Federal Skilled-Trades AI Pilot — Proposal

A demonstration proposal Author: Brad M. Lindsey — independent engineer · Master Electrician · Master HVAC Technician **Version:** 1.0 — 2026-06-02 **Companion to:** the policy white paper *AI-Augmented Skilled Trades* and the one-page brief.

> **Calibration banner.** Every outcome figure in this document is a **target to be measured**, not a result achieved. No physical or operational validation is claimed. The proposal de-identifies all facility and system specifics; the author writes as an independent practitioner and implicates no employer, contract, or named facility.

---

1 — Purpose

Demonstrate, at a single willing federal facility and at low cost, that a **provenance-governed AI system** can preserve critical-infrastructure troubleshooting knowledge as the workforce turns over — without ever presenting an ungoverned or unverifiable claim to an operator. The demonstration is designed to answer the one question that gates any AI deployment near critical systems: *how do we know it isn't confidently wrong?*

2 — Background (one paragraph)

Federal building maintenance backlogs more than doubled to $370 billion (FY2017–FY2024) and are now a GAO High-Risk item; the skilled-trades workforce faces ~2.1 million unfilled positions by 2030.[1][2] The deepest loss is tacit: undocumented, system-specific troubleshooting expertise that retires with the worker and is rebuilt from scratch at every contract turnover. AI can now capture that reasoning in the field — but only governed AI is safe to deploy near safety-critical systems. The methodology underlying this pilot is published and open-source.[3]

3 — Objectives

1. **Capture** expert troubleshooting workflows from veteran tradespeople on a defined set of systems, in auditable form, before that expertise departs.
2. **Convert** captured field experience into a searchable, provenance-anchored knowledge base.
3. **Assist** a less-experienced operator in diagnosing and resolving real faults faster, with every answer carrying its evidence trail.

[1] GAO-25-108400 (2025); GAO Director testimony, House Appropriations, Apr 9 2025; GAO High-Risk Series GAO-25-107743 (2025).

[2] JLL skilled-trades talent research (April 2026); job-growth figures +9.5%/+8.1% through 2034 per JLL (BLS Occupational Outlook rounds to 9%/8%). Figures verified in `FIGURES_VERIFIED_2026-06-02.md`.

[3] *Lindsey Provenance Discipline* (`lindsey-provenance`, MIT, github.com/bradmlindsey/lindsey-provenance); companion papers *One Operator, Nine Trunks, Seven Weeks* and *Zero of Forty-Nine* (Zenodo DOIs per RELEASE_SEQUENCE).

4. **Govern** throughout: no answer is presented as authoritative without its provenance; the system flags low-confidence or unverifiable cases rather than asserting them.

4 — Scope

In scope: one facility; a small, bounded set of building systems (e.g., essential electrical, generator/ATS, HVAC air-handling and VFDs); knowledge capture from designated veteran experts; assisted-diagnosis sessions with designated less-experienced operators; measurement against the metrics in §6.

Out of scope: any role in live energized work; any autonomous action on equipment; any presentation of the AI as a substitute for licensed personnel; enterprise or multi-facility rollout; any procurement commitment.

5 — Candidate site types (generic — no specific facility implied)

- A military or veterans' healthcare facility with NEC 517 essential electrical systems and an aging maintenance workforce.
- A GSA-managed federal building with a documented deferred-maintenance backlog.
- A federal civilian campus facing near-term retirement of key trades personnel.

Selection criteria: a willing facility sponsor; a documented backlog or retirement exposure; at least one veteran expert and one less-experienced operator available to participate; a systems set that is safety-relevant but boundable for a short demonstration.

6 — Measured outcomes (targets, defined in advance)

Metric	How measured	Why it matters
Mean time-to-diagnose	Time for a defined set of common faults, expert baseline vs. assisted-novice	Direct test of whether captured knowledge shortens the gap
Equipment downtime	Downtime on covered systems, pre-pilot baseline vs. pilot period	The operational cost the backlog represents
Training time-to-competency	Time for a new operator to reach a defined competency bar on a covered system	Whether day-one fluency is real
Knowledge-capture coverage	Share of a designated expert's troubleshooting workflows captured in auditable form	Whether the loss is actually being prevented

Metric	How measured	Why it matters
Governance integrity	Rate at which the system correctly flags low-confidence / unverifiable answers instead of asserting them	The trust question, measured directly

Success criteria are set **with the facility sponsor before the pilot begins**, and results are reported whichever way they come out. A null or negative result is a published public good: it bounds what governed AI can and cannot yet do near critical systems.

7 — Governance and safety guardrails

- **Provenance on every claim.** Each captured workflow and each answer carries an append-only, cryptographically anchored evidence trail. Traceable and auditable by facility staff and inspectors.
- **Calibration.** The system distinguishes what it has evidence for from what it is inferring, and says so. A claim advances only as far as its evidence.
- **De-identification.** Facility identity, system specifics, and any sensitive details are scrubbed; the demonstration uses generic descriptors in all reporting.
- **Human-in-command.** The AI assists licensed personnel. It never acts on equipment, never enters the loop on live energized work, and is never presented as authoritative without its evidence attached.
- **Conflict-of-interest hygiene.** The author participates as an independent practitioner. No employer, contract, or specific facility is implicated, and the pilot is structured to keep it that way.

8 — Method and timeline (illustrative)

1. **Setup (weeks 1–2):** sponsor agreement; site and systems selection; baseline measurement; success criteria locked.
2. **Capture (weeks 3–6):** structured sessions with veteran experts; workflows recorded with provenance; knowledge base assembled.
3. **Assist + measure (weeks 7–10):** assisted-diagnosis sessions with less-experienced operators on real faults; metrics collected against baseline.
4. **Report (weeks 11–12):** results compiled, de-identified, and published — including governance-integrity findings — regardless of outcome.

(Timeline is illustrative; it scales to the sponsor's constraints. The pilot is deliberately small.)

9 — Cost posture

The methodology is open-source (MIT) and the toolchain is lightweight. Scope is a single facility and a bounded systems set. The cost to the government is **facility**

access, participant time, and measurement — not a development contract or a software procurement. This is a demonstration, not a deployment.

10 — What success unlocks

If the metrics move and governance integrity holds, the government has a **low-cost, auditable, publishable** method for preserving critical-infrastructure knowledge as its workforce turns over — one that answered the trust question before scaling, instead of after. If they don't, the measurement tells the government something true and useful about the current limits of governed AI near critical systems. Either outcome is information.

Author's disclosure. This proposal and the methodology it rests on are LLM-collaborative work produced under that same provenance discipline — built to its own standard.

Calibration attestation. Every claim here is held to the proof-state on record as of June 2, 2026. Macro figures are cited to public primary sources; every pilot outcome is a target to be measured, not a result achieved. No fabricated figures, vendor-supplied targets, or unvalidated outcome claims appear.