PRINCIPLE

Fast for the parts, *right for the person.*

SCOPE

An AI ***agent*** integration strategy for executive IT hardware provisioning in aerospace and defense.

SCENARIO

8:52 A.M.

The message lands: the board meeting starts in eight minutes, the executive's hearing aid battery is nearly depleted, and the charger remains on his kitchen counter.

9:01 A.M.

The appropriate charger arrives discreetly, meeting momentum carries on uninterrupted, and the entire process remains fully confidential and seamless.

CONTENTS

PREPARED BY	FUNCTION	ISSUE	PAGE	DRAWING SET
Zach Geller	Executive IT Support · N. Texas Principal Lead	Sep 2026	01 of 06	ZG·CAP·8.1

Executive summary

I am the North Texas principal lead for executive IT support at an aerospace and defense company. Much of the work is getting the right peripherals into people's hands before they travel; those requests reach me personally. It takes one to four business days from the ask to hardware moving.

I am proposing an agent built on Microsoft 365 Copilot that takes intake and fulfillment for non-controlled commodity gear, hands a request to a person on any of five triggers, and never guesses. The rollout is Crawl, Walk, and Run, with dates on Crawl. Three measures cover it: time to serve for operations, containment for adoption, and a hard zero on silent failures for quality and compliance, with the money carried as avoided trip failures. However, all of it rests on the people who text me today choosing the agent instead.

01 Organizational context and problem statement

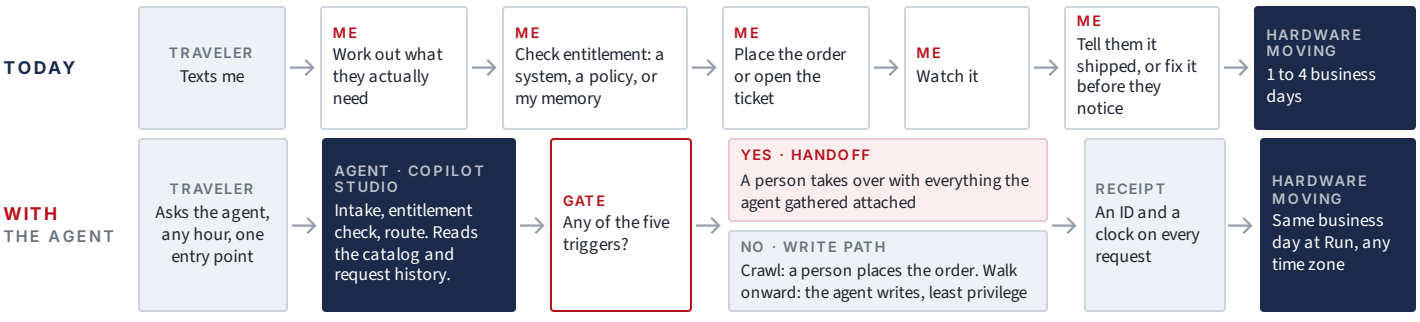
I work in aerospace and defense, in an executive IT support team that sits inside business systems and transformation. This plan concerns one process. The hardware provisioning loop runs from the moment somebody asks for a piece of gear to the moment that gear is in their hands.

Today it is me. Figure 01 sets it beside the version with the agent in it. Somebody texts, usually the day before they fly. I work out what they need, since it is often not what they asked for. Then I check entitlement: part of the answer is in a system, part in policy, and part in my memory. I place the order or open the ticket, watch it move, and tell them when it ships. When something goes sideways I fix it. Most of the time they never hear about it.

That last part matters. When the executive on the cover sends his message about a charger, what he is asking for is ten quiet minutes to read the deck while somebody handles the problem behind him. Nobody else in the room needs to know anything happened. That discretion is the real product. The charger is simply what makes it possible. People in these roles text the person they trust because they want the problem gone, so whatever replaces that text has to preserve the same feeling or it will sit unused.

The volume is 10 to 100 requests a week by text and walk-up. More arrive in ways nobody logs. Time to serve runs one to four business days. Working harder does not fix it: every step passes through one person's availability. When I travel, requests wait on a time zone. When the population grows, the queue grows with it. My own hours are the only lever left. However, none of the work is difficult. It simply depends on whether I am reachable. That kind of dependency does not divide.

FIG 01 THE LOOP *The same request both ways.*



Catalog · asset · procurement. Governed connectors, every record tagged agent-originated, **nightly reconciliation**.

02 AI use case, technology choice, and data readiness

Why an agent and not a form? A form fails the same way the current loop fails. The request is ambiguous, the entitlement answer is spread across systems, policy, and memory, and the action lands in three systems of record that each expect a person clicking through a screen. A workflow only automates a path somebody has already mapped. This job needs something that can take an unclear request at nine at night, work out what the person needs, assemble the entitlement answer from wherever it lives, act in the systems of record, and know when to hand off with everything attached. Language models settled the understanding part. The rest is what makes it an agent.

FIG 02 HUMAN IN THE LOOP *The five triggers that hand a request to a person.*

1 MORE THAN ONE
Two or more items, or a full kit.

2 OVER THE LINE
Estimated value above roughly $500 cumulative over 30 days so an order cannot be split to stay under it. A person signs.

3 TOO FAST
Several requests from one person in a short window. Something is wrong or something is urgent. Both need a person.

4 ADDRESS MISMATCH
Ship-to does not match what the backend holds. Catching a wrong address before it ships costs far less than after.

5 NOT STANDARD
Custom, non-catalog, or anything the rules do not cover. Outside the catalog it is a judgment call.

THE HANDOFF A trigger does not reject a request. It goes to a person **with everything attached. The agent never refuses when a human could say yes.**

The platform is Microsoft 365 Copilot: governed, seat-licensed, already in the stack, and inside the tenant boundary with our data. I looked at a self-hosted open-weight model first. It costs less per token and gives more control over behavior, but it also hands me the security posture, the patching, and the model lifecycle. In this sector that is a second job, so the tenant boundary decides it for the closed-source platform.

The architecture starts as a single agent at Crawl and becomes three at Walk. One assistant takes the request, checks entitlement, and routes it. At Walk it splits into separate intake, entitlement, and fulfillment agents, once there is enough real traffic to show where the seams belong and where the gaps show.

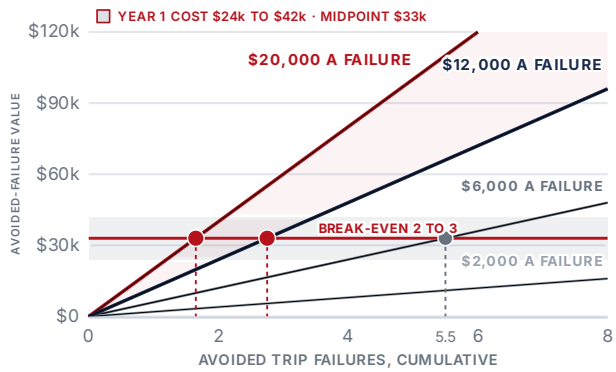
On integration, the design assumes orchestration in Copilot Studio, governed connectors into the three systems of record for catalog, asset, and procurement, and a Model Context Protocol interface where needed. That is a proposal for what we would build. The agent holds no credential a person would not hold. The final middleware choice belongs to our platform and security teams.

Three data feeds have to be ready before Crawl opens. The catalog needs current part numbers and stock status. The request history needs cleaning, since it doubles as context for the agent and as the replay set for testing. The entitlement rules need writing down. That is the hard one: a real share of them exist today only as things I know.

03 Cost considerations

COST CATEGORY	ESTIMATE AND BASIS
CapEx · build, licensing, infrastructure	$18k to $32k. 120 to 200 internal build hours at our internal hourly rate, plus platform and security review. No new licenses. Estimate from comparable internal builds.
OpEx · subscription, compute, maintenance	Nothing extra in Year 1. Copilot seats are already licensed for everybody in scope. The model usage comes with the seat.
Staffing and training	$6k to $10k. Roughly 60 hours of delivery and materials across two tiers, reusing the lunch-and-learn format already running.
Year 1 total	**$24k to $42k. Almost entirely internal labor. Midpoint near $33k. The gear itself is not in this number.**
Year 2 onward, each year	**$6k to $12k a year for maintenance, monitoring, and governance time. No new licensing.**
What pays it back	**Avoided trip failures. Break-even at roughly 2 to 3 avoided failures against the Year 1 midpoint.**

FIG 03 PAYBACK *Avoided-failure value against Year 1 cost.*



The cost this project prevents is a failed trip. I have priced one at roughly $12,000 to $20,000: the airfare and lodging spent, the executive's hourly rate across the trip hours, and a cost for the meeting that did not happen. The last term does the damage. Against a Year 1 midpoint near $33,000, the build pays for itself somewhere between two and three avoided failures, as Figure 03 shows.

The platform is seat-priced, so there is no per-call bill. If it were metered, a typical request runs about 2,000 input tokens against 400 output: roughly a penny a transaction at representative enterprise rates and five to fifty dollars a year at our volume. Token pricing is not the constraint at this size.

04 Security plan

Aerospace and defense sets the starting condition: the agent never touches export-controlled technical data. The scope fence in Identify is how that holds. Above it the plan runs on NIST CSF 2.0, the AI Risk Management Framework for model-specific risk, and the OWASP Top 10 for LLM Applications for the threats. Figure 04 shows the controls across the six functions. Most sit in Identify and Protect. Section 07's standing group is Govern.

FIG 04 SECURITY *The controls across the six NIST CSF 2.0 functions.*

GOVERN	IDENTIFY	PROTECT	DETECT	RESPOND	RECOVER
IT, security, procurement, and me as a standing group.	Four internal data types. Scope fence to non-controlled gear, classification check at intake.	Least privilege on the write path. No standing credential. Five triggers as approval gates.	Receipt with an ID and a clock on every request. Nightly reconciliation against downstream.	I triage operational errors. Data, access, or controlled-item events go to the incident process.	No new system of record. If the agent is pulled, requests fall back to the manual path.

IDENTIFY covers the four things the agent reads: the commodity catalog, request history, the entitlement rules, and employee shipping addresses. All of it is internal business data already inside the governed Microsoft 365 tenant. No new data store gets created. A classification check at intake enforces the fence, so it never depends on an assumption.

One thing that fence does not cover is the medical device named in the request on the cover. People travel with them. The agent uses that detail to find the part, then a redaction step strips it before anything reaches the audit log. That is the control for sensitive information disclosure, OWASP LLM02. The record shows what happened without showing what it was for. That protects the person's privacy and their dignity. The log still keeps the request ID, the timestamps, the trigger that fired, the catalog class, and the rule that decided, so a GDPR explanation request can be answered without the medical detail ever entering the record.

PROTECT is least privilege on the write path, with no credential the agent holds that a person would not. The five triggers in Figure 02 act as approval gates. That is the control for excessive agency, OWASP LLM06. The agent acts on its own only when all five are clear. The same design limits prompt injection, LLM01: the agent only writes catalog selections downstream and never free text, so hostile instructions buried in a request can at worst trigger an escalation.

FIG 05 AI-SPECIFIC THREATS *Two failures and their controls.*

Threat 01 · an answer that is not true reaches the record

LOOKS LIKE The agent produces a part number, a quantity, or an entitlement that reads perfectly well and is wrong. It lands in procurement as fact. The next person down the line treats it as one.

CONTROL **It never types into a system of record.** It selects from the catalog. Anything it cannot resolve to a catalog entry becomes an escalation instead of a guess. Every record it does write is tagged agent-originated so an auditor can tell later.

PROTECT IDENTIFY LLM09 · CATALOG-ONLY WRITES, NO FREE TEXT

Threat 02 · the request that looked fine and did nothing

LOOKS LIKE Someone asks. Everything looks good. Nothing gets recorded, sent, logged, or started. A week later they ask where it is.

CONTROL **Two layers.** Every request closes with a receipt carrying an ID and a clock. If no fulfillment event matches that ID inside the window, it escalates on its own. Overnight, a sweep compares intake against the downstream systems and puts any orphan back on the queue.

DETECT RESPOND RESPONSE CLOCK + NIGHTLY RECONCILIATION

DETECT is the pair of controls in Figure 05: the response clock and the nightly sweep behind it for anything the clock misses.

On Respond, I own triage for operational errors. Anything touching data exposure, access, or a controlled item goes into the security incident process the company already runs. I want that line bright: the worst moment to decide whether something counts as an incident is while it is happening. Recover is short: nothing here creates a new system of record. If the agent is pulled, requests fall back to the manual path.

05 Change management and training

ADKAR fits because adoption happens one person at a time. Every traveler has to choose the agent over texting me. ADKAR shows which part of that choice is failing. Awareness comes through the lunch-and-learn already on the calendar. Desire has to be earned on speed. Knowledge and ability come from two training tiers. Reinforcement is the receipt. From Kotter I borrow one idea: short-term wins earn each phase gate. Appendix B has the full mapping.

I own adoption. I know how that sounds: I am the person being taken out of the loop, so handing ownership to a service owner is one of the things Walk has to show. Nobody's job goes away. There is one of this role and demand is climbing. The agent frees time for exceptions, hands-on executive support, and transformation projects I cannot get to today. The KPI set deliberately carries no headcount metric.

COMMUNICATION runs through three things: the lunch-and-learn, the card in the kit, and the receipt after each request. Training runs in two tiers. Travelers get ten minutes, one path, and one card, inside that same session in month one. The support team gets a full session covering what the agent does, where it stops, and how to take over a handoff.

Shadow AI is already governed by our corporate acceptable-use policy. What this project adds is the part a policy cannot do alone. People paste request details into whatever assistant is already open because it is faster than the sanctioned path, so making the governed path the fastest one is the real control. Prohibition without a better option does not stop the behavior. It pushes the behavior somewhere nobody can see.

FEEDBACK comes three ways: one tap when the request closes, a short survey after the hardware lands, and a control at every step that hands the request to a person and records why. The survey is the real verdict. Every use of the handoff control is a labeled example of the agent failing.

RESISTANCE arrives in a predictable order. Convenience comes first: texting me works, so the bar is beating a text message. Trust comes second. That failure does not look like one. Somebody tries the agent once, something goes wrong in a small way, and next time they order their own charger overnight and say nothing. No complaint gets filed. The requests simply stop arriving. So the standing group watches the mix of requests as well as the volume. A category that goes quiet gets investigated before anybody counts it as a saving. The receipt was built as a security control. It doubles as the confirmation people can believe.

06 Scaling strategy

FIG 06 SCALING *Three phases. Crawl has a date. Walk and Run open on gates.* COMMITTED GATED

Crawl
DAYS 0 TO 90

SCOPE
Intake only. One pilot group, one item class. The agent captures, validates, and routes. A person places every order.

HUMAN IN THE LOOP
Every request. Nothing leaves without a human touching it.

GATE TO THE NEXT PHASE
Two consecutive clean weeks at normal volume: no orphaned requests, no missed escalation.

Walk
GATE-OPENED · MONTHS 4 TO 8

SCOPE
The write path opens. The agent splits into intake, entitlement, and fulfillment. Catalog widens by item class.

HUMAN IN THE LOOP
Five triggers stop it. Everything else runs unattended.

GATE TO THE NEXT PHASE
Two more clean weeks with the write path live, reconciliation still returning zero.

Run
GATE-OPENED · MONTH 9 ON

SCOPE
Full non-controlled commodity catalog, every supported site, every time zone.

HUMAN IN THE LOOP
Exceptions and samples. People see what the agent stopped on, plus a monthly slice of what it did not.

STANDING GATE
Monthly review, quarterly scope. Widening scope goes back through the group.

The hard part is the write: the moment a validated request has to land in procurement, asset, and ticketing systems built on the assumption that a human with credentials is clicking through a form. The silent failure lives in that same moment: the write that fails quietly. Crawl keeps a person in the write path entirely. The reconciliation sweep runs before that path opens, so nobody waits for something to go missing. In Figure 06, if the gates clear without a stall, Walk opens in months four to eight and Run from month nine.

07 Governance and compliance

By name and jurisdiction, ITAR (22 CFR 120 to 130) and the EAR (15 CFR 730 to 774) govern export-controlled technical data in the United States. CMMC and NIST SP 800-171 set the contractual security baseline for defense work, while NIST CSF 2.0 and the NIST AI RMF give the security and AI risk structure. Where we operate in the EU, the AI Act's transparency obligations apply to employees interacting with an AI system. Article 50(1) took force in August 2026. The GDPR applies to anything the agent holds about a person. Section 04's redaction keeps enough non-medical detail to answer an explanation request. In the United States, the Texas Data Privacy and Security Act and the California Consumer Privacy Act reach the personal data in scope: who asked, where it ships, and their request history.

The scope fence is the compliance argument. Holding the agent to non-controlled commodity gear means ITAR and EAR technical data never enters its context. The classification check at intake turns that from a claim into a control. The agent runs inside an enclave already covered by CMMC and 800-171. Nothing in this design extends them.

TESTING works by replay. Real historical requests run through the agent in a sandbox, scored against what actually happened, so it becomes an A/B comparison against the manual process on the same cases. Replay runs before go-live and then monthly as a regression check. The evaluation set costs nothing. Appendix C shows how it is scored. Replay and the fairness check below carry the AI RMF's Measure function.

Before launch, somebody on the team also tries to talk the agent into leaking a shipping address or a travel schedule. Whatever gets through gets fixed and added to the test set.

GOVERNANCE sits with a standing group of IT, security, procurement, and me, meeting monthly on metrics and exceptions, quarterly on scope, and on a phase gate whenever its condition is met.

On **TRANSPARENCY**, the agent says what it is at the start of every interaction and says what it cannot do. The disclosure sits behind a continue button. The boundary is the part that builds trust. Every record it writes downstream is tagged agent-originated, so an auditor can tell agent from person later. Anyone can ask for a human at any point, without justifying it and without waiting longer for it.

On **FAIRNESS**, the risk is plain. Somebody who writes a clean, well-formed request gets fast service. Somebody who writes in a hurry, in a second language, or in an unusual way gets escalated more often. Left alone, the agent would reward fluency. Fluency is not evenly distributed. So we watch escalation rate per requester instead of in aggregate. Any clustering counts as a defect in the agent.

08 Key performance indicators

FIG 07 KPIS *Baselines, targets, and how often we check them.*

TIME TO SERVE

≤1 day REQUEST TO HARDWARE MOVING

BASELINE 1 to 4 business days. It stretches with travel. An estimate pending validation.

TARGET Within one business day at Run in any time zone.

MONTHLY | OPERATIONAL

CONTAINMENT

70 % CLOSED WITHOUT A HUMAN TOUCH

BASELINE Zero. Today every request passes through one person.

TARGET 40% at Walk, 70% at Run. The rest are meant to reach a person.

MONTHLY | ADOPTION

SILENT FAILURES

0 REQUESTS WITH NO DOWNSTREAM EVENT

BASELINE Unknown. That is the problem. Nothing counts them today.

TARGET Zero in every phase. The agent does not run without it.

CONTINUOUS | QUALITY AND COMPLIANCE

FINANCIAL · CARRIED IN SECTION 03 Avoided trip failures counted from Crawl. Baseline: not counted today. Target: clear the 2-to-3 break-even inside Year 1.

WHEN A NUMBER MISSES A miss **holds the next phase gate closed** until the number comes back. The group decides whether the gate or the design was wrong.

Each number in Figure 07 answers one question. Time to serve tells me whether the agent beats a text. Containment tells me whether people choose it. The silent-failure count tells me whether the write path can be trusted. The money answer is in section 03, once the pilot supplies the avoided-failure count.

09 Conclusion

The business case is small: a few tens of thousands of dollars, mostly my own hours, against a loop that runs at the speed of one person's availability and a failure mode that costs somebody a trip. The path to value is unglamorous: keep a person in the write path through Crawl, hold two clean weeks at normal volume, and open the next phase only when the numbers say so.

The one thing that decides it is whether the sanctioned path becomes faster and more trusted than a text to me at nine at night; if it does not, this is theater and the requests come back to my phone.

One last thing that I want in writing now instead of worked out after a bad week: when the agent gets something wrong, that is mine; I would not want to build it under any other arrangement.

Appendix · everything behind the plan

Outside the graded pages. Section E holds ideas I am still thinking about.

A WHERE THE NUMBERS COME FROM

FIGURE USED	WHERE IT COMES FROM	HOW IT GETS CHECKED
Time to serve and weekly volume	A year of requests, from memory. Nobody logged them	One pull of twelve months of tickets before Crawl. The spread matters more than the average: the slow requests are the ones that cost somebody a trip
CapEx of $18k to $32k	120 to 200 build hours at our internal hourly rate	Finance confirms the rate. Platform and security confirm the hours after design review
No extra running cost	Copilot seats are already licensed for everybody in scope	Licensing confirms the coverage and tells me whether any add-on applies
$12,000 to $20,000 per avoided trip failure	Airfare and lodging spent, plus hourly rate times trip hours, plus a cost for the missed meeting	Agree the number with the sponsoring organization before anyone builds a case on it
Break-even at 2 to 3 avoided failures	The $33k Year 1 midpoint divided by the per-failure range	Recompute from the confirmed CapEx and the agreed per-failure number. The pilot supplies the count
70 percent containment at Run	A goal I picked. There is no data behind it yet	Replace it with whatever Crawl replay produces and say so when I do
Badge readers on the touchscreens	Not priced. A concept in E.3, absent from section 03	Price the readers and their integration with platform before the concept enters the cost table

B HOW PEOPLE MAKE THE SWITCH — ADKAR

STAGE	WHAT IT MEANS HERE	MECHANISM
Awareness	*People know the path exists*	The lunch-and-learn already on the calendar and the card in the kit
Desire	*They want it more than texting me*	Speed. It has to beat sending me a message
Knowledge	*They know how to use it*	Tier one: ten minutes and one path. Tier two: a full session for the support team
Ability	*They can do it at 9 p.m. the night before a flight*	One entry point, no sign-in (a badge tap, E.3), and a way out to a person at every step
Reinforcement	*It works and they keep choosing it*	The closeout receipt on every request. Kotter's short-term wins carry each phase gate

C HOW THE TESTING WORKS — REPLAY

01 The set is the trailing twelve months of real requests, cleaned, with shipping addresses de-identified for test use.

02 Each request gets scored on what the agent proposed against what actually happened: a match, a harmless difference, or an error. A person reads the harmless differences. Some are the agent getting something right that I got wrong.

03 Every historical case that should have hit one of the five triggers has to hit it. Missing one blocks the release. It does not simply lower the score.

04 Before go-live, somebody also tries to talk the agent into leaking a shipping address or a travel schedule. Whatever gets through gets fixed and added to the set.

05 Replay runs in a sandbox before go-live and monthly afterward. Nothing ships that scores below the month before it.

E.1 TECHBAG IQ CONCEPT — LIVE

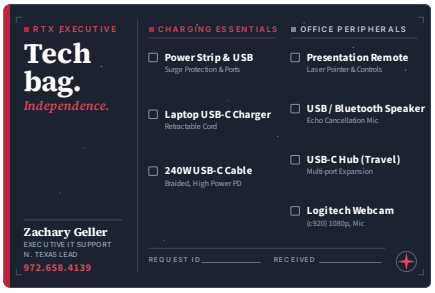
zach.technology/techbag-iq/index.html

I built this during Assignment 2.1 and host it on my own domain with what the course taught me about getting a vibe-coded app online. It is a rough first draft of the back-end GUI, a starting point for the touchscreen the agent would sit behind, and a sketch of what could be tracked.

As built: a command center, a kit builder from the Executive Tech Bag template, a fulfillment queue, an audit trail, and analytics panels.

Treat it as scaffolding. It predates this design and needs reworking to it: the five triggers, the receipt, the three KPIs, and the scope fence are not in it yet.

E.2 THE CARD IN THE BAG AND RECEIPT



The card that travels in the kit doubles as the packing list and, with a request ID and two timestamps, as the receipt. Same object, three jobs.

E.3 TAP YOUR BADGE, SKIP THE TYPING — CONCEPT

The badge that opens the door and releases a print job could start a request. After one tap the screen already knows who is asking, their role, site, kit, and ship-to. The traveler confirms instead of explaining. That takes the anxiety out of talking to a machine. The wrong escalations shrink too. The address check runs against the badge record instead of typed text.

Every tap is a measured interaction: time to serve from the tap, handoff and repeat use per requester, an identity on the record. More use makes more data. More data makes the agent easier to use. That is ITIL 4's continual improvement loop worked by the standing group.

Guardrails first. A tap shows the badge is present. It does not show who is holding it, so anything over the line still needs a second factor. The screen shows a first name and a kit. It never shows an address. It clears on a second tap or a timeout.

E.4 ONE MORE THOUGHT · A VENDING MACHINE — CONCEPT

Maybe a vending machine is another angle on this: the same badge tap on a machine in the hallway, stocked with the true grab-and-go commodities, keyboards, mice, chargers, cables. The agent keeps the kits, the judgment calls, and the shipping. The machine covers the person who simply needs a mouse at 7 a.m. and does not want a conversation. Same catalog, same receipt, same data. It is worth pricing next to the badge readers.

D SOURCES

IN ORDER OF FIRST APPEARANCE · WEB ADDRESSES WITHOUT THE PROTOCOL

FRAMEWORKS AND STANDARDS

01 NIST. *The NIST Cybersecurity Framework (CSF) 2.0.* CSWP 29, Feb 2024. doi.org/10.6028/NIST.CSWP.29

02 NIST. *AI Risk Management Framework (AI RMF 1.0).* AI 100-1, Jan 2023. doi.org/10.6028/NIST.AI.100-1

03 NIST. *AI RMF: Generative AI Profile.* AI 600-1, Jul 2024. doi.org/10.6028/NIST.AI.600-1

04 NIST. *Protecting CUI in Nonfederal Systems and Organizations.* SP 800-171 Rev. 3, May 2024. doi.org/10.6028/NIST.SP.800-171r3

05 U.S. DoD. *CMMC Program,* 32 CFR Part 170 (pub. Oct 15, 2024, eff. Dec 16, 2024); DFARS rule, 48 CFR, Case 2019-D041 (pub. Sep 10, 2025, eff. Nov 10, 2025).

06 OWASP. *Top 10 for LLM Applications 2025.* Nov 18, 2024. Threat naming in §04. owasp.org/www-project-top-10-for-large-language-model-applications

REGULATION

07 ITAR, 22 CFR Parts 120 to 130. U.S. Dept. of State, DDTC.

08 EAR, 15 CFR Parts 730 to 774. U.S. Dept. of Commerce, BIS.

09 Regulation (EU) 2024/1689 (AI Act), Art. 50, transparency obligations, applicable from 2 Aug 2026; unchanged by the 2026 Digital Omnibus.

10 Regulation (EU) 2016/679 (GDPR), Art. 5 (minimisation, storage limitation) and Art. 17 (erasure).

11 Texas Data Privacy and Security Act, Tex. Bus. & Com. Code ch. 541 (H.B. 4, 2023), eff. Jul 1, 2024.

12 California Consumer Privacy Act of 2018, Cal. Civ. Code § 1798.100 et seq., as amended by the CPRA (2020).

CHANGE, SERVICE MANAGEMENT, PLATFORM

13 Hiatt, J. M. *ADKAR: A Model for Change in Business, Government and Our Community.* Prosci, 2006.

14 Kotter, J. P. *Leading Change.* Harvard Business School Press, 1996.

15 AXELOS. *ITIL Foundation: ITIL 4 Edition.* TSO, 2019. Continual improvement, used in E.3.

16 Microsoft. *Data, privacy, and security for Microsoft 365 Copilot.* Microsoft Learn, accessed Aug 2026. learn.microsoft.com/copilot/microsoft-365/microsoft-365-copilot-privacy

17 Microsoft. *Connect your agent to an existing MCP server.* Copilot Studio docs, Microsoft Learn, accessed Aug 2026. learn.microsoft.com/microsoft-copilot-studio/mcp-add-existing-server-to-agent

18 Desai, Z. "MCP is now generally available in Microsoft Copilot Studio." *Microsoft Copilot Blog,* May 29, 2025.

19 Model Context Protocol. *Specification.* Accessed Aug 2026. modelcontextprotocol.io/specification

COURSE MATERIALS · MIT PROFESSIONAL EDUCATION, APPLIED AGENTIC AI FOR ORGANIZATIONAL TRANSFORMATION, 2026

20 Module 8, Video 8.1, "The Capstone." Transcript MPE-DTR_R3_M8_V1. The brief.

21 *Capstone Project Template* (updated). Section structure and the "must demonstrate" checks.

22 Geller, Z. *Assignment 1.1: Evaluating the Cost of AI Systems,* Jul 18, 2026. Token-level cost illustration, §03.

23 Geller, Z. *Assignment 2.1: Vibe Coding,* Jul 18, 2026. Open versus closed platforms, §02; origin of TechBag IQ, E.1.

24 Geller, Z. *Assignment 3.1: Conceiving and Programming of Agents,* Aug 3, 2026. First version of the provisioning assistant.

25 Geller, Z. *Assignment 4.1: AI Risks and Security Plan,* Aug 15, 2026. NIST structure and the Detect gap behind the silent-failure control.

26 Geller, Z. *Assignment 5.1: AI and the Design Process,* Aug 23, 2026. The Travel Tech Bag agent and the current-state workflow, §01.

27 Geller, Z. *Assignment 6.1: KPIs for AI,* Aug 24, 2026. Time to serve, handoff rate, and repeat use, §08.

28 Geller, Z. *Assignment 7.1: Governance Plan,* Aug 25, 2026. The 8:52 a.m. scenario, the principle, the never-refuse rule, the leak test.

GENERATIVE AI ASSISTANCE · DISCLOSED PER ACADEMIC INTEGRITY POLICY

29 Anthropic. *Claude.* 2026. Used to draft and edit prose against my own voice profile, build the layout and figures, verify arithmetic and cross-references, and audit the draft against the capstone template. All content directed, reviewed, and approved by the author.

30 Google. *Gemini.* 2026. Used for an independent review of the technical design and a draft workflow graphic. Suggestions were evaluated individually and selectively adopted; the graphic was redrawn instead of used.