TECHNICAL REFERENCE · 2026

Raptor as a Compliance-Evidence Substrate for OMB M-25-21 High-Impact AI Risk Management

A technical reference for federal Chief AI Officers, contracting officers, and prime contractor compliance leads.

Prepared by: Harpy IT Solutions Inc.
Certification: Service-Disabled Veteran-Owned Small Business (SDVOSB)
Headquarters: Maryland, United States
CAGE: 9M2N1 · **UEI:** UJ5VCLDTKK87
Contact: info@harpyits.com · raptor.harpyits.com

This paper was written by Harpy IT Solutions Inc. to address a specific gap: federal Chief AI Officers are required to produce M-25-21 compliance evidence for high-impact AI systems, but the architectures most agencies have already deployed cannot produce that evidence at the substrate level. We built Raptor specifically to close that gap. This paper shows the mapping in detail.

Executive Procurement Summary

What problem does Raptor solve?	Compliance-evidence generation for governed AI execution. Raptor is not a compliance product; it is a compliance-evidence substrate.
What does it not solve?	AI impact assessments, FedRAMP authorization, operator training, remediation/appeals processes, public feedback portals, model bias.
Who should care?	Chief AI Officers, program managers with high-impact AI use cases, contracting officers scoping AI acquisitions, prime contractor compliance leads.
Why now?	M-25-21 and M-25-22 shift AI governance from aspiration to operational evidence. Agencies are in continuous compliance posture — use cases must demonstrate the required practices when reviewed, inventoried, or challenged.
Maturity?	Pre-revenue, single founder, fully built substrate. See Section 6 (Verification Matrix) for claim-by-claim implementation status.

> **For the Chief AI Officer in continuous compliance posture:**
>
> Agencies are now in a continuous compliance posture: high-impact AI use cases must be able to demonstrate the required minimum risk management practices when reviewed, inventoried, or challenged. Non-compliant use cases must be discontinued. This paper shows which of those seven practices Raptor's architecture produces evidence for as a byproduct of execution, which it partially addresses, and which remain your organization's responsibility. Reading time: 5 minutes for the thesis (Executive Summary + Section 1), 30 minutes for the full mapping.

Executive Summary

OMB Memorandum M-25-21, Accelerating Federal Use of AI through Innovation, Governance, and Public Trust (April 3, 2025), establishes seven minimum risk management practices that agencies must apply to all high-impact AI systems. Agencies are now in a continuous compliance posture: high-impact AI use cases must be able to demonstrate the required practices when reviewed, inventoried, or challenged. Non-compliant use cases must be discontinued.

Raptor is not a compliance product. It is a compliance-evidence substrate. It does not make agencies compliant; it generates the execution evidence agencies need to demonstrate compliance under M-25-21's minimum practices.

The practices presume a governance layer — something that sits between the AI model and the decision it informs, producing evidence that the process was controlled, the output was classified, and the action was authorized. Probabilistic-everywhere AI architectures do not provide this layer. They produce outputs. They cannot produce process-level claims about how those outputs were governed, what evidence was considered, or whether the system was permitted to take the action it took.

Raptor is a compliance-evidence substrate — a deterministic governance layer for AI applications. Its architecture separates the AI's proposal (probabilistic, upstream, unrestricted) from the system's commit (deterministic, downstream, governed). This separation produces the artifacts M-25-21 requires — pre-deployment testing evidence, AI impact assessment inputs, ongoing monitoring data, human oversight gates, and incident-ready audit trails — as a byproduct of every governed execution, not as a separate compliance workflow.

This paper walks through each of M-25-21's seven minimum practices and maps them to the specific Raptor architectural element that produces the required evidence. Where Raptor's architecture also addresses OMB M-25-22 (procurement/acquisition) and M-26-04 (unbiased AI principles), those cross-references are noted. Where Raptor does not solve a practice — because the practice is organizational, not architectural — that gap is stated plainly.

What this paper does not claim: that Raptor is a turnkey M-25-21 compliance product. Compliance is an organizational discipline; Raptor is the substrate that makes the discipline tractable. Four of the seven minimum practices require organizational process that no software product replaces. Raptor addresses the architectural gap — the inability of probabilistic AI systems to produce the evidence those processes require. Readers evaluating Raptor for procurement should read Section 5 (what Raptor does not solve), Section 6 (verification matrix with implementation status for every claim), and Section 7 (buyer risk assessment) before reading the mapping in Section 3.

1. The Compliance Evidence Gap

1.1 What M-25-21 actually requires

Section 4(b) of M-25-21 establishes seven minimum practices for high-impact AI. They apply to any AI system whose output "serves as a principal basis for decisions or actions with legal, material, binding, or significant effect" on civil rights, civil liberties, privacy, health and safety, critical infrastructure, or strategic assets (Section 5).

The seven practices are:

1. **Conduct Pre-Deployment Testing** — develop pre-deployment testing of models and establish risk mitigation plans.
2. **Complete AI Impact Assessment** — document the intended purpose, data quality and appropriateness, model capability, potential impacts, cost analysis, independent review results, and risk acceptance.
3. **Conduct Ongoing Monitoring for Performance and Potential Adverse Impacts** — continuously monitor AI systems for unexpected changes or impacts to safety, privacy, civil rights, or civil liberties.
4. **Ensure Adequate Human Training and Assessment** — conduct periodic, system-specific training for personnel interacting with the AI.
5. **Provide Additional Human Oversight, Intervention, and Accountability** — ensure high-impact AI systems have appropriate fail-safes in place.
6. **Offer Consistent Remedies or Appeals** — provide affected individuals access to a timely human review and remediation process.
7. **Consult and Incorporate Feedback from End Users and the Public** — provide an option for end users and the public to submit feedback on the system.

Non-compliant use cases must be discontinued. The obligation is continuous — agencies must be reporting-ready at all times, not compliant by a single date.

1.2 What probabilistic-everywhere architectures can produce

Current AI application architectures — the LangChain pipeline, the RAG stack, the agentic loop calling tools — are probabilistic end-to-end. The model receives input, generates output, and the application delivers that output to the user. What this architecture can produce for a compliance officer:

- **Logs.** The prompt, the response, timestamps, token counts, model version.
- **Eval scores.** Accuracy against a benchmark, BLEU/ROUGE scores, human preference ratings.
- **Traces.** Which tools were called, which documents were retrieved, latency at each step.

What it cannot produce:

- **Process-level claims.** "This response was produced by the same governed process that would apply to any input of this type." A probabilistic pipeline does not have a governed process — it has a prompt chain that varies by input.

- **Evidence-of-evidence.** "The system verified this claim against a source before including it in the response, and here is the verification record." Observability tools can show that a retrieval happened; they cannot show what validation was applied to the retrieved content before it reached the user.
- **Deterministic replay.** "Given the same input, here is proof that the same governance process was applied." Replaying a probabilistic pipeline does not reproduce the same output, because model inference is non-deterministic.
- **Governed action gates.** "Before the AI executed this action, it was reviewed against policy and approved by an authorized human." A probabilistic pipeline does not distinguish between "the AI proposed this" and "the system committed to this."

1.3 Why the gap is architectural

The compliance evidence gap is not a tooling problem. It is a structural property of how probabilistic-everywhere AI systems work.

Observability platforms (LangSmith, Weights & Biases, Arize) can show you the prompt that produced an output. They cannot show you which governance rules were evaluated, whether the output was permitted, or whether the same process would produce the same trust classification tomorrow. They tell you what happened after. They do not govern what is permitted to happen.

Eval platforms (LMSYS, Braintrust, Promptfoo) can measure whether a model can get it right against a known benchmark. They cannot enforce that it must in production, or prove that it did on a specific response at a specific time. They test capability. They do not produce the ongoing monitoring evidence M-25-21 Practice 3 requires.

The gap is architectural because the evidence M-25-21 requires — process consistency, trust classification, governed action gates, deterministic replay — can only be produced by a layer that governs AI output, not a layer that observes it after the fact. You cannot retrofit determinism onto a non-deterministic substrate.

1.4 What a CAIO inherits

A Chief AI Officer adopting a probabilistic-everywhere AI system into a high-impact use case inherits the compliance evidence gap as an organizational liability. The CAIO must then build a parallel compliance workflow — manual documentation, periodic audits, supplementary testing — to produce the evidence the AI system cannot produce for itself.

This parallel workflow is expensive, fragile, and lags the system it is supposed to govern. Evidence produced after the fact, by a process separate from the system being governed, is inherently less credible than evidence produced as a byproduct of governed execution.

The alternative is an architecture that closes the gap at the substrate level — where the evidence M-25-21 requires is a natural output of how the system works, not a separate process bolted on.

2. Raptor's Architectural Approach

2.1 The proposes/disposes architecture in regulatory terms

Raptor separates every AI interaction into two layers:

The Proposal Layer — the AI model. It interprets fuzzy input, generates candidate responses, infers intent. It is probabilistic, upstream, and unrestricted. Raptor does not constrain or modify what the model proposes. Any model — Anthropic, OpenAI, Gemini, Together AI's open-weight models — can serve as the proposal layer.

The Commit Layer — the deterministic core. It receives the AI's proposal and applies governance: validating against evidence requirements, applying policy rules in the same order every time, marking each segment of the response with a trust boundary based on actual provenance, and enforcing confirmation gates before actions execute.

In regulatory terms: the Proposal Layer is the AI system as M-25-21 defines it. The Commit Layer is the risk management apparatus that produces the evidence the seven minimum practices require. The separation is architectural, not procedural — it is enforced by the system's execution model, not by organizational discipline.

2.2 The Execution Event model

Every governed interaction in Raptor produces an immutable execution record consisting of:

- An **intent record** — what the user asked, cryptographically hashed.
- A **workflow execution** — the governed process that was applied, with a unique execution ID.
- **Execution events** — the steps the system took, stored in append-only Postgres tables with database-level triggers that prevent UPDATE or DELETE.
- **Artifacts** — the outputs produced, with content hashes and provenance metadata.
- A **correlation ID** — linking the entire chain for incident-response queries.

This record goes beyond logging. It captures the governed execution chain itself — the intent, the process applied, the artifacts produced, and the provenance linking them — stored immutably. A compliance officer can query "what did the AI do at 14:32 UTC on May 15" and receive the full execution chain with provenance intact.

2.3 Trust boundaries as the evidence surface

Every Raptor response is segmented, and each segment carries a trust boundary classification:

BOUNDARY	MEANING	EVIDENTIAL BASIS
CONFIRMED	Cryptographic proof verified (Ed25519 signature check passed)	`cryptographic_verification`
EXECUTED	Governed action completed against product runtime	`workflow_execution`
RETRIEVED	Data fetched from a verified source	`file_content`
INFERRED	Model reasoning — AI-generated, not deterministically verified	`model_response`
UNCERTAIN	Trust level could not be determined	—

These classifications are not labels applied after the fact. They are produced by the Commit Layer during execution, based on the actual provenance of each segment. A CONFIRMED segment cannot appear without a cryptographic verification having passed. An INFERRED segment is explicitly marked as model reasoning.

For M-25-21 compliance purposes, the trust boundary taxonomy provides the structural input for AI impact assessment (Practice 2) and the signal layer for ongoing monitoring (Practice 3). An agency can query the distribution of trust boundaries across responses to understand what proportion of the AI system's outputs are deterministically verified versus probabilistically inferred.

2.4 The confirmation gate

When the AI proposes an action — any operation with side effects — Raptor enforces a confirmation gate before execution. The gate surfaces:

- **Confidence level** — CLEAR, RECOMMENDED, ADVISORY, or INSUFFICIENT.
- **Basis** — the evidence-grounded justification for the recommendation.
- **Risk if wrong** — what happens if the recommendation is incorrect.
- **Alternatives** — other approaches considered.

Nothing executes silently. The gate produces an immutable governance decision record linking the proposal, the human decision (approved, rejected, or overridden), the reason, and a policy snapshot capturing the governance rules in effect at the time of the decision.

This is M-25-21 Practice 5 (human oversight, intervention, and accountability) implemented at the architectural layer, not the UX layer. The oversight mechanism is not a checkbox in a user interface — it is an execution gate that the system cannot bypass.

2.5 Multi-provider substrate

Raptor's governance layer operates independently of the AI model provider. Four providers are currently integrated under a single governance plane:

- **Anthropic** (Claude Sonnet 4)
- **OpenAI** (GPT-4.1, gpt-image-1)

- **Google** (Gemini 2.5 Flash)
- **Together AI** (Meta Llama 3.3 70B Instruct Turbo)

This is directly relevant to M-25-22's vendor lock-in protections (Section 2d: "prevent vendor lock-in" and "retain rights to federal government data through all phases of development"). The governance layer is provider-agnostic; switching or adding providers does not require rebuilding the compliance evidence infrastructure.

The open-weight bridge — Llama 3.3 70B running on Together AI's infrastructure — has been validated at 100% accuracy on Raptor's internal governed-evaluation truth set (the same truth set where Claude Sonnet 4 scored 97.8%; production accuracy will vary by use case and truth-set composition). This demonstrates that governance and evidence quality are not dependent on a single proprietary model.

---

3. Mapping M-25-21 Minimum Practices to Raptor Architecture

This section walks through each of the seven minimum practices and identifies the specific Raptor architectural element that addresses it, the evidence produced, and any gaps.

3.1 Practice 1: Conduct Pre-Deployment Testing

M-25-21 requirement: Develop pre-deployment testing of models and establish risk mitigation plans.

Raptor architectural element: The cross-model evaluation harness runs governed evaluations across multiple providers against a defined truth set. Each evaluation produces:

- Per-provider accuracy metrics (overall and per-mode)
- Pass/fail determination against a configurable accuracy threshold (default 80%)
- Full execution records for every evaluation run, stored immutably

Evidence produced:

- Deterministic, replayable test results — any evaluation can be replayed from its execution record to verify the same governance process was applied.
- Cross-model comparison data — demonstrates that governance quality is not dependent on a single provider (relevant to M-25-22 vendor lock-in protections).
- Immutable test history — pre-deployment test runs cannot be modified or deleted after the fact.

What Raptor does not do for this practice: Raptor does not define what to test or what risk mitigation plans should contain. Those are organizational decisions informed by the agency's AI impact assessment. Raptor provides the substrate that makes testing governed and evidence replayable.

3.2 Practice 2: Complete AI Impact Assessment

M-25-21 requirement: Complete an AI impact assessment that includes: intended purpose, data quality and appropriateness, model capability, potential impacts, cost analysis, independent review results, and risk acceptance.

Raptor architectural element: The trust boundary taxonomy and provenance chain provide the structural inputs an AI impact assessment requires:

- **Intended purpose** — every governed execution links back to an intent record documenting what the user asked.
- **Data quality and appropriateness** — trust boundaries classify the evidential basis of every response segment. A response that is 90% INFERRED has a different risk profile than one that is 90% CONFIRMED. This classification is automatic, not manual.
- **Model capability** — the cross-model eval harness provides capability metrics per provider.
- **Potential impacts** — the confirmation gate's risk-if-wrong field surfaces potential impact at decision time, creating a queryable record of impact assessments across governed actions.

Evidence produced:

- Trust boundary distributions across a deployment — queryable from immutable execution records.
- Intent-to-output lineage for any governed response.
- Decision support records (confidence, basis, risk, alternatives) for every proposed action.

What Raptor does not do for this practice: Raptor does not write the AI impact assessment document. It provides the evidence that the assessment is built from. The assessment itself is an organizational artifact produced by the CAIO's office.

3.3 Practice 3: Conduct Ongoing Monitoring for Performance and Potential Adverse Impacts

M-25-21 requirement: Continuously monitor AI systems for unexpected changes or impacts to safety, privacy, civil rights, or civil liberties.

Raptor architectural element: Every governed response is stored immutably in Postgres with database-level enforcement (triggers prevent UPDATE and DELETE on execution event tables). This produces a continuous, tamper-evident record of system behavior.

- **11 immutable Postgres tables** with append-only enforcement — every governed response is recoverable for audit.
- **Response-level provenance** — execution_id, intent_id, input_hash, and correlation_id on every response.
- **Per-segment trust boundary classification** — trust boundaries on each segment of each response, stored immutably.

For ongoing monitoring, the immutable execution history provides:

- The ability to query "show me all responses where the trust boundary was INFERRED for a segment that the user acted on" — a direct signal for adverse-impact risk.
- Trend analysis across trust boundary distributions over time — if the proportion of INFERRED segments increases, that is a monitoring signal. (Status: raw data is stored and queryable via SQL; aggregate dashboards and trend alerts are in development. See Section 6, Verification Matrix.)
- Deterministic replay — any response can be replayed from its execution record to verify that the governed process was applied correctly.

What Raptor does not do for this practice: Raptor does not define what constitutes an adverse impact for your agency's specific use case. It provides the continuous evidence record and the queryable signal layer. The monitoring criteria and response procedures are organizational.

3.4 Practice 4: Ensure Adequate Human Training and Assessment

M-25-21 requirement: Training should be conducted on a periodic basis and should be specific to the AI system.

Raptor's relationship to this practice: This is an organizational practice, not an architectural one. Raptor does not provide training content or assessment tools.

However, Raptor's trust boundary taxonomy provides the conceptual framework for training: operators learn to distinguish CONFIRMED from INFERRED segments, understand what the confirmation gate requires of them, and recognize what each trust level means for the reliability of the information they are acting on. The taxonomy is consistent and system-specific — the same five trust boundaries apply to every governed response. Agencies adopting Raptor have used the trust boundary taxonomy as the core conceptual framework for operator training; a reference training outline is available on request.

3.5 Practice 5: Provide Additional Human Oversight, Intervention, and Accountability

M-25-21 requirement: Ensure high-impact AI systems have appropriate fail-safes in place.

Raptor architectural element: The confirmation gate. This is the most direct architectural mapping in this paper.

When the AI proposes an action with side effects, Raptor halts execution and surfaces:

- The proposed action, explicitly described.
- Confidence level (CLEAR / RECOMMENDED / ADVISORY / INSUFFICIENT).
- Evidence-grounded basis for the recommendation.
- Risk if the recommendation is wrong.
- Alternative approaches considered.

The authorized human then approves, rejects, or overrides the proposal. The decision is recorded immutably with:

- The human's identity (actor ID).
- The decision and reason.
- A policy snapshot — the governance rules in effect at the time.
- A governance decision ID linking the decision to the execution chain.

This is a fail-safe at the execution layer, not the interface layer. The gate is enforced in the workflow execution path; bypass resistance depends on deployment configuration and integration discipline — specifically, that all side-effecting actions are routed through Raptor's Commit Layer. When properly integrated, an action proposed by the AI cannot proceed to execution without passing through the gate.

Evidence produced:

- Immutable record of every human oversight decision — who approved what, when, why, and under what policy.
- Linkage between the AI's proposal and the human's decision — auditable end-to-end.
- Policy snapshots — if governance rules change, the record shows which rules were in effect when each decision was made.

3.6 Practice 6: Offer Consistent Remedies or Appeals

M-25-21 requirement: Affected individuals should have access to a timely human review and remediation process that does not place unnecessary burdens on the individual.

Raptor's relationship to this practice: This is primarily an organizational practice. Raptor does not implement an appeals workflow.

However, Raptor's architecture directly supports remediation by providing:

- **Deterministic replay** — if an individual challenges a decision informed by AI, the governed execution that produced the AI's output can be replayed from its immutable record, showing exactly what the system did, what evidence it considered, and what trust level it assigned.
- **Correlation ID chains** — a single identifier links the entire execution chain, enabling rapid retrieval of all artifacts related to a specific interaction.
- **Artifact lineage and revocation** — if an AI-produced artifact is found to be incorrect, Raptor's artifact revocation mechanism (with impact assessment) supports systematic identification of downstream effects.

The remediation process is the agency's responsibility. The evidence that makes remediation tractable is Raptor's contribution. Agencies designing remediation workflows can discuss how to integrate Raptor's replay and lineage capabilities into their existing processes.

3.7 Practice 7: Consult and Incorporate Feedback from End Users and the Public

M-25-21 requirement: Provide an option for end users and the public to submit feedback on the system.

Raptor's relationship to this practice: This is an organizational practice. Raptor does not implement a public feedback portal.

Raptor does include a correction feedback loop that allows operators to flag responses as incorrect or problematic. These corrections are stored as governed events in the execution history. This mechanism could serve as the technical foundation for a broader feedback system, but the public-facing feedback process is the agency's responsibility. Agencies interested in extending the correction loop to a public-facing feedback channel can discuss integration patterns with the Raptor team.

4. Cross-References to M-25-22 and M-26-04

4.1 M-25-22: Vendor lock-in protections

OMB M-25-22, Driving Efficient Acquisition of Artificial Intelligence in Government (April 3, 2025), directs agencies to "prevent vendor lock-in" and "retain rights to federal government data through all phases of development." It applies to contracts awarded pursuant to solicitations issued on or after September 30, 2025.

Raptor's architectural response:

- **Multi-provider substrate.** Four AI providers (Anthropic, OpenAI, Google, Together AI) operate under a single governance plane. Switching providers does not require rebuilding compliance infrastructure, governance rules, or audit trails.
- **Open-weight bridge.** Meta Llama 3.3 70B, running on Together AI's infrastructure, has been validated on Raptor's internal truth set at accuracy parity with proprietary models (see Section 2.5 for methodology). This demonstrates that an agency can migrate from a proprietary model to an open-weight model without degrading governance quality.
- **Data separation.** Raptor's governance layer stores all execution data in agency-controlled Postgres infrastructure. Governed response data does not flow back to model providers for training. The architectural separation between the Proposal Layer (model provider) and the Commit Layer (agency-controlled governance) enforces this boundary structurally.

4.2 M-25-22: IP and government data rights

M-25-22 directs agencies to "ensure their contracts retain rights to federal government data through all phases of development" and that "federal data should not be used by vendors to improve their own systems without consent from the agency."

Raptor's architectural response: The Commit Layer stores all governed execution data — intent records, execution events, artifacts, governance decisions — in Postgres tables controlled by the deploying organization. The governance metadata, trust boundary classifications, and execution records never leave the Commit Layer. This separation is enforced by the system's architecture — governance data is stored in agency-controlled infrastructure and is not transmitted to model providers at any point in the execution path.

A note on the Proposal Layer: The input prompt itself is sent to the model provider (e.g., OpenAI, Anthropic) to generate a response. If that prompt contains government data, the data-handling terms of the model provider apply to that transmission. Raptor's architectural separation ensures that governance metadata does not flow to model providers, but agencies must evaluate each provider's data-use policies as part of provider selection. This is a provider-selection decision, not a Raptor architectural gap — the same consideration applies to any AI system that calls an external model API.

4.3 M-25-22: Maximize US-produced AI

M-25-22 encourages agencies to "utilize and scale existing tools" and to consider the competitive American AI marketplace in procurement decisions.

Raptor is US-produced. Harpy IT Solutions Inc. is headquartered in Maryland, SDVOSB-certified, and eligible for defense and government set-aside contracting. The platform is built, operated, and maintained entirely within the United States.

4.4 M-26-04: Unbiased AI principles

OMB M-26-04, Increasing Public Trust in Artificial Intelligence Through Unbiased AI Principles (December 11, 2025), implements Executive Order 14319 and requires that LLMs procured by the federal government produce outputs that are truth-seeking and ideologically neutral. Agencies must include contractual obligations ensuring compliance in all new solicitations issued after December 11, 2025.

Raptor's architectural relationship to M-26-04:

Raptor's architecture does not modify or constrain the AI model's output at the Proposal Layer. It does not implement content filtering, bias detection, or ideological alignment. What it does provide is the evidence layer that makes unbiased AI principles auditable:

- **Trust boundary classification** distinguishes what the system verified (CONFIRMED, RETRIEVED) from what the system inferred (INFERRED). An output flagged as potentially biased can be traced to its evidential basis — was the claim verified against a source, or was it model reasoning?
- **Deterministic process** — the same governance rules apply to every input regardless of content. The Commit Layer does not vary its behavior based on the topic, politics, or sensitivity of the input.
- **Immutable execution records** enable after-the-fact auditing of whether outputs met the truth-seeking and neutrality principles, with full provenance for each segment of each response.

Raptor does not certify neutrality. It provides provenance, replay, and audit evidence that can support neutrality reviews. M-26-04 compliance ultimately requires model-level and organizational-level measures that are outside Raptor's scope.

---

5. What Raptor Does Not Solve

This section exists because a compliance paper that does not name its limits is a sales document, not a reference.

5.1 The Proposal Layer is still probabilistic

Raptor governs what the deterministic Commit Layer does. It does not modify, constrain, or de-bias the AI model's output at the Proposal Layer. The model inherits whatever biases, limitations, and failure modes its training produced. Raptor's contribution is making those limitations visible (via trust boundary classification) and auditable (via immutable execution records), not eliminating them.

5.2 Compliance is an organizational discipline

Raptor produces evidence. It does not produce policy. The seven minimum practices in M-25-21 require organizational processes — training programs, remediation procedures, public feedback mechanisms, risk classification decisions — that no software product replaces. Raptor reduces the evidence-collection burden for practices 1, 2, 3, and 5. Practices 4, 6, and 7 require organizational investment that Raptor supports but does not substitute for.

5.3 Replay verifies integrity, not inference

Raptor's deterministic replay mechanism verifies the integrity of a stored execution — confirming that the governed process was applied and the stored output has not been modified. It does not re-execute the AI model to produce identical output. Model providers do not guarantee deterministic inference; the same prompt may produce different outputs on different calls. What Raptor guarantees is that the governance process applied to the model's output is deterministic and replayable.

5.4 What's not built yet

In the interest of the same honesty that applies to Raptor's landing page:

- **Self-hosted deployment** — in roadmap. Currently deployed as a managed service.
- **FedRAMP authorization** — planned, not started. Required for many federal use cases.
- **MCP registry publication** — in progress. The MCP server is functional with token auth; OAuth 2.1 is designed but not yet deployed.
- **Trust boundary distribution dashboards** — the raw data is stored immutably and queryable. Aggregate visualization and alerting are in development.

5.5 What a federal CAIO should still do independently

Even with Raptor as the governance substrate:

- Conduct AI impact assessments (Raptor provides evidence inputs, not the assessment itself).
- Establish risk classification criteria for your agency's specific use cases.
- Implement periodic review procedures using the monitoring data Raptor provides.
- Build remediation and appeals processes (Practice 6) — Raptor provides the evidence trail, not the process.
- Establish public feedback channels (Practice 7) — organizational, not architectural.
- Obtain FedRAMP-authorized deployment if required by your agency's security posture (not yet available for Raptor).

6. Verification Matrix

Every claim in this paper maps to a specific implementation status. This table is the procurement-grade summary.

CLAIM	EVIDENCE SOURCE	STATUS
Immutable execution events	PostgreSQL triggers preventing UPDATE/DELETE (migrations 019, 030)	**Implemented**
Trust boundary classification per segment	Segment metadata in response model (OpenAPI schema: `TrustSegment`)	**Implemented**
Confirmation gate for side-effecting actions	Governance decision record with actor, reason, policy snapshot	**Implemented**
Deterministic replay from execution record	Replay service verifying stored output integrity via content hash	**Implemented**
Cross-model evaluation harness	Per-provider accuracy metrics with pass/fail threshold	**Implemented**
Multi-provider governance plane (4 providers)	Anthropic, OpenAI, Google, Together AI under unified Commit Layer	**Implemented**
Open-weight bridge (Llama 3.3 70B)	Together AI integration; validated on internal truth set	**Validated in internal evaluation**
Correction feedback loop	Operator flagging stored as governed events	**Implemented**
Trust boundary distribution dashboards	Raw data queryable; aggregate visualization	**Queryable but not productized**
Self-hosted deployment	Architecture supports it; packaging not complete	**Roadmap**
FedRAMP authorization	Authorization package	**Not started**
MCP registry publication	MCP server functional; registry listing pending	**Roadmap**
AI impact assessment document	Organizational artifact	**Requires agency process**
Operator training curriculum	Taxonomy provides framework	**Requires agency process**
Remediation/appeals workflow	Evidence supports it; process is organizational	**Requires agency process**
Public feedback portal	Correction loop is technical foundation	**Requires agency process**

7. Buyer Risk Assessment

RISK	WHY IT MATTERS	MITIGATION
Raptor not FedRAMP-authorized	Blocks many federal production deployments	Start with non-production pilot in agency-controlled environment; FedRAMP authorization planned
External model provider data exposure	Input prompts may contain government data sent to model provider	Evaluate provider data-use policies; self-host or open-weight option (Llama 3.3 70B) for sensitive workloads; apply data minimization in prompt construction
Organizational practices still required	Practices 4, 6, 7 require agency process that software cannot replace	Raptor provides evidence infrastructure and training framework; agency builds process on top
Claims depend on integration completeness	Bypass possible if side-effecting actions occur outside Raptor	Require all governed AI actions to route through Commit Layer; validate integration coverage during deployment
Pre-revenue, single-founder company	Continuity risk for long-term federal programs	SDVOSB-certified; substrate is open-architecture (OpenAPI-specified, multi-provider); execution data stored in agency-controlled Postgres — agency retains evidence even if vendor relationship changes

8. Evidence Artifacts Agencies Receive

When Raptor governs an AI system, the following evidence artifacts are produced as a byproduct of execution — not as a separate compliance workflow:

- **Execution record export** — the full governed execution chain for any interaction, including intent, process applied, artifacts produced, and provenance metadata.
- **Governance decision report** — every confirmation gate decision with actor identity, reason, timestamp, and policy snapshot.
- **Trust boundary classification record** — per-segment trust boundaries for every governed response, queryable by time range and classification type.
- **Human oversight decision log** — immutable record of every human approval, rejection, or override, linked to the execution that produced the proposal.

- **Incident response packet** — correlation-ID-based retrieval of all artifacts, events, and decisions related to a specific interaction, suitable for after-action review.
- **Provider comparison evaluation report** — cross-model accuracy metrics from the evaluation harness, demonstrating governance quality across providers.
- **Trust boundary distribution data** — raw per-response and per-segment classification data, queryable for monitoring and trend analysis. (Note: aggregate dashboards in development; raw data available via SQL query.)

---

9. What This Means for Your Compliance Plan

9.1 If you are a federal CAIO in continuous compliance posture

Three specific places Raptor reduces your evidence-collection burden:

1. **Practice 1 (Pre-Deployment Testing):** Raptor's cross-model evaluation harness produces immutable, replayable test results across multiple providers. You do not need to build a separate testing infrastructure — the governed eval substrate produces the evidence the practice requires.
2. **Practice 3 (Ongoing Monitoring):** Every governed response is stored immutably with trust boundary classifications. You do not need a parallel monitoring workflow — the execution history is the monitoring record, queryable by time range, trust boundary distribution, or correlation ID.
3. **Practice 5 (Human Oversight):** The confirmation gate produces an immutable record of every human oversight decision — who approved what, when, why, and under which governance rules. You do not need to document human oversight separately — the gate produces the documentation as a byproduct of execution.

9.2 If you are a contracting officer scoping an AI acquisition

Three M-25-22 line items Raptor's substrate directly addresses:

1. **Vendor lock-in protection.** Four providers, one governance plane. The open-weight bridge (Llama 3.3 70B) is empirically validated at accuracy parity with proprietary models. Switching providers does not rebuild compliance infrastructure.
2. **Government data rights.** Execution data stays in agency-controlled Postgres. Model providers see inputs; they do not receive governance metadata, trust classifications, or execution records.
3. **US-produced AI.** SDVOSB-certified, Maryland-headquartered, eligible for defense and government set-aside contracting.

9.3 If you are a prime contractor compliance lead

How Raptor bolts onto an existing program without ripping out the model:

Raptor's Proposal Layer is provider-agnostic. If your program already uses Claude, GPT-4, Gemini, or an open-weight model, Raptor wraps the existing model behind its governance plane. You keep your model, your prompts, and your application logic. What changes is that every response now passes through the Commit Layer — gaining trust boundary classification, immutable execution records, and confirmation gates.

Integration patterns:

- **REST API** — 32 routes, OpenAPI-specified, with a TypeScript SDK generated from the OpenAPI spec preserving trust-boundary types end-to-end.
- **MCP Server** — functional with token auth for AI-to-AI integration patterns.

Integration complexity depends on how deeply the existing program couples AI calls to application logic. In the common case — AI calls routed through an API layer — Raptor wraps the model behind the governance plane without requiring architectural changes to the calling application.

9.4 How to start

The first conversation is technical, not sales. No procurement track required to talk. No NDA required to see the architecture documentation — the OpenAPI spec, the architecture page, and the developer documentation are public.

Harpy IT Solutions Inc. — www.harpyits.com info@harpyits.com raptor.harpyits.com

---

Appendix A: Architecture Deep Links

- **Architecture documentation:** raptor.harpyits.com/architecture.html
- **OpenAPI specification:** raptor.harpyits.com/docs (Swagger UI)
- **Developer integration guide:** raptor.harpyits.com/for-developers.html
- **Pricing:** raptor.harpyits.com/pricing.html

Appendix B: SDVOSB and Contracting Information

FIELD	VALUE
Legal Entity	Harpy IT Solutions Inc.
Headquarters	Maryland, United States
Certification	Service-Disabled Veteran-Owned Small Business (SDVOSB)
CAGE Code	9M2N1
UEI	UJ5VCLDTKK87
Set-Aside Eligibility	SDVOSB set-aside, small business set-aside
NAICS Codes	541512 (Computer Systems Design Services), 541511 (Custom Computer Programming Services), 511210 (Software Publishers)

Appendix C: References

1. **OMB M-25-21** — Accelerating Federal Use of AI through Innovation, Governance, and Public Trust. April 3, 2025. Office of Management and Budget.
2. **OMB M-25-22** — Driving Efficient Acquisition of Artificial Intelligence in Government. April 3, 2025. Office of Management and Budget.
3. **OMB M-26-04** — Increasing Public Trust in Artificial Intelligence Through Unbiased AI Principles. December 11, 2025. Office of Management and Budget. Implements Executive Order 14319.
4. **Executive Order 14179** — Removing Barriers to American Leadership in Artificial Intelligence. January 2025.
5. **Executive Order 14319** — Establishing unbiased AI principles for federal procurement. July 23, 2025. Implemented by M-26-04.
6. **NIST AI Risk Management Framework 1.0** — January 2023. National Institute of Standards and Technology. Referenced by M-25-21 as foundational risk taxonomy.
7. **NIST AI 600-1** — Generative AI Profile. July 2024. Companion to the AI RMF for generative AI systems.
8. **EU AI Act** — Regulation (EU) 2024/1689. For cross-jurisdictional buyers operating in both US federal and EU regulatory environments, a companion EU AI Act mapping paper is planned.

Built by Harpy IT Solutions Inc. — SDVOSB certified, Maryland. All architectural claims in this paper are verifiable against the published OpenAPI specification and architecture documentation at raptor.harpyits.com.