that an AI-recorded meeting is now a discoverable record, and privilege, attorney-client, and M&A confidentiality issues attach accordingly.

Copyright and IP in AI Work Products

Major active matters as of May 2026 include *Disney and Universal v. Midjourney* (filed June 2025); *Warner Bros. Discovery v. Midjourney* (filed September 2025); *The New York Times v. OpenAI and Microsoft* (motion to dismiss largely denied in March 2025; summary-judgment proceedings pending); *Bartz v. Anthropic* (settled for approximately \$1.5 billion in 2025, the largest copyright settlement in U.S. history); the UK High Court ruling in *Getty v. Stability AI* (November 2025) holding that model weights are not "infringing copies" under UK copyright law, with Getty granted permission to appeal in December 2025; and the Munich Regional Court ruling in *GEMA v. OpenAI* (November 2025) holding that training on copyrighted lyrics violates German copyright.⁴² The input-side risk (training on copyrighted material) and the output-side risk (producing material that reproduces protected expression) are legally distinct. Executive use of AI for memo drafting, marketing content, or internal communications is primarily an output-side risk: the question is whether the output substantially reproduces protected expression. The answer is jurisdictionally uneven.

Attorney-Client Privilege When Using AI

ABA Formal Opinion 512 (July 2024) holds that informed client consent is required before using client information in self-learning AI tools, and that boilerplate engagement-letter consent is insufficient.⁴³ In a 2025 case of first impression, *United States v. Heppner* (S.D.N.Y., Rakoff, J.), the court held that an attorney's exchanges with a consumer AI tool were not protected by attorney-client privilege or the work-product doctrine.⁴⁴ The practical implication for executives is that pasting privileged or sensitive material into a consumer AI tool may waive protections that would otherwise apply; enterprise-grade AI agreements with appropriate confidentiality terms are not the same as consumer products, but the distinction must be made deliberately, not assumed.

Putting the Three-Layer Model to Work in Your AI Tool

The Field Reference above carries the conceptual scaffolding of the Three-Layer Model. This appendix carries the operational counterpart — what each layer looks like when you actually configure a project in the AI tool you use at work, and how to translate the same canonical setup across tools without one walkthrough per vendor.

The translation discipline is deliberate. AI tool UIs change quarterly; a static walkthrough of Claude Projects, ChatGPT Custom GPTs, Gemini Gems, Microsoft 365 Copilot Agents, Salesforce Agentforce, Glean, Writer, or your firm's internal LLM gateway is stale before the term ends. The canonical tutorial below ages on the Three-Layer Model itself, which is stable. The translation prompt that follows lets the AI tool you are configuring do its own translation — and lets you exercise the AI Work Brief discipline on the act of setting up the project. The meta-move is the lesson. You are about to use the Three-Layer Model on itself.

Tool defaults vary substantially across vendors, tiers, regions, and enterprise contracts — particularly on training, retention, and data residency. The canonical tutorial names the executive controls that matter; current values require verification against vendor documentation.

The Canonical Tutorial

The Three-Layer Model has three layers. Setting up a project in any AI tool is the work of placing AI Work Brief content into the right layer.

The Prompt Layer

The prompt layer is the persistent role configuration. It is the standing contract that governs what the AI is being asked to be, every time you open the project. It does not change task-by-task.

Place these AI Work Brief fields here: **Role** (the perspective the system should adopt — e.g., "you are a board pre-read editor for a mid-market industrial company"), **Audience** (who reads the output — board, executive committee, working team), **Constraints** (what the system should never do — invent citations, recommend compensation actions, advise on legal questions, exceed an output length), and **Review** (the criteria a human will use to judge quality — Context Quality Test, factual sourcing, structured output).

Do not place these fields in the prompt layer: the **Decision** the work supports (different per task), the **Task** itself (different per task), the **Data boundary** (specific to the data you are about to share). Those belong in the per-conversation layer.

The most common executive mistake at the prompt layer is overloading. The prompt layer should carry the standing contract — three to seven sentences is usually enough — not every instruction the system might ever need. A persistent role that has grown to two pages by accumulation is a project hygiene problem; some of those instructions belong in context, some belong in per-conversation prompts, and some have aged out of usefulness. Module 7 covers the hygiene routine.

The Context Layer

The context layer is what the AI tool can read while doing the work. In practical terms, it is some combination of attached files, knowledge bases, retrieval against a connected document store, project knowledge, and source documents pasted into the conversation. It is the briefing packet.

Place these AI Work Brief fields here: **Context** (the background the model needs to do this work — policy documents, prior board materials, financial pack, account history, contract terms, regulatory baseline), and the **provenance markers** that let the reviewer tell where the model's facts came from.

The Context Quality Test from earlier in the Field Reference is an operational checklist at this layer. Apply it explicitly when configuring the context layer:

- **Relevant.** Each context document materially helps the task. Documents added "in case it's useful" are usually noise.
- **Sufficient.** A capable human reviewer would have everything they need. Missing context is the most common reason a model produces plausible-but-thin output.
- **Isolated.** Distracting, contradictory, stale, or restricted material has been excluded. A context layer that mixes the current and the superseded version of the same policy will produce blended answers.
- **Economical.** The smallest packet that solves the task. Bloat is not free — it raises cost, increases latency, and degrades retrieval quality. Context rot is documented; longer is not better.¹²

- **Provenance-aware.** The reviewer can tell where each fact came from. Mark which sources are authoritative and which are supporting.

Three operational decisions sit at this layer that the conceptual model in the body section did not surface:

- **Attach vs. retrieve.** Most enterprise AI tools support both attaching a small set of documents directly to the project and connecting to a larger document store the AI retrieves from. Attach for documents the AI must read every time. Retrieve for larger corpora where relevance varies by task.
- **Freshness ownership.** Someone owns whether the context is current. A project anchored to a 2024 regulatory baseline is producing 2024 answers regardless of how well the harness is built. Name the freshness owner before the project goes into recurring use.
- **Permissions.** The context the AI can read should match the permissions of the user the AI is acting on behalf of. If your tool's retrieval is permission-aware (it pulls only documents the user could open directly), this is automatic. If not, the executive must hand-curate the attached set, because the AI does not know what the user is not allowed to see.

The Data Sensitivity Tiers earlier in the Field Reference apply at this layer. Sanitize at the source. Prompt-level instructions to "keep this confidential" are not the same as data-architecture controls.

The Harness Layer

The harness layer is the operating scaffolding around the prompt and the context — the tools the AI can use, the approval gates, the logs, the evaluation routines, and the points where a human re-enters. Module 3 owns the harness layer in depth. For the executive setting up a project at the Module 2 level, the harness layer involves a smaller set of decisions.

Place these AI Work Brief fields here: **Authority Ladder placement** (Assistant, Copilot, Workflow Agent, Supervised Digital Worker, Multi-Agent Process), and the **re-entry conditions** (low confidence, sensitive data, irreversible action, cross-system action, threshold breach).

Three operational decisions sit at this layer:

- **Tools.** What the AI can do beyond producing text. Most consumer-grade configurations restrict this. Most enterprise configurations permit some combination of web search, code execution, document creation, calendar access, email drafting, CRM read or write, or connection to internal systems. The decision of which tools the AI may use is the most consequential harness-layer decision because it is the difference between answer risk and action risk.
- **Memory and persistence.** Whether the AI carries information across conversations, and how widely. Memory that surfaces a prior conversation's context inside an unrelated current task is a context-isolation failure waiting to happen. Memory disabled by default; memory enabled deliberately for projects that genuinely benefit.
- **Logging and review.** Whether the project's interactions are logged for later review by the executive, the team, or the governance function. For workflows that meet Module 4's risk-tier criteria, logging is a control, not a convenience.

Five-Step Build Procedure

Across all three layers, the build procedure is the same:

1. Create the project in your tool.
2. Paste the **Role**, **Audience**, **Constraints**, and **Review criteria** from your AI Delegation Card into the persistent configuration (the prompt layer).
3. Attach or connect the **Context** sources from your Delegation Card. Apply the Context Quality Test before saving.
4. Configure the **harness** — Authority Ladder placement, tool permissions, memory, and logging — to match the level the Delegation Card specifies.
5. Run a representative task. Compare the output to the Delegation Card's success criteria. Iterate the project setup, not the task.

How Professionals Organize AI Work

The people who get consistent results from these tools converge on a recognizable working architecture, and it costs nothing to copy.

One project per job. A separate workspace for each distinct role the person plays, so the contexts never blur. A professor's set might include an executive-MBA teaching project, a consulting project, a whitepaper-writing project, and an adversarial-review project whose entire job is to argue against a draft before it ships. That last one matters more than it looks: the system's default posture is agreeable, and the executive needs one place where it is not.

Each project carries a small set of curated knowledge files: a voice and style guide, the canonical current version of each key document, and a few exemplars of what good output looks like. Canonical documents are versioned, with a version log, so the system is never working from three conflicting drafts of the same thing. Each recurring workflow has one owned master, not a hundred forks. And finished deliverables stay out of the project's knowledge files, so retrieval stays clean. The system should draw on the current canon, not on every artifact ever produced.

None of this is technical. All of it is the discipline you would apply to a well-run shared drive, applied to a system that actually reads the files.

The Translation Prompt

Paste this prompt into the AI tool you are configuring. It will translate the canonical tutorial above into step-by-step instructions specific to that tool. The prompt is itself an AI Work Brief — its structure should look familiar.

``` ROLE — You are a translator from a canonical AI-tool-setup tutorial to a tool-specific walkthrough. The user will name their AI tool. You will produce step-by-step instructions for that tool.

CONTEXT — The canonical tutorial below covers the Three-Layer Model (prompt layer, context layer, harness layer) and a five-step procedure for setting up an AI Delegation Card project. The user has read it.

[Paste the canonical tutorial above]

**TASK** — Translate the canonical tutorial into step-by-step instructions for the following tool: [name your tool]. For each instruction, name the specific UI element, menu, configuration field, or feature in this tool where the action happens.

**OUTPUT FORMAT** — Produce a numbered list matching the canonical tutorial's structure. Group your instructions under the same three layer headings (Prompt Layer, Context Layer, Harness Layer). End with the five-step build procedure adapted to this tool.

**EVIDENCE** — For each instruction, indicate whether you are confident the feature exists and behaves as described, or whether it is uncertain. If your training cutoff predates a recent UI change, say so. Recommend that the user verify data residency, training, and retention defaults against current vendor documentation — defaults vary by tier, region, and contract and your information may not be current.

**CONSTRAINTS** — Do not recommend this tool over others. Do not invent features. If a step in the canonical tutorial does not map cleanly to this tool, explain the difference rather than forcing a fit. Do not advise on the user's enterprise data classification, privacy posture, or governance configuration — those decisions belong to the user's IT and legal teams.

**REVIEW** — The user will check your output against the Context Quality Test from Module 2 — Relevant, Sufficient, Isolated, Economical, Provenance-aware. Anticipate that test as you respond.  
...

The annotation worth holding onto: every field in this prompt is a field from the AI Work Brief. The prompt is a working example of the discipline you are applying. If you find the prompt persuasive — if it produces useful translation output — that should sharpen your sense of what the discipline is for.

### **Worked Example: A Translation for Claude Projects**

What the prompt produces, given "Claude Projects" as the named tool, looks roughly like this. The example is calibration, not canon — your own translation will be more current than this one because it will use whatever the tool's UI looks like the day you run it.

**Prompt Layer.** In Claude, the persistent role lives in the **Project Instructions** field, accessed by creating a new Project from the side menu and clicking "Set instructions." Paste your Role, Audience, Constraints, and Review criteria here. Project Instructions are applied to every conversation in the Project automatically.

**Context Layer.** Claude Projects support both attached **Project Knowledge** files (uploaded directly; readable across all conversations in the Project) and per-conversation file uploads (readable only in that conversation). Attach your standing context — policies, templates, source documents — to Project Knowledge. Use per-conversation uploads for documents that change task-by-task. Apply the Context Quality Test before saving Project Knowledge.

**Harness Layer.** Claude's tool access depends on the tier. Consumer Claude offers limited tools (web search, file creation in some configurations); Claude for Work and Enterprise tiers expose connectors, MCP integrations, and computer use under admin configuration. Memory is opt-in and operates per-Project. Logging beyond conversation history typically requires Enterprise admin configuration.

*Verification flagged.* Default behavior for training on conversation content varies between consumer Claude and Team / Enterprise tiers; verify against current Anthropic documentation before pasting confidential material.

Two things to notice. First, the translation respects the canonical tutorial's three-layer structure rather than reorganizing around the tool's own menu hierarchy. Second, the verification flag at the bottom is the AI doing what the translation prompt asked it to do — explicitly hedging where its training cutoff or your tier-specific configuration may diverge from current reality. That hedge is a feature, not a defect of the output.

## Verification Discipline: Where Translation Is Least Reliable

The translation prompt asks the AI to flag uncertainty. It will not always do so reliably. Four areas warrant your own verification, regardless of what the translation says:

- **Data residency, training defaults, retention.** Vary by tool, tier, region, and contract. The AI rarely has current information on enterprise-tier defaults. Verify against vendor documentation and your own contract.
- **Enterprise admin and tenant settings.** Configured by your IT team, not by you in the tool's user-facing UI. The translation may describe features your tenant does not enable, or omit features your tenant has activated. Confirm with IT before assuming.
- **Recently shipped features.** Anything more recent than the AI's training cutoff is unreliable. Ask the tool what its training cutoff is before trusting the translation on new functionality.
- **Tool-specific governance integrations.** DLP rules, audit logging, sensitivity-label handling, agent identity, and permission-aware retrieval are configured at the enterprise layer. Confirm with the team that owns governance, not via translation.

The translation discipline does not replace these conversations. It produces a starting point that should make the conversations with IT, legal, and security shorter, because you arrive with a structured request rather than a vague one.

## Try This Before Class — Extended

The "Try This Before Class" prompt in the body of Module 2 asked you to write a request first as a one-line prompt, then as an AI Work Brief, run both, and compare. The extended version, for students with access to enterprise AI tooling:

1. Read the canonical tutorial above.
2. Run the translation prompt for the AI tool you use most at work.
3. Configure a Project — or whatever the tool calls its persistent workspace — for one recurring workflow from your AI Delegation Card workbook prompt.
4. Run one task through the configured Project.
5. Bring three things to class: the translation output, your Project configuration, and the task output.

The cohort debrief works best when the room has run translations for several different tools against the same canonical tutorial. Bring what you produced and be ready to explain where the translation was useful, where it required correction, and where the canonical tutorial itself failed to anticipate something the tool exposed.

---

## Part 3: Card Specification

---

### The AI Delegation Card

---

The AI Delegation Card is drafted in session for a real recurring executive workflow — either from the student's own organization or from a chosen company — using the template in the Field Reference. A complete card is practical enough that a manager could use it as the starting point for a responsible pilot conversation: a specific workflow, a plausible authority level, and a well-defined human approval gate with re-entry conditions. There is no post-class submission.

### Capstone Build and Next-Module Handoff

---

For students on the Blog 3 Transform path, the AI Delegation Card feeds directly into the capstone artifact kit. The card's workflow specification, authority level, human-review point, data-boundary language, and evidence-standard language together form the prioritized-use-case section of the final memo.

Module 3 picks up the harness layer: what technology stack, trusted-context architecture, model routing, tool access, observability, and sourcing choices are required to make the workflow specified on the Delegation Card actually dependable in production.

### Works Cited and Sources

---

1. Aerospace America, "Lockheed's AI Factory and the Externalization of Defense AI," 2026; lockheedmartin.com program documentation, 2025–2026.
2. Lockheed Martin, "Lockheed Martin Launches Astris AI to Drive AI Adoption Across Defense Industrial Base," press release, December 16, 2024; astrisai.com.
3. Stanford Institute for Human-Centered AI, *2026 AI Index Report*, Stanford University, April 2026.
4. McKinsey & Company, *The State of AI in 2025: Agents, Innovation, and Transformation*, November 2025.
5. Tobi Lütke, internal memo to Shopify employees, "Reflexive AI usage is now a baseline expectation," subsequently published by Lütke on X, April 7, 2025; coverage in *Business Today*, April 8, 2025.
6. Aaron Levie, remarks at TechCrunch Disrupt 2025, October 29, 2025.
7. U.S. Securities and Exchange Commission, "SEC Charges Two Investment Advisers with Making False and Misleading Statements About Their Use of Artificial Intelligence," press release 2024-36, March 18, 2024.
8. U.S. Securities and Exchange Commission, *In the Matter of Presto Automation, Inc.*, Administrative Proceeding File No. 3-22413, Release No. 33-11352, January 14, 2025. <https://www.sec.gov/files/litigation/admin/2025/33-11352.pdf>
9. U.S. Securities and Exchange Commission and U.S. Department of Justice, parallel actions against Albert Saniger (Nate, Inc.), April 9, 2025; Holland & Knight, "SEC and DOJ Warm Up to Enforcement over AI Washing," July 2025.
10. The Conference Board / ESGAUCE, S&P 500 AI risk-factor disclosure analyses, October 2025 and April 22, 2026 (full-year 2025 update: 12% in 2023 to 83% in 2025); EY Center for Board Matters, "Cyber and AI Oversight Disclosures," October 2025.

11. Cornerstone Research, *Securities Class Action Filings: 2024 Year in Review*, 2025, and *2025 Year in Review*, January 2026 (16 AI-related filings in 2025); New York State Bar Association, "Regulating AI Deception in Financial Markets," 2025.
12. Hong, Troynikov, and Huber (Chroma Research), "Context Rot: How Increasing Input Tokens Impacts LLM Performance," July 2025.
13. Microsoft and LinkedIn, *2024 Work Trend Index Annual Report: AI at Work Is Here. Now Comes the Hard Part*, May 2024.
14. IBM Security and Ponemon Institute, *Cost of a Data Breach Report 2025*, July 2025.
15. Harmonic Security, *2025 Sensitive Data in AI Tools Analysis*, 2025.
16. *Bloomberg News*, "Samsung Bans Staff's AI Use After Spotting ChatGPT Data Leak," May 2, 2023.
17. TokenMix, "Structured Output and JSON Mode Guide 2026"; TianPan, "Beyond JSON Mode: Getting Reliable Structured Outputs from LLMs in Production," October 29, 2025.
18. Lee, Tankelevitch, Hofman, and colleagues (Microsoft Research and Carnegie Mellon University), "The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects from a Survey of Knowledge Workers," CHI 2025, February 2025.
19. Kos'myna and colleagues (MIT Media Lab), "Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for Essay Writing Task," arXiv preprint 2506.08872, June 2025.
20. "Exploring Automation Bias in Human-AI Collaboration: A Review and Implications for Explainable AI," *AI & Society* (Springer), 2025.
21. European Union, Regulation (EU) 2024/1689 (Artificial Intelligence Act), Articles 13, 14, and 50, Official Journal of the European Union, July 12, 2024; European Commission, AI Act application timeline ([digital-strategy.ec.europa.eu](https://digital-strategy.ec.europa.eu/en/policies/artificial-intelligence-act)); European Commission, Digital Omnibus on AI, proposed November 2025 — political agreement May 7, 2026; European Parliament endorsement June 16, 2026; Council approval June 29, 2026.
22. Gartner, "Predicts 2025: Building Trust in AI Requires Governance, Risk Management, and Investment," 2025.
23. Truist Financial Corporation, "Truist Launches AI-Enabled Receivables Platform," PR Newswire, February 3, 2026; *American Banker*, "How Truist and Western Alliance Deploy AI to Fight Fraud," June 4, 2025.
24. *Consumer Goods Technology*, "How Bic, Kimberly-Clark Use AI to Democratize Data," October 2025.
25. Damien Charlotin, AI Hallucination Cases Database, 2025; *CalMatters* reporting on *Noland v. Land of the Free*, September 2025.
26. *FindLaw*, "DOJ Attorney's AI-Generated Brief Sparks Sanctions Threat After Pro Se Plaintiff Uncovers Fabricated Quotes," 2025.
27. *Moffatt v. Air Canada*, 2024 BCCRT 149, British Columbia Civil Resolution Tribunal, February 14, 2024.
28. U.S. National Vulnerability Database, CVE-2025-32711 (EchoLeak), 2025; Aim Labs disclosure, June 2025; Reddy and colleagues, "EchoLeak: The First Real-World Zero-Click Prompt Injection Exploit in a Production LLM System," arXiv preprint 2509.10540, September 2025.
29. *Bloomberg News*, "Klarna Turns From AI to Real-Person Customer Service," May 8, 2025; CX Dive, coverage of Klarna Q3 2025 earnings, November 2025.
30. Australian Department of Employment and Workplace Relations report on the Targeted Compliance Framework, 2025; *The Register* and *Sydney Morning Herald* coverage, October 6–7, 2025.
31. MIT Sloan Management Review, "Continuous Learning with AI: Aflac's Shelia Anderson," *Me, Myself, and AI* podcast, 2023; *CMSWire*, "Human First: How Aflac Combines AI With Authentic Connections," 2025.
32. Centers for Disease Control and Prevention, *AI Strategy and Vision*, [cdc.gov/ai](https://cdc.gov/ai), 2025–2026 updates; *Nextgov / Federal Computer Week*, "The CDC Placed Early Bets on AI...," December 2025.
33. U.S. National Institute of Standards and Technology, *AI Risk Management Framework (AI RMF 1.0)*, January 2023.

34. U.S. National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile* (NIST AI 600-1), July 26, 2024.
35. OWASP Foundation, *OWASP Top 10 for Large Language Model Applications, 2025 Edition*, November 2024.
36. International Organization for Standardization, *ISO/IEC 42001:2023 Artificial Intelligence — Management System*, December 2023.
37. Bommasani, Klyman, Longpre, and colleagues (Stanford CRFM), *The 2025 Foundation Model Transparency Index*, December 2025.
38. Coalition for Content Provenance and Authenticity, *Content Credentials Specification, v2.2*, May 2025.
39. Eugene Yan, "Evaluating the Effectiveness of LLM-Evaluators," eugeneyan.com, 2024-2025; Zheng and colleagues, "Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena," NeurIPS 2023.
40. Clarifai, "AI Cost Controls: Budgets, Throttling, and Model Tiering," 2026; MindStudio, "Best AI Model Routers for Multi-Provider LLM Cost Optimization," 2026.
41. GeekWire, "One Year After Its Rocky Launch, Microsoft's Windows Recall Still Raises Security Red Flags," 2026; *CX Today*, "Microsoft Copilot Bug Raises CX Email Security Concerns," 2025.
42. AI Lawsuit Tracker, [ailawsuittracker.com](https://ailawsuittracker.com), 2026; *Bartz v. Anthropic* settlement documents (\$1.5B; preliminary approval September 25, 2025; final fairness hearing May 14, 2026); *Getty Images v. Stability AI Limited* [2025] EWHC 2863 (Ch), November 4, 2025 (permission to appeal granted December 16, 2025); *GEMA v. OpenAI*, LG München I 42 O 14139/24, November 11, 2025 (on appeal); *Thaler v. Perlmutter*, certiorari denied March 2, 2026 (human-authorship requirement left intact).
43. American Bar Association, Standing Committee on Ethics and Professional Responsibility, Formal Opinion 512: Generative Artificial Intelligence Tools, July 2024.
44. *United States v. Heppner*, U.S. District Court for the Southern District of New York, 2025 (Rakoff, J.); analysis in *Harvard Law Review* blog, March 2026.
45. Amazon Web Services, *Georgia-Pacific Optimizes Operator Efficiency Using Generative AI on AWS*, AWS Case Study, 2024; *Manufacturing Digital*, "KONE & Georgia-Pacific, elevators & paper, gen AI & results," 2024.
46. Sherry Ruan, Jacob O. Wobbrock, Kenny Liou, Andrew Ng, and James A. Landay, "Comparing Speech and Keyboard Text Entry for Short Messages in Two Languages on Touchscreen Phones," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2018 (reported by Stanford in 2016 as "Speech Is 3x Faster than Typing"). English speech input was 3.0 times faster than keyboard typing with a 20.4 percent lower error rate. The comparison is against a phone keyboard; the margin against a full-size keyboard is smaller, but the direction holds.
47. Mingqian Zheng, Jiaxin Pei, Lajanugen Logeswaran, Moontae Lee, and David Jurgens, "When 'A Helpful Assistant' Is Not Really Helpful: Personas in System Prompts Do Not Improve Performances of Large Language Models," *Findings of the Association for Computational Linguistics: EMNLP 2024* (arXiv:2311.10054). The survey literature reaches the same conclusion for objective tasks: Sander Schulhoff et al., "The Prompt Report: A Systematic Survey of Prompting Techniques," arXiv:2406.06608, 2024. A smaller set of studies reports gains from role-play prompting on selected reasoning benchmarks; the accuracy evidence is mixed at best, while the register effect is consistent.

## 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 2: Executive AI Literacy*. 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](mailto:jerry@profkane.com)