

FEDERAL MODERNIZATION PRACTICE · RESEARCH REPORT

The Federal Integration Layer

Why middleware and iPaaS decide how — and how well — federal AI gets deployed; the legacy-ERP integration problem underneath; and what to evaluate before the platform decision locks in.

02	Executive summary
03	Where federal AI actually runs
05	What iPaaS solves that models can't
07	Legacy ERP is the load-bearing constraint
09	Where federal AI pilots actually stall
10	What to evaluate
11	What this rules in and out
12	Endnotes & sources

● Executive summary

Most discussion of federal AI moves between two surfaces — the policy layer (OMB memoranda, the NIST AI Risk Management Framework, the executive orders) and the model layer, where the public conversation tracks which foundation models agencies selected and which contract just got awarded. The layer between them, where AI actually runs when it runs, gets almost no attention. It is the integration layer, and the platforms operating in it are the most important federal technology category nobody is writing about.

~49%

of federal AI pilot stalls trace to integration — the largest single cause⁵

~11%

of stalls trace to the model layer — a small minority⁵

4

integration-layer capabilities the foundation model does not provide

~10yr

the vendor relationship a platform decision locks in for, once made

This report sets out why the integration layer — iPaaS-class middleware over legacy ERP and systems of record — decides whether federal AI deployments ship, and what agencies should evaluate before a platform decision locks in for a decade. Its argument is in five parts:

- **The model is not the architecture.** A foundation model is a stateless reasoning engine that produces tokens. The middleware is what turns those tokens into governed, auditable actions across federal systems.
- **Four capabilities matter that the model does not provide.** System connectivity at federal breadth, state management across long-running workflows, governance and audit, and federal-grade controls all live in the integration layer.
- **Legacy ERP is the load-bearing constraint.** The systems an agent must touch were built over decades to be touched by

humans and point-to-point integrations — not by an agent integrating the constellation at runtime.

- **Pilots stall in integration, not in the model.** Integration and governance together account for the majority of federal AI pilot stalls; the model layer accounts for a small minority.
- **The cost reality is the inverse of the visibility.** The foundation model is the most visible line item and one of the smallest; integration is the largest and is rarely scoped on its own.

"The federal AI procurement that matters most is not the foundation-model award. It is the integration-platform commitment underneath, which costs less to procure and matters more to outcomes."

A note on figures. Layer shares, stall distributions, and cost patterns in this report are drawn from FCI Advisory's engagement base across federal middleware, integration, and agentic AI program design, FY24-Q4 through FY26-Q1.⁵ They are illustrative of the directional pattern, not point-precise extracts.

● 01 Where federal AI actually runs

The technical architecture of a federal AI deployment is mostly not the model. It is the integration platform that lets the model touch anything. Federal agencies do not run AI on greenfield infrastructure; they run AI inside organizations with decades of accumulated systems — legacy ERP, thirty-year-old custom applications, identity infrastructure that predates cloud, records environments built before the iPhone.

An agentic AI system in a federal context has to reason across those systems, take actions inside them, retrieve and update records held inside them, and produce auditable trails of everything it did. The foundation model does none of that. The model is a reasoning engine that produces tokens. The middleware is what turns those tokens into actions in a federal environment. The category name for this middleware is iPaaS — integration platform as a service — a complete platform for moving data between systems, managing API contracts, transforming records as they flow, and orchestrating workflows that cross system boundaries.⁴

THE FEDERAL AI STACK: WHERE ATTENTION GOES VS. WHERE THE WORK HAPPENS

Layer	What it is	Engineering effort to deploy
06	Foundation model — commodity reasoning layer	Lowest. Most discussed; least decisive for whether a deployment ships.
05	Cloud infrastructure — FedRAMP-authorized compute	Moderate; largely solved procurement.
04	Governance & audit — trail, reversibility, defensibility	High; federal-specific.

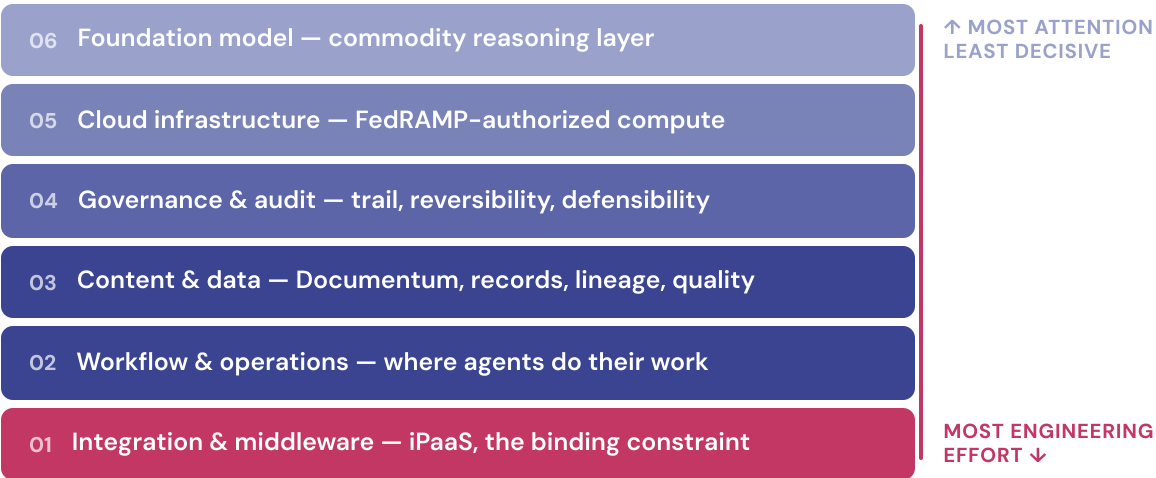
Layer	What it is	Engineering effort to deploy
03	Content & data — Documentum, records, lineage, quality	High; the retrieval surface.
02	Workflow & operations — where agents do their work	High.
01	Integration & middleware — iPaaS, the binding constraint	Highest. Least attended; does the most work.

Layers ordered from most-attended (top) to least-attended (bottom). Engineering effort runs the opposite direction. FCI Advisory observation across federal AI deployments, FY24-Q4 through FY26-Q1.⁵

FIGURE 01 · ATTENTION VS. ENGINEERING EFFORT

Attention runs top-down. Engineering effort runs the opposite way.

The most-discussed layer — the foundation model — is the least decisive for whether a deployment ships. The integration layer is the binding constraint.



Layers ordered from most-attended (top) to least-attended (bottom). Engineering effort runs the opposite direction. FCI Advisory observation across federal AI deployments, FY24-Q4 through FY26-Q1.⁵

Each iPaaS platform has been quietly added to the federal vendor mix over the last several years as agencies have figured out that integration is a binding constraint on every modernization effort. No single platform dominates across federal sectors —

defense weights on-premise and air-gapped options, civilian agencies weight FedRAMP-High coverage, and infrastructure operators weight legacy-system integration depth. The federal iPaaS market is genuinely contested, and the platform decisions being made now will lock agencies into roughly ten-year vendor relationships.

● SECTION 02 · THE CAPABILITIES

What iPaaS solves that models can't

02

● 02 What iPaaS solves that models can't

Foundation models, even the best ones, are stateless reasoning engines. Federal AI deployments require something fundamentally different: persistent, governed, auditable interaction with operational systems. Four capabilities matter at federal scale, and the model delivers none of them.

FIGURE 02 · FOUR CAPABILITIES THE MODEL DOES NOT PROVIDE

The model produces tokens. The middleware turns them into governed, auditable action.

Each of these lives in the integration layer, not the model — and each is a prerequisite for a federal deployment.

System connectivity at federal breadth

One workflow may read a Documentum repository, write an Oracle HR system, query a mainframe records system, and notify a collaboration suite — each on a different identity system and audit standard.

State management

Models hold no state between calls. Federal workflows are stateful — eligibility unfolds over weeks, procurement over months. The middleware holds the state the model needs to act consistently.

Governance and audit

Every action must be auditable, reversible, explainable. The middleware records the decision, ties it to the operator's identity, and makes the chain queryable when an inspector asks.

Federal-grade controls

FedRAMP boundaries, data-classification rules, rate limits, and circuit breakers that prevent a runaway agent from creating thousand-record errors. The model has no awareness of any of this.

FCI Advisory framework, derived from federal AI integration engagement observation.

FOUR INTEGRATION-LAYER CAPABILITIES THE MODEL DOES NOT PROVIDE

Capability	Why it lives in the middleware, not the model
System connectivity at federal breadth	A single workflow may read from a Documentum repository, write to an Oracle HR system, query a mainframe-backed records system, post to a ticketing platform, and notify a collaboration suite — each governed by a different identity system and auditable to a different standard.

Capability	Why it lives in the middleware, not the model
State management	Models hold no state between calls. Federal workflows are stateful — a benefits-eligibility process unfolds over weeks; a procurement spans months. The middleware holds the state the model needs to act consistently across that timeline.
Governance and audit	Every agent action must be auditable, reversible, and explainable. The model produces a decision; the middleware records it, captures context, ties it to the operator's identity, and makes the chain queryable when an inspector asks. Without that layer, the agent is unauditable — which means unauthorized.
Federal-grade controls	The middleware is where FedRAMP boundaries are enforced, data-classification rules applied, and rate limits and circuit breakers prevent a runaway agent from creating thousand-record errors. The model has no awareness of any of this.

FCI Advisory framework, derived from federal AI integration engagement observation.

The cost inversion most procurements miss

Federal AI procurement narratives center on model cost, because foundation models are visible, branded, and easy to compare. Integration platforms are commodity-named, less branded, harder to compare, and usually scoped as a line item buried inside a larger statement of work. The cost reality is the inverse of the visibility: the foundation model is a small share of total federal AI program cost, and integration is the largest. Vendor selection optimized against model benchmarks is optimizing the smallest line item against the largest source of risk.

The foundation model is the most visible, branded, and discussed component of a federal AI program — and one of the smallest cost lines. The largest line item, integration, is rarely scoped on its own.

● 03 Legacy ERP is the load-bearing constraint

The reasoning surface for a single federal workflow is not exotic. It is the same constellation an experienced agency employee navigates manually every day. What changes when an AI agent is inserted is that the navigation has to be machine-traversable in real time, the agent has to maintain transactional consistency across systems that do not share a transaction model, the audit trail has to span every system the agent touched, and the failure modes have to be recoverable without losing operational state. None of this is a model capability. All of it is middleware capability.

The systems the agent touches were not built to be touched by an agent. They were built, in sequence, over twenty or thirty years, to be touched by human users following procedures and by point-to-point integrations carrying specific transactions between specific systems.¹ The agent is now expected to integrate the constellation at runtime — a class of integration problem the underlying systems were not engineered to support.

INTEGRATION COMPLEXITY BY SYSTEM AGE (ILLUSTRATIVE)

System cohort	Est. share of estate ⁵	Agent-readiness assessment
Pre-2005	~22%	Highest integration complexity. Often requires a middleware abstraction layer or replacement before an agent can touch it safely.
2005–2015	~46%	Moderate; integration patterns predate modern API conventions. Workable with deliberate adapter work.
Post-2015	~32%	Lowest; more API-addressable, though still requiring governed orchestration.

Three federal system-age cohorts with integration-complexity assessment. Share figures reflect FCI's engagement observation; the directional shape is consistent across federal civilian and defense.⁵

FIGURE 03 · INTEGRATION COMPLEXITY BY SYSTEM AGE

Two-thirds of the estate predates modern API conventions.

The oldest cohort carries the highest integration complexity — and often needs an abstraction layer before an agent can touch it safely.



Three federal system-age cohorts with integration-complexity assessment. Share figures reflect FCI's engagement observation; the directional shape is consistent across federal civilian and defense.⁵

Pre-agentic vs. agentic integration architecture

Pre-agentic federal integration was largely point-to-point or hub-and-spoke. A specific transaction moved from System A to System B on a known schedule, through a known adapter, with a known schema. The integration map was a collection of pipes, each carrying defined traffic. The iPaaS modernization wave of the late 2010s consolidated many of those pipes onto common platforms without fundamentally changing the shape.

Agentic workloads do not fit the pipe model. An agent does not move a defined transaction on a schedule; it reasons across the system landscape in response to a runtime query, decides which systems to read from and write to, and operates against multiple systems within a single workflow. The integration layer has to support arbitrary read access (not just scheduled extracts), runtime authorization decisions (not just pre-provisioned trust), transactional rollback across systems that do not share a transaction manager, and audit logging that spans every system the agent touched in a single decision. The middleware has to support a runtime reasoning mesh, not scheduled transactions.

What changes with the agent. The constellation already exists — agency employees navigate it manually every day. What changes is that the navigation has to be machine-traversable in real time, with transactional consistency, audit coverage, and recoverable failure across systems that do not share a transaction model. None of this is a model capability.

SECTION 04 · THE EVIDENCE

Where federal AI pilots actually stall

04

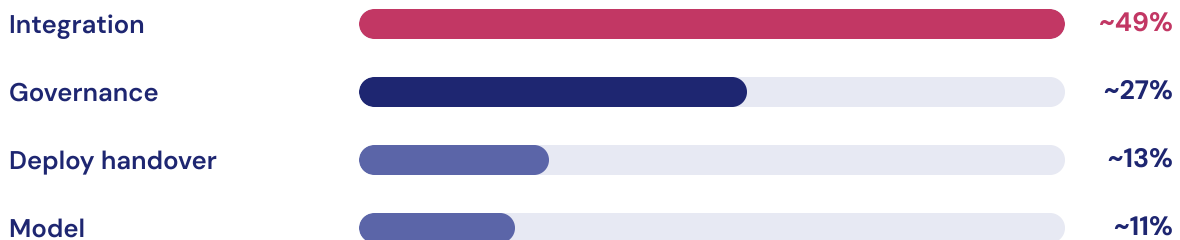
04 Where federal AI pilots actually stall

Pilots that stall do not stall evenly across the stack. They cluster at the stages where the integration work is hardest — data access from legacy systems, transactional consistency across multiple systems of record, identity and authorization plumbing across federated environments, and the production handover where the pilot has to operate against the real workflow rather than a sanitized test case.

FIGURE 04 · WHERE PILOTS STALL, BY CAUSE

Integration and governance account for the majority. The model is a small minority.

Programs scoped to the model and not the middleware are scoping to the easy part — and to the layer where pilots rarely fail.



Distribution of federal AI pilot stall causes, FY24–Q4 through FY26–Q1. Percentages reflect FCI's engagement observation; the directional shape is consistent across federal sectors and AI use cases.⁵

WHERE FEDERAL AI PILOTS STALL, BY CAUSE (ILLUSTRATIVE)

Stall cause

Share⁵

Layer

Stall cause	Share ⁵	Layer
Integration	~49%	Data access, transactional consistency, identity plumbing, deploy handover.
Governance	~27%	Audit trail, reversibility, defensibility, authorization.
Deploy handover	~13%	Moving from sanitized test case to the real workflow.
Model	~11%	The reasoning layer — a small minority of stalls.

Distribution of federal AI pilot stall causes, FY24-Q4 through FY26-Q1. Percentages reflect FCI's engagement observation; the directional shape is consistent across federal sectors and AI use cases.⁵

The model layer accounts for a small minority of pilot stalls; integration and governance together account for the majority. The pattern is consistent across federal sectors, agency sizes, and AI use cases. Programs that fund the integration work alongside model selection have substantially higher ship rates than programs that fund the model and treat integration as a downstream concern. This is not surprising once stated — but it is not what most federal AI procurement scoping reflects today.

The model layer accounts for a small minority of pilot stalls. The integration layer is where federal AI program risk actually lives. Programs scoped to the model and not the middleware are scoping to the easy part.

● 05 What to evaluate

Stall analysis should drive vendor evaluation. A procurement that scores vendors primarily on model benchmarks is scoring the layer where pilots rarely fail. A procurement that scores on integration depth is scoring the layer where pilots actually fail. The evaluation criteria below match where the risk lives.

INTEGRATION-LAYER EVALUATION CRITERIA FOR A FEDERAL AI PROCUREMENT

Criterion	What to look for
-----------	------------------

Criterion	What to look for
Federal connector inventory	Pre-built, supported connectors for the agency's actual systems of record — ERP, ECM, identity, mainframe-backed records — not a generic adapter SDK.
FedRAMP boundary handling	How the platform enforces FedRAMP boundaries, data classification, and authorization at the integration layer, at the authorization level the agency requires.
Legacy-system adapter patterns	Demonstrated patterns for the pre-2005 and 2005–2015 cohorts — abstraction layers over systems that predate modern APIs.
Audit-trail completeness	Whether the platform captures an audit trail that spans every system the agent touched in a single decision, queryable months later.
Cross-system transaction handling	Rollback and partial-completion recovery across systems that do not share a transaction manager.
Records-disposition automation	Whether the platform can route AI-generated records into a records-managed environment with schedules attached.

FCI Advisory framework, derived from federal middleware and integration program design.⁴⁵

A vendor with strong integration and content credentials and a competent model partner will outperform a vendor with the reverse — every time. The model is becoming a baseline; the integration depth is the differentiator. The same logic appears in FCI's companion report on federal workforce AI: replace model-benchmark scoring with integration-depth scoring, and the evaluation starts measuring the layer that decides delivery.

● What this rules in and out

Four strategic conditions reshape what federal technology leadership should be funding and evaluating:

- **The middleware layer is where federal agentic AI succeeds or fails.** Foundation-model performance is becoming a baseline; the differentiator is integration depth. The decision to fund the middleware work belongs in the AI program plan, not in a separate IT modernization track.
- **Legacy ERP is the load-bearing constraint.** Federal HR, finance, and case-management ERP systems carry the agency's most consequential workflows and the oldest integration patterns. Agencies deploying agentic AI without a deliberate strategy for the legacy-ERP integration layer are scoping past the constraint that will bind the program.
- **Stall analysis should drive vendor evaluation.** Procurements that score on integration depth, federal connector inventory, FedRAMP boundary handling, and legacy-system adapter patterns are scoring the layer where pilots actually fail.
- **The remediation curve is non-linear.** The middleware work done in the months before a deployment is fundamentally easier than the same work done after a deployment has stalled — post-stall remediation happens under deadline pressure, with active users and accumulated technical debt that pre-deployment work would not have created.

The decision is not whether to deploy agentic AI inside HR, finance, or case-management workflows — the procurements are in market. It is whether the integration and middleware layer underneath is being scoped, funded, and governed as a first-class engineering concern, or whether the program will discover the gap at deployment time and pay for it twice.⁵

Put this thinking to work.

FCI Advisory helps federal agencies design and deliver the integration layer underneath AI — including as an authorized partner for the Boomi Enterprise Platform on GSA. McLean, Virginia.

● Endnotes & sources

1. Federal Enterprise Architecture frameworks document the typical federal ERP and case-management landscape; agency-level systems of record are inventoried in Federal IT Dashboard data and agency enterprise-architecture reports.

2. OMB Memorandum M-25-21, "Accelerating Federal Use of AI Through Innovation, Governance, and Public Trust," April 3, 2025.
3. NIST AI Risk Management Framework (AI RMF 1.0), National Institute of Standards and Technology, January 2023. See also OMB Memorandum M-24-10, "Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence," March 28, 2024.
4. iPaaS-class federal middleware patterns are catalogued through FedRAMP-authorized integration-platform offerings.
5. Aggregate observations cited in this report are drawn from FCI Advisory's engagement base across federal middleware, integration, and agentic AI program design, FY24-Q4 through FY26-Q1. Stall-stage shares, layer shares, and complexity figures are illustrative of the directional pattern, not point-precise extracts.