

# Buy Public Service AI with Evidence You Can Challenge

## *A Public Procurement Management Method for Independent Scrutiny*

Phil Hatch | Akholi

October 23, 2026

Public procurement managers should require evidence that the authority can inspect and challenge independently of the supplier's explanation. Evaluation should test the service's actual function, limitations, decision records, and correction route before dependence becomes material.

## Specify the Evidence Alongside the Service

The procurement manager should identify the public decision or service outcome the agent will support and the evidence the authority needs to defend it. A supplier can demonstrate fluent interaction while leaving the authority unable to explain a consequential action or investigate a complaint.

The UK government AI playbook highlights transparency, legal support, and lifecycle assurance in procurement.<sup>1</sup> It is guidance for its public-sector context. The manager should determine the applicable local procurement and public-law requirements rather than assume one global evaluation model.

---

<sup>1</sup> Government Digital Service, *Artificial Intelligence Playbook for the UK Government* (Government Digital Service, February 10, 2025), *Contractual issues, Intellectual property, including copyright, and Buying AI*, <https://www.gov.uk/government/publications/ai-playbook-for-the-uk-government/artificial-intelligence-playbook-for-the-uk-government-html>.

Define the intended users, permitted actions, affected people, source information, acceptance criteria, and escalation. State the evidence required in the bid and during delivery. Leaving those requirements until contract negotiation can make otherwise attractive offers impossible to compare fairly.

## Separate Supplier Assertion from Observable Performance

Ask bidders to identify which claims are demonstrated, independently tested, modeled, or merely planned. Inspect the conditions and population behind a benchmark. A result from another service or configuration should remain evidence within its stated scope, not proof of the authority's intended use.

The study reports completion signals inconsistent with results and visibility fragmented across firms.<sup>2</sup> These banking accounts do not measure public-sector performance. They identify why procurement should require evidence from authoritative service outcomes rather than rely solely on agent self-reporting.

For a hypothetical permit-support service, test whether the agent uses the correct current rule, identifies missing information, preserves the applicant's request, and routes a reserved decision to an authorized official. A persuasive answer should not compensate for an action outside the service's authority.

## Use a Challengeable Evidence Requirement

Figure 1 proposes four requirements: observable result, reconstructable decision basis, independent review access, and corrective response. Use representative and difficult cases, including ambiguity, prohibited action, failed tools, and human correction. Record limitations in the evaluation conclusion.

Specify secure access to the relevant configuration, records, source materials, and personnel. The authority need not demand unrestricted disclosure of proprietary technology to obtain meaningful evidence. It does need a practical way to challenge material claims and involve authorized reviewers.

---

<sup>2</sup> Phil Hatch, *Agentic AI Outsourcing at Tier 1 Financial Institutions: Full Study Report*, version 1.8.2 (Akholi, forthcoming manuscript dated November 1, 2026), Part 3, AI Technology, sec. 4.1, "Accuracy of the Tools' Own Success and Failure Reports," indicator C172, <https://doi.org/10.2139/ssrn.7164559>; Hatch, *Agentic AI Outsourcing*, Part 2, sec. 5.1, "Agent Actions Visible Only Inside Each Firm's Own Tools."

GSA's AI policy addresses lifecycle governance and accountability within that agency's scope.<sup>3</sup> It is not a universal procurement rule. It provides another example of the need to connect AI acquisition with ongoing operating responsibility rather than treat the purchase as the end of assurance.

Figure 1. The challengeable procurement evidence record

| Evidence requirement | Bid and service demonstration                | Procurement decision                             |
|----------------------|----------------------------------------------|--------------------------------------------------|
| <b>Result</b>        | Authoritative accepted outcome               | Does the service perform the public function     |
| <b>Basis</b>         | Sources authority and material interventions | Can a consequential case be explained            |
| <b>Challenge</b>     | Qualified independent access and review      | Can material claims be tested                    |
| <b>Correction</b>    | Restriction remedy and durable records       | Can the authority respond and remain accountable |

Author's proposed operating framework. Adapt to the service and applicable jurisdiction; not an empirically validated standard.

Source: Phil Hatch, proposed framework developed for this paper.

## Preserve Local Public Accountability

EU, UK, Swiss, US, and other public bodies face different duties concerning administrative decisions, accessibility, equality, privacy, records, transparency, and review. Qualified owners should establish which apply. A supplier's standard contract should not define the authority's public-law responsibilities by default.

Assess whether affected people can understand the relevant process and obtain a usable correction or review route. Distinguish the AI service's recommendation from the authority's legally responsible decision. Internal supplier escalation should not be represented as a statutory review mechanism.

<sup>3</sup> US General Services Administration, "Accelerating Responsible Use of Artificial Intelligence at GSA," US General Services Administration, March 11, 2026, CIO 2185.1C, purpose and applicability, <https://www.gsa.gov/directives-library/accelerating-responsible-use-of-artificial-intelligence-at-gsa>.

Evaluate continuity and successor access. The authority may need to explain historical decisions after the provider relationship ends. Require usable exports, record retention, transition assistance, and appropriate rights, with practical tests rather than broad promises alone.

## Award Against Evidence and Maintain the Right to Challenge

A reasonable objection is that extensive evidence requirements favor large incumbents. Make them proportionate to consequence and publish clear, consistent criteria. Allow different technical approaches to demonstrate the same necessary result. Avoid unnecessary certifications or proprietary requirements that do not answer the public-service question.

Before award, reconcile the proposed service, evaluation findings, evidence rights, operating responsibilities, and price. Identify conditions that must be satisfied before production. During delivery, repeat relevant tests after material change and track correction of significant findings.

The manager should purchase a service the authority can understand and challenge throughout its life. Independent scrutiny is an operating capability supported by records, access, expertise, and action. It cannot be replaced by a favorable demonstration or an assurance statement whose scope excludes the actual public function.

## Evidence and Source Note

The study is a nonrepresentative cross-section of 850 vendor delivery-team respondents on 150 engagements serving Tier 1 financial institutions, collected during January–June 2026. Neither banks nor vendors corroborated the accounts. It establishes reported experience, not causation, legal breach, or market prevalence. Its application here is a mechanism to investigate, not an estimate of conditions in other industries. The author-supplied version 1.8.2 is dated November 1, 2026 and is cited as a forthcoming manuscript consulted October 23, 2026.<sup>4</sup>

This is an analytical decision paper based on the supplied study and selected primary public sources checked through October 23, 2026. The original contribution is a public procurement evidence record linking award evaluation to independent challenge, correction, and durable public accountability. The proposed method has not been independently validated. Legal requirements depend on the entity, activity, jurisdiction, and effective provisions; supervisory observations, research, and professional commentary have distinct authority.

## Bibliography

Government Digital Service. *Artificial Intelligence Playbook for the UK Government*. Government Digital Service, February 10, 2025. <https://www.gov.uk/government/publications/ai-playbook-for-the-uk-government/artificial-intelligence-playbook-for-the-uk-government-html>.

Hatch, Phil. *Agentic AI Outsourcing at Tier 1 Financial Institutions: Full Study Report*. Version 1.8.2. Akholi. Forthcoming manuscript dated November 1, 2026. Author-supplied manuscript consulted October 23, 2026. <https://doi.org/10.2139/ssrn.7164559>.

US General Services Administration. “Accelerating Responsible Use of Artificial Intelligence at GSA.” US General Services Administration. March 11, 2026. <https://www.gsa.gov/directives-library/accelerating-responsible-use-of-artificial-intelligence-at-gsa>.

---

<sup>4</sup> Hatch, *Agentic AI Outsourcing*, Introduction, “Executive Summary” and “Evidentiary Basis and Its Limits.”