



Securing Tomorrow's Missions Today.



Before the Build: A Fixed-Fee AI Use-Case Readiness and Governance Assessment for One Use Case

Cloud AI/ML Services: find out whether one AI use case is buildable, what evidence OMB M-25-21 will ask for, and what a pilot should cost, before the agency spends pilot money.

AvalonTechServices.com

contact@AvalonTechServices.com

Executive Summary	2
Why AI Pilots Stall	3
Why It Concentrates in Government	3
The Mandates and the Calendar	4
What the Assessment Does	5
Deliverables	6
Where the Engagement Stops	6
Evidence the CAIO and the AO Can Use	7
How an Engagement Runs	8
What the Agency Provides	8
What Is Not Included	9
How Avalon Staffs It	9
Pricing and How to Buy	10
Illustration: One Use Case, 14 Open Evidence Items	11
The Evidence Gap	11
Data Readiness	12
What the Assessment Costs	12
Common Pitfalls	13
Conclusion and Next Step	13
Appendices	14
Appendix A: Glossary	14
Appendix B: Illustration Assumptions	14
Appendix C: References	15

Executive Summary

Most agencies now have access to capable AI models through cheap enterprise licenses and managed AI services inside FedRAMP Certified clouds. What most lack is a working capability: a model connected to mission data, evaluated on that data, and documented well enough that the Chief AI Officer (CAIO) can stand behind it and the ISSO can let it run. Pilots stall because the data was not ready, security came in at the end, or nobody agreed on what success looked like.

OMB Memorandum M-25-21 (April 2025), which replaced M-24-10, raises the stakes. It keeps agency AI use-case inventories, assigns CAIO roles, and sets minimum risk-management practices for high-impact AI: testing before deployment, an impact assessment, ongoing monitoring, human oversight, and paths for notice and remedy. M-25-22, which replaced M-24-18, governs how agencies buy AI, with an emphasis on data rights, testable claims and avoiding lock-in.

This paper is for mission program managers who own the backlog, CAIOs and CIOs who sponsor AI work, Chief Data Officers, ISSMs and ISSOs who guard the boundary, and the CORs who buy the work. It describes Avalon Technology Services' Cloud AI/ML Services, which starts with one engagement:

- **A fixed-fee readiness and governance assessment for one named AI use case.** Two to four weeks, on the agency's own cloud. It defines the use case and its success metrics, supports the CAIO's high-impact determination, checks whether the source data is usable, measures the evidence gap against M-25-21 minimum practices and the NIST AI Risk Management Framework, and ends with a scoped pilot statement of work and a go or no-go recommendation. Market range \$15K to \$40K.
- **A conditional follow-on.** A firm-fixed-price pilot for the same use case, built and evaluated inside the agency tenant, priced in the statement of work the assessment produces.

The illustration in this paper works through one representative use case: summarizing incoming case files for an intake team. Of 18 illustrative minimum-practice evidence items, 4 are documented, 6 are partial and 8 are absent. Of four source systems, one is not ready. The assessment runs three weeks at a planning fee of \$28.4K, and it shows that deferring the weakest source to a later use case shortens the data readiness phase of the pilot.

Avalon builds nothing in this first order. It documents what exists and what is missing. The CAIO makes the risk-tier and waiver decisions, the agency's assessors assess, and the Authorizing Official (AO) decides whether any AI capability operates.

Why AI Pilots Stall

The same phrases come up in agency after agency, and each one points to a gap that appears before any model is built.

What the program office says	What it usually means
"We have a pilot that never went anywhere."	No success metric was agreed at the start, so nobody could say whether the pilot worked
"Our data isn't ready for AI."	Nobody has checked access approvals, quality, data rights and location for the specific sources one use case needs
"Security won't approve anything."	The ISSO saw the design late, with no boundary description, logging plan or documentation to review
"We bought the licenses but nobody's built anything with them."	The model layer is in place; the engineering that connects it to mission data inside the boundary is not
"We can't hire data scientists."	The agency needs outside engineers for a bounded job and a handoff, not a permanent team
"Everyone's already using ChatGPT on their own."	Ungoverned use is filling a gap that a sanctioned, documented capability should fill

Why It Concentrates in Government

- **Talent is scarce.** Federal ML engineering talent is hard to hire, and recent workforce reductions cut in-house technical capacity.
- **Every capability crosses an authorization process.** A model endpoint is part of a system the AO has to accept, so documentation is not optional.
- **Only authorized services are usable.** The work has to run on AI services inside the agency's FedRAMP Certified cloud offering, at the right impact level.

The Mandates and the Calendar

Driver	What it requires	What it means for this engagement
OMB M-25-21 (April 2025; replaced M-24-10)	Use-case inventories, CAIO roles, minimum risk-management practices for high-impact AI	The evidence gap is measured against these practices, using the checklist the CAIO office supplies
OMB M-25-22 (replaced M-24-18)	How agencies acquire AI: data rights, testable claims, avoiding vendor lock-in	The pilot statement of work is written to these terms
NIST AI RMF 1.0 (Jan 2023) and NIST AI 600-1 Generative AI Profile (July 2024)	Voluntary guidance organized as Govern, Map, Measure and Manage, with risks specific to generative AI	The structure for the governance evidence; there is no AI RMF certification
NIST SP 800-53 Rev. 5 and SP 800-218A	Security controls for the hosting system; secure development practices for AI models	Evidence for AC, AU, CM, SI and SA controls, including SA-11 and SI-4, is planned into the pilot
FedRAMP Consolidated Rules (Jun 25, 2026)	Cloud offerings are FedRAMP Certified in classes A through D; certification covers the provider's offering only	Only certified AI services in the agency tenant are used; the agency still authorizes its own system
OMB M-26-14 (May 2026; replaced M-21-31)	A risk-based logging model	Logging for model endpoints and prompts follows the agency's logging decisions
DoD CSRMC (Sep 24, 2025)	Construct announced; RMF, DoDI 8510.01 and eMASS remain in use	Defense use cases follow the component's RMF process at IL2, IL4 or IL5; IL6 is out of scope

The assessment gets bought at predictable moments: a CAIO asking every office for real use cases; licenses in place with nothing mission-facing shipped; a stalled pilot; a

shadow-AI incident or inspector general finding; a new CAIO or CIO; a workforce reduction that turned a backlog into a crisis; and the FY2028 budget build, where a scoped pilot with a cost estimate is easier to defend than an idea.

What the Assessment Does

This is neither AI strategy slides nor a platform sale. It takes one named use case and answers three questions an engineer and an ISSO can both check: is it buildable on this data, what evidence will it need, and what should the pilot cost?

Workstream	Activities
Use-case definition	Working sessions with the mission owner to fix the use case, the users, the decision it supports and two to four success metrics (for example accuracy, precision and recall, or groundedness for generative AI)
Risk tier	A documented recommendation on whether the use case is likely high-impact under M-25-21, for the CAIO's determination
Data readiness	A desk-level check of each source system: access approved, quality checked, rights and privacy approvals cleared, and already in the agency cloud
Governance evidence gap	Each item on the CAIO's minimum-practice checklist tagged documented, partial or absent, organized by AI RMF function and the AI 600-1 profile
Platform fit	Which certified AI services in the agency tenant can host the pilot, and what logging and access the ISSO will need to see
Pilot scoping	A capped statement of work, a schedule tied to agency dependencies, and a consumption estimate for compute and inference on the agency's own bill

Deliverables

Deliverable	Format
Use-Case Feasibility & Prioritization Report: the use case, metrics, risk-tier recommendation and go or no-go	Word and PDF, with a readout to the sponsor
Data Readiness Assessment: source inventory, readiness checks, access and rights gaps, pipeline outline	Word and PDF
Evidence-gap register: each minimum-practice item, status, owner and the pilot task that closes it	Excel or CSV
Scoped pilot statement of work with a consumption estimate	Word, ready for the contracting officer

Where the Engagement Stops

The assessment deploys into nothing and builds nothing. If the agency has no landing zone, that is Avalon's separate Cloud Architecture & Engineering service. If the data must first move to the cloud, that is Cloud Migration. Controlling AI compute spend after go-live is FinOps work, and placing a capability in a shared government cloud is Gov-Shared Cloud Services work. Writing the system security plan or running the authorization belongs to the ATO and NIST RMF services. Each is scoped and priced on its own.

Evidence the CAIO and the AO Can Use

Framework element	What the assessment produces
M-25-21 minimum practices	A status for every checklist item, with the gap, owner and the pilot task that closes it
AI RMF Govern and Map	Purpose, context, users, accountable official and risk-tier recommendation
AI RMF Measure	Success metrics and an evaluation plan on representative agency data
AI RMF Manage and AI 600-1	Monitoring, human review and generative AI risks the pilot has to address
800-53 Rev. 5 (SA-11, SI-4, AU, AC, CM)	The evidence the pilot must produce, planned in advance for the ISSO

The accurate way to state the value is narrow: the engagement documents the technical practices M-25-21 requires for a high-impact use case and plans the evidence that the agency's assessors and AO will need. Determinations of compliance, waivers and authorization remain with the agency and its designated officials.

Accurate	Not accurate
An evidence-gap register for one use case	Makes the agency M-25-21 compliant
A risk-tier recommendation	The CAIO's risk-tier or waiver decision
Planned evidence for the ISSO and assessor	An ATO, or a promise that a pilot will receive one
Work inside FedRAMP Certified services	A FedRAMP Certified solution from Avalon
An evaluation plan with stated metrics	A guarantee of model accuracy

How an Engagement Runs

Step	Timing	What happens
Kickoff	Week 1	Confirm the use case, users and sponsor; brief the ISSO and the CAIO office; receive the minimum-practice checklist and source list
Use-case and data sessions	Weeks 1–2	Working sessions with the mission owner and data steward; success metrics drafted; readiness checks per source
Evidence gap and platform fit	Weeks 2–3	Checklist items tagged; certified services and logging reviewed with the ISSO; risk-tier recommendation drafted
Readout	Week 3	Reports, register and pilot statement of work delivered; go or no-go discussed with the sponsor

If the answer is go, the pilot follows the same phases the assessment scoped: data readiness and environment setup (3 to 6 weeks), then pilot build and evaluation on real agency data (6 to 10 weeks). An honest miss at the end of evaluation is a valid exit: a documented negative result at pilot price is cheaper than a quiet failure at program price. Production hardening and transition are separate, later orders.

What the Agency Provides

- A named product owner or mission subject matter expert for 2 to 4 hours a week.
- A data steward who can describe each source system and its access, sharing and privacy approvals.
- The ISSO or ISSM and the CAIO office engaged in week 1, and the agency's minimum-practice checklist.
- Read access to existing documentation; no production data leaves the agency.
- A decision of record on generative AI acceptable use for the pilot's users, or a named owner who will make it.
- Agency-caused delays move the schedule day for day. Tenant accounts for any pilot take 2 to 6 weeks to provision, so the pilot statement of work dates them.

What Is Not Included

Included	Separate work
One named use case	A portfolio of use cases or an AI strategy
A desk-level data readiness check	Building a data platform or pipelines
Platform fit within the existing tenant	A landing zone, migration or shared-cloud placement
A consumption estimate for the pilot	A cost audit of the agency's cloud bill
Advice on certified services	Resale of licenses, models or GPU capacity
Unclassified work in the agency tenant	Classified or IL6 work

How Avalon Staffs It

Avalon takes one engagement at a time, works business hours, and works on unclassified systems only, inside the agency's own tenant. Avalon's founder serves as delivery lead, with one named AI/ML lead engineer, whose certification on the agency's cloud platform is a condition of the engagement, and a compliance specialist for the M-25-21 and AI RMF mapping. Avalon has no past federal engagements under this service. FAR 15.305 says an offeror without a record of past performance may not be evaluated favorably or unfavorably on that factor; what protects the agency is a named team, a fixed price, a short schedule and agency ownership of every artifact. Avalon never assesses a system it built or operates.

Requirement	How it is handled
AI/ML lead engineer	Named personnel whose platform certifications and investigation level match the agency's requirements, listed in the proposal, hired or subcontracted
Compliance specialist	Named in the proposal, with experience mapping evidence to NIST frameworks
Civilian access	Public Trust investigation at the tier the agency requires, as a condition of the engagement
Sensitive data	No CUI, PII or PHI leaves the agency; any handling the agency requires is agreed before award
Data rights	The agency owns all reports, code, prompts and documentation from the first day
Insurance and contract vehicles	Any coverage the solicitation requires is a condition of award; Avalon does not hold a GSA Schedule or a governmentwide contract today

Pricing and How to Buy

Engagement	Market-norm range	Avalon
Readiness and governance assessment, one use case, 2–4 weeks	\$15K–\$40K	Firm fixed price
Scoped pilot, one use case, 10–16 weeks (conditional)	\$90K–\$250K	Firm fixed price, \$90K–\$180K
Pilot to production, 6–9 months	\$250K–\$750K	Not offered today
Multi-use-case program	\$750K–\$2.5M+ a year	Not offered today

The assessment and the pilot both sit under the \$350,000 simplified acquisition threshold, so either can be bought as an open-market purchase. Purchases between

the \$15,000 micro-purchase threshold and the SAT are generally set aside for small business when two or more capable small businesses are expected to compete. AI work inside a larger modernization task order can be subcontracted to Avalon by the prime. Avalon does not hold a GSA Schedule or a governmentwide contract today.

Cloud compute and model inference run on the agency's own cloud bill by design, so Avalon's price is never the total cost of the capability. The pilot statement of work carries a consumption estimate as a courtesy line, and usage is reported at each phase gate. Performance-based pricing is declined: model performance is not a clean basis for a contingency fee.

Illustration: One Use Case, 14 Open Evidence Items

The figures below are a planning illustration for one representative use case, not a quote and not a result from any engagement. Every checklist item, status, source and hour is an assumption, and the real checklist comes from the agency's CAIO office.

The use case: a generative AI assistant that summarizes incoming case files so an intake team can triage them faster. Because the summaries could influence decisions about individuals, the assessment recommends treating it as likely high-impact, and the CAIO decides.

The Evidence Gap

Practice area	Documented	Partial	Absent
Pre-deployment testing	0	1	2
AI impact assessment	1	1	1
Ongoing monitoring	0	1	2
Human oversight and training	1	1	1
Notice, remedy and feedback	1	1	1
Documentation and inventory	1	1	1
Total	4	6	8

14 of 18 items are open. The pattern is typical of a use case that has a sponsor but no build: the purpose, the accountable official, the inventory entry and the appeal path exist

on paper, while everything that depends on a running system (evaluation on real data, written limitations, drift and adverse-impact monitoring, operator training, a model card) is absent. That is not a failure; it is the pilot's task list, and each open item gets an owner and a closing task.

Data Readiness

Source system	Access approved	Quality checked	Rights cleared	In cloud	Ready
Case management database	Yes	Yes	Yes	Yes	100%
Scanned case-file repository	Yes	No	Yes	Yes	75%
Policy manual library	Yes	Yes	Yes	No	75%
Correspondence mailbox archive	No	No	No	Yes	25%

Three sources are usable with work: the scanned repository needs a quality check and the policy library needs to move into the cloud. The correspondence archive fails three of four checks, including rights clearance. The model sizes the pilot's data readiness phase at three weeks plus one for each partly ready source and two for each source that is not ready, within the 3 to 6 week range. With all four sources the phase reaches the 6-week cap and the pilot (data readiness plus build and evaluation) runs 12 to 16 weeks. Deferring the correspondence archive to use case #2 cuts the phase to 5 weeks and the pilot to 11 to 15 weeks, and removes the riskiest rights question from the first build. The pilot statement of work freezes the source list.

What the Assessment Costs

Scenario	Weeks	Hours	Rate-band range	Planning fee
Short	2	108	\$15.4K–\$22.4K	\$18.9K
Base	3	162	\$23.1K–\$33.6K	\$28.4K
Long	4	216	\$30.8K–\$44.8K	\$37.8K

The build is an AI/ML lead at 1.0 FTE (\$150 to \$220 an hour), a compliance specialist at 0.25 FTE (\$120 to \$170) and a delivery lead at 0.1 FTE (\$125 to \$175), at 40 hours a week. The base case is 162 hours over three weeks, \$23.1K to \$33.6K across the bands and \$28.4K at planning rates, inside the \$15K to \$40K market range. A two-week scope runs \$15.4K to \$22.4K, just above the micro-purchase threshold. A four-week scope runs \$30.8K to \$44.8K; its upper end breaks the \$40K ceiling, so a four-week assessment needs rates near the bottom of the band or a trimmed delivery line. Every case sits far under the simplified acquisition threshold.

Common Pitfalls

Pitfall	How to avoid it
"While you're at it, add another use case"	Freeze the use-case register; the next idea is use case #2 and its own order
Overselling compliance	Use the accurate phrasing; the CAIO, assessors and AO decide
Agency dependencies discovered late	Date accounts, data access and SME time in the statement of work
The curated-demo trap	Evaluate on representative real data, even if the demo looks less impressive
Unbudgeted compute	Consumption estimate at kickoff and usage reported at each gate
Building on a model that disappears	Model-portable design and agency-owned prompts, code and weights, written into every statement of work

Conclusion and Next Step

Most failed AI pilots fail before the model is built. A short assessment that answers the data, success and security questions for one use case costs a fraction of a pilot, and a no-go answer is worth as much as a go. The CAIO gets an honest evidence picture, the ISSO sees the design in week 1, and the program office gets a pilot statement of work it can buy.

The natural first step is the fixed-fee readiness and governance assessment: one named use case, two to four weeks, under the simplified acquisition threshold, on the agency's own cloud, ending in a feasibility report, a data readiness assessment, an evidence-gap register and a scoped pilot statement of work. To discuss a use case, contact Avalon Technology Services at contact@AvalonTechServices.com.

Appendices

Appendix A: Glossary

CAIO

Chief AI Officer, the agency official M-25-21 makes responsible for AI governance, the use-case inventory and high-impact determinations.

Evidence-gap register

A list of minimum-practice evidence items for one use case, each tagged documented, partial or absent, with an owner and a closing task.

Groundedness

For generative AI, how well an answer is supported by the source documents it was given.

High-impact AI

Under M-25-21, AI whose output serves as a principal basis for decisions with significant effects on rights or safety; subject to minimum risk-management practices.

Retrieval-augmented generation (RAG)

A generative AI pattern that retrieves relevant agency documents and passes them to the model as context for each answer.

Appendix B: Illustration Assumptions

All values are planning assumptions for one representative use case. None is drawn from a specific agency or a completed engagement.

Assumption	Value	Basis
Use case	Generative AI summaries of incoming case files for an intake team	Planning assumption
Evidence checklist	18 items in six practice areas; 4 documented, 6 partial, 8 absent	Illustrative; the CAIO office supplies the real list
Source systems	Four, each checked for access, quality, rights and cloud location	Planning assumption
Data readiness phase	3 weeks, +1 per partly ready source, +2 per not-ready source, capped at 6	Model rule within the KB's 3–6 weeks
Build and evaluation phase	6–10 weeks	KB phase range
Team loading	AI/ML lead 1.0, compliance 0.25, delivery lead 0.1 FTE; 40 hours a week	Planning assumption
Rates	\$150–\$220, \$120–\$170, \$125–\$175 an hour; planning rates \$185, \$145, \$150	KB AI/ML card and category rate card
Price check	Base fee \$28.4K	Inside the \$15K–\$40K market range

Appendix C: References

1. **Office of Management and Budget**, *M-25-21, Accelerating Federal Use of AI through Innovation, Governance, and Public Trust*, April 2025
<https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-21-Accelerating-Federal-Use-of-AI-through-Innovation-Governance-and-Public-Trust.pdf>
2. **Office of Management and Budget**, *M-25-22, Driving Efficient Acquisition of Artificial Intelligence in Government*, April 2025
<https://www.whitehouse.gov/omb/information-for-agencies/memoranda/>
3. **National Institute of Standards and Technology**, *AI Risk Management Framework (AI RMF 1.0)*, January 2023
<https://www.nist.gov/itl/ai-risk-management-framework>

4. **National Institute of Standards and Technology**, *NIST AI 600-1, Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*, July 2024
<https://doi.org/10.6028/NIST.AI.600-1>
5. **National Institute of Standards and Technology**, *SP 800-53 Rev. 5, Security and Privacy Controls for Information Systems and Organizations*
<https://csrc.nist.gov/pubs/sp/800/53/r5/upd1/final>
6. **National Institute of Standards and Technology**, *SP 800-218A, Secure Software Development Practices for Generative AI and Dual-Use Foundation Models*
<https://csrc.nist.gov/pubs/sp/800/218/a/final>
7. **FedRAMP**, *FedRAMP program and Consolidated Rules*
<https://www.fedramp.gov>
8. **Office of Management and Budget**, *M-26-14, Ensuring Effective and Efficient Agency Logging and Network Visibility to Defend Against Evolving Cyber Threats*, May 2026
<https://www.whitehouse.gov/wp-content/uploads/2026/05/M-26-14-Ensuring-Effective-and-Efficient-Agency-Logging-and-Network-Visibility-to-Defend-Against-Evolving-Cyber-Threats.pdf>
9. **Acquisition.gov**, *FAR 15.305, Proposal evaluation*
<https://www.acquisition.gov/far/15.305>
10. **Centre Law & Consulting**, *Inflation Updates to FAR Thresholds: What Changed on October 1, 2025*
<https://centrelawgroup.com/blog/inflation-updates-to-far-thresholds-what-changed-on-october-1-2025/>