

Zero-Dollar Token AI

Rapid, compliant path to deploy production-ready AI platform - spanning predictive, generative, and agentic capabilities

Thank you for your interest in exploring this content.

Carahsoft is the **Trusted Government IT Solutions Provider**® supporting a broad portfolio of industry-leading technologies through NASA SEWP V, ITES-SW2, NASPO ValuePoint and a wide range of other contract vehicles.

As the **Master Government Aggregator**®, Carahsoft connects government agencies, industry partners, and technology providers to deliver innovative, mission-focused solutions.

In partnership with H2O.ai, we provide technology solutions that drive modernization, strengthen operations, and ensure compliance with evolving government standards.



To learn more about how Carahsoft can support your technology needs, please visit carahsoft.com



Explore More Resources:
carah.io/H2O.ai-resources



Join Events & Webinars:
carah.io/H2O.ai-events



Discover Technology Solutions:
carah.io/H2O.ai



Learn About Procurement:
carah.io/H2O.ai-contracts



Connect With Our Team:
H2O@carahsoft.com
(888) 606-2770



ZERO \$ TOKEN AI

Seven use cases inside for your review

THIS PROPOSAL OFFERS A RAPID, COMPLIANT PATH TO DEPLOY A PRODUCTION-READY AI PLATFORM— SPANNING PREDICTIVE, GENERATIVE, AND AGENTIC CAPABILITIES— BEFORE END-OF-YEAR FUNDING DEADLINES, BUNDLED WITH A 12-MONTH ENABLEMENT PROGRAM AND EMBEDDED FORWARD DEPLOYED ENGINEER SERVICES TO ENSURE MISSION-ALIGNED OUTCOMES.

Carahsoft, Cygnus, H20.ai and K3 Clouds

Re: Unsolicited Proposal — H2O.ai Enterprise AI Platform: End-of-Year Rapid Deployment Options

Carahsoft, Cygnus, K3 Clouds, and H2O.ai respectfully submit this unsolicited proposal as a coordinated team that brings together the software publisher, public-sector distribution, reseller support, and implementation leadership needed to move quickly from acquisition through deployment. Carahsoft serves as the public-sector distributor and contract vehicle leader, Cygnus provides reseller flexibility and customer engagement support, K3 Clouds leads solution architecture and delivery as the systems integrator and H2O.ai consulting practice, and H2O.ai provides the enterprise AI platform and product expertise.

Together, this team offers federal agencies a practical path to adopt predictive, generative, and agentic AI within the current fiscal year while maintaining security, compliance, and mission alignment. Across acquisition planning, platform licensing, solution deployment, training, and scale-out support, the team is structured to deliver both the technology and the hands-on execution required to achieve measurable outcomes quickly.

This proposal outlines H2O.ai's capabilities to rapidly deploy a production-ready enterprise AI platform — spanning predictive, generative, and agentic AI — within your agency's remaining fiscal year timeline. Carahsoft believes that H2O.ai's Enterprise AI Suite would be an ideal solution for agencies seeking to operationalize AI before end-of-year funding deadlines while maintaining full compliance with federal security requirements.

H2O.ai provides a fully sovereign, FedRAMP High Authorized deployment in which no agency data ever leaves your authorization boundary. The platform is equally available for on-premises/air-gapped deployment or via FedRAMP High cloud environments — giving your agency maximum flexibility regardless of infrastructure posture.

This proposal presents two deployment options (On-Premises and FedRAMP High Cloud) at three software licensing levels (\$250K, \$500K, and \$1M) each bundled with a corresponding Forward Deployed Engineer (FDE) and 12-month enablement Statement of Work with defined milestones, deliverables, hours, and FTE allocations. Packages are sized to match agency scale, user count, and use-case complexity.

Carahsoft is submitting as a top-ranked GSA Multiple Award Schedule (MAS) Contractor (47QSWA18D008F), and the Public Sector reseller for H2O.ai Software. As a top-ranked GSA MAS Contractor, Carahsoft has delivered best value solutions to government clients for over 18 years.

Please feel free to contact us directly at 703-431-1224 / jason.richards@carahsoft.com 571-719-8256 / kent.warren@k3clouds.com / 571-533-7726 / jeffery.phelan@h20.ai 804-513-1779 doug.cornell@cygtech.us / with any questions or communications about our proposed solution.

Thank you for your consideration,

Jeff Phelan, Jason Richards, Kent Warren & Doug Cornell
H20.ai | H20 Team Lead Carahsoft | GM K3 Clouds, LLC. | CRO Cygnus

Table of Contents

Executive Summary	1
Proposal Audience	1
Objectives	2
Key Problems Solved	2
Benefits	2
Deployment Options	4
Option A: On-Premises Deployment	4
Option B: FedRAMP High Cloud Deployment	4
Software Licensing Tiers	5
Representative AI Use Cases for Federal Agencies	5
Use Case 1: Intelligent Document Processing & Compliance Review	5
Use Case 2: Predictive Maintenance & Supply Chain Optimization	6
Use Case 3: Fraud, Waste, and Abuse Detection	7
Use Case 4: Knowledge Management & Institutional Intelligence	8
Use Case 5: Risk Scoring & Decision Support	9
Use Case 6: Automated Reporting, Briefing Generation & Workflow Acceleration	10
Use Case 7: SIEM Cost Optimization & Security Data Triage	11
Applicable Agencies	13
Use Case Summary Matrix	13
Use Case Alignment to Package Tiers	14
Bundled Solution Packages (Software + Services)	14
Package 1: Small — \$350,000 (12 Months)	15
Package 2: Medium — \$700,000 (12 Months)	16
Package 3: Large — \$1,400,000 (12 Months)	18
12-Month Enablement Statement of Work	21
SOW Scope, Milestones, and Deliverables	21
Deliverable Summary by Package	22
Hours and FTE Summary	22
FDE Role, Responsibilities, and Qualifications	22
Acceptance Criteria	24
Security and Compliance	25
Company Backgrounds	25
Solution Provider: H2O.ai	25

H2O.ai Zero-Dollar Token AI Unsolicited Proposal

Distributor and contract vehicle leader: Carahsoft Technology Corp.	25
Implementation: K3 Clouds (H2O.ai and Cribl)	26
Reseller: Cygnus.....	26
In Summary.....	27

Executive Summary

Carahsoft Technology Corp. understands that federal agencies face a critical window to obligate end-of-fiscal-year funds on high-impact modernization initiatives. Carahsoft and H2O.ai are proposing the H2O.ai Enterprise AI Suite — a unified platform for predictive, generative, and agentic AI — deployable within weeks and designed to deliver measurable mission value within the first 90 days.

Each package in this proposal is a single, bundled Time and Material procurement combining H2O.ai software licensing with a fully defined 12-month enablement Statement of Work (SOW). The SOW includes a dedicated Forward Deployed Engineer (FDE), defined milestones, measurable deliverables, and acceptance criteria — providing the contracting officer with a clear basis for award and performance evaluation.

Key features of the H2O.ai Enterprise AI Suite include:

- Predictive AI (Driverless AI / H2O-3) — Automated machine learning for forecasting, classification, anomaly detection, and risk scoring with full explainability.
- Generative AI (Enterprise h2oGPTe) — Enterprise-grade LLM/RAG platform supporting 50+ languages, 80+ file types, document intelligence, and multi-model orchestration.
- Agentic AI (H2O Vertical Agents) — Autonomous multi-step reasoning agents with tool calling, function execution, and human-in-the-loop oversight for complex workflows.
- Forward Deployed Engineer (FDE) — A senior AI engineer embedded directly in your agency environment to design, deploy, integrate, tune, and sustain AI solutions in production.
- 12-Month Enablement SOW — Structured milestones, deliverables, knowledge transfer, and adoption acceleration with defined acceptance criteria.
- FedRAMP High Authorized & On-Premises — Fully sovereign deployment with no data leaving your authorization boundary; FIPS-validated cryptography; SOC 2 Type 2.

Proposal Audience

These representative AI use cases align directly with the Department of Homeland Security's mission to safeguard the nation, secure borders and cyberspace, and enhance resilience against threats and disasters. Intelligent document processing and compliance review can accelerate adjudication of immigration benefits, enforcement case files, grants, and regulatory actions, reducing backlogs while improving consistency with complex statutory and policy requirements that drive fair and timely outcomes for individuals and partners. Predictive maintenance and supply chain optimization can strengthen the reliability of critical DHS assets—such as screening equipment, vehicles, sensors, and communications infrastructure—by forecasting failures and optimizing spares and logistics, helping border, transportation, and emergency operations stay mission-ready. Risk scoring and decision support can prioritize inspections, enforcement actions, vetting, and threat investigations with transparent, explainable models, ensuring limited investigative and operational resources are focused on the highest-risk individuals, cargo, facilities, and cyber threats. Automated reporting and briefing generation can turn operational data and intelligence into ready-to-use reports and situational briefings for leadership, Congress, interagency partners, and state and local stakeholders, reducing administrative burden on

frontline staff so they can spend more time on operations. SIEM cost optimization and security data triage can enhance DHS's cyber defense posture across vast hybrid environments by improving signal-to-noise in security operations while controlling SIEM costs, enabling DHS components to preserve forensic visibility and compliance while reinvesting savings into mission-critical prevention, protection, and response capabilities.

Objectives

H2O.ai proposes to rapidly deploy its Enterprise AI Suite within your agency's environment — delivering production AI capabilities before end-of-year funding deadlines. The objectives are:

- Deploy a production-ready enterprise AI platform (predictive, generative, and agentic) within 30–60 days of contract award.
- Deliver initial AI use cases into production within the first 90 days, aligned to agency mission priorities.
- Embed a Forward Deployed Engineer to accelerate adoption, ensure production outcomes, and build internal agency AI capability.
- Provide a 12-month enablement program with defined milestones that transitions AI ownership to agency personnel.
- Establish a governed, compliant AI foundation that supports expansion in subsequent fiscal years.

Key Problems Solved

The H2O.ai Enterprise AI Suite addresses the following common federal agency challenges:

- End-of-year funding deadlines — Agencies need to obligate funds on high-value AI initiatives quickly, with minimal procurement friction and rapid time-to-value.
- AI pilot-to-production gap — Many agencies have experimented with AI but lack the embedded engineering support to move from pilot to sustained production operations.
- Data sovereignty and compliance — Agencies require AI platforms that operate entirely within their authorization boundary, meeting FedRAMP High, FISMA, and agency-specific security requirements.
- Workforce AI readiness — Agency staff need hands-on enablement and knowledge transfer to sustain AI systems independently after initial deployment.
- Fragmented AI tooling — Agencies often have disconnected point solutions for ML, NLP, and document processing; H2O.ai unifies predictive, generative, and agentic AI under one governed platform.
- Scalability from pilot to enterprise — Agencies need licensing models that scale without per-user or per-query fees, enabling broad adoption without budget surprises.

Benefits

The H2O.ai Enterprise AI Suite will help your agency:

- Obligate end-of-year funds on a proven, FedRAMP High Authorized AI platform with rapid deployment timelines (30–60 days to production).
- Accelerate mission outcomes by deploying AI use cases in the first 90 days with embedded FDE support.

H2O.ai Zero-Dollar Token AI Unsolicited Proposal

- Eliminate per-user and per-query licensing fees — H2O.ai licenses by provisioned capacity (GPU memory) with unlimited users.
- Maintain full data sovereignty — all processing occurs within your authorization boundary; H2O.ai does not train on customer data.
- Build lasting internal AI capability through structured 12-month enablement with defined milestones and knowledge transfer.
- Establish a governed AI foundation compliant with OMB M-24-10, NIST AI RMF, and agency-specific AI policies.
- Scale from initial deployment to enterprise-wide AI adoption in subsequent fiscal years without re-architecture or re-licensing.
- Proposed Solution — The H2O.ai Platform

H2O.ai proposes its Enterprise AI Suite — a unified platform spanning predictive machine learning, generative AI with retrieval-augmented generation, and autonomous agentic AI — as the foundation for your agency's AI modernization.

Included Modules	Supported Use Cases	Description
Predictive AI	Fraud, Waste, and Abuse; Insider Threat; Cyber Automation; Case Management; Burden Reduction; Document Modernization; Supply Chain Optimization	Automated machine learning — classification, regression, time-series forecasting, anomaly detection, NLP. Full explainability, automatic feature engineering, and model governance.
Generative AI	Enterprise Rag; Document Management & Automation; Chat Bots; Policy & Procedures Analysis; Agency Modernization; Citizen Services Automation	Enterprise LLM/RAG platform — 50+ languages, 80+ file types, document intelligence, multi-model orchestration, fine-tuning, prompt engineering, and guardrails.
Agentic AI	Systems Automation & Orchestration; Advanced Research; Data & API integrations; Cross application reasoning; Decision Reasoning, Automation & Routing	Autonomous multi-step reasoning agents with tool calling, function execution, MCP integration, and human-in-the-loop oversight for complex mission workflows.

Licensing Model

H2O.ai is licensed on a capacity basis (by provisioned GPU memory) with:

- No per-user fees — unlimited named users across the agency
- No per-query or per-transaction fees
- No token-based pricing
- Standard annual support and maintenance included
- Pricing identical across deployment models (cloud, on-premises, air-gapped)

Deployment Options

H2O.ai supports two primary deployment configurations for federal agencies. Both options deliver identical platform capabilities and are priced equivalently — the choice depends on your agency's infrastructure posture, security requirements, and operational preferences.

Option A: On-Premises Deployment

For agencies requiring full physical control of AI infrastructure, H2O.ai deploys directly on agency-owned or agency-controlled hardware. This option supports fully air-gapped environments with no external network connectivity required after initial installation.

- Deployed on agency-owned GPU servers (validated designs on Dell AI Factory with NVIDIA, Cisco infrastructure, or equivalent)
- Fully air-gapped capable — no internet connectivity required for operation
- Container-based deployment (Docker, Kubernetes, Helm) for portability and repeatability
- Supports DoD SRG IL4/IL5 environments
- Agency retains full physical and logical control of all data, models, and infrastructure
- Local LLM inference (vLLM, Ollama) with no external API calls
- Compatible with agency SIEM, ITSM, and observability tooling

Option B: FedRAMP High Cloud Deployment

For agencies leveraging authorized cloud environments, H2O.ai deploys within FedRAMP High Authorized infrastructure (AWS GovCloud, Azure Government, or Google Cloud). H2O.ai itself holds FedRAMP High Authorization, ensuring the full platform stack meets the highest civilian security baseline.

- H2O.ai platform is FedRAMP High Authorized (not just "FedRAMP Ready")
- Deployed within agency's Virtual Private Cloud — data never leaves authorization boundary
- Native support for AWS GovCloud, Azure Government, and Google Cloud (Vertex AI)
- SOC 2 Type 2 certified; FIPS-validated cryptography
- Rapid provisioning — production environment available within 2–3 weeks of contract award
- Elastic scaling for burst workloads without hardware procurement
- Managed platform updates and security patching within authorization boundary
- Available through AWS Marketplace, Azure Marketplace, and GCP Marketplace for streamlined procurement

Software Licensing Tiers

H2O.ai offers three software licensing tiers representing Small, Medium, and Large platform configurations. Each tier is a 12-month annual subscription licensed by provisioned GPU memory capacity with unlimited users and no per-query fees.

Tier	Annual License	GPU Memory	Platform Modules Included	Ideal For
SMALL	\$250,000	Base allocation Up to 8 GPUs	Enterprise h2oGPTe (Generative/RAG) Driverless AI (Predictive ML) Document AI Standard Support	Pilot, single-division
MEDIUM	\$500,000	Enhanced allocation Up to 16 GPUs	All Small PLUS: H2O Vertical Agents (Agentic AI) MLOps (Governance) LLM Studio (Custom Models) Premium Support	Multi-division, production agentic AI
LARGE	\$1,000,000	Maximum allocation Up to 32 GPUs	All Medium PLUS: H2O Wave (AI App Layer) Custom Agent Development Advanced Fine-Tuning Enterprise Support + Dedicated CSM	Enterprise-wide transformation

Representative AI Use Cases for Federal Agencies

The following seven use cases represent high-impact, cross-cutting AI applications that are relevant across Federal Civilian agencies, the Department of Defense, and the Intelligence Community. Each use case leverages the H2O.ai Enterprise AI Suite — combining predictive, generative, and agentic AI capabilities — to automate manual processes, accelerate decision-making, and modernize mission operations. These use cases are designed to demonstrate rapid time-to-value within the first 90 days of deployment and can be tailored to any agency's specific data, policies, and mission requirements.

Use Case 1: Intelligent Document Processing & Compliance Review

Problem

Federal agencies process millions of documents annually — contracts, invoices, applications, reports, correspondence, and regulatory filings. Manual review of these documents is slow, error-prone, and resource-intensive. Staff spend days reviewing 100–250 page document packages, cross-referencing them against regulations, contract terms, and policy requirements. Backlogs grow, compliance deadlines are missed, and improper payments go undetected. Agencies face Prompt Payment Act penalties, audit findings, and mission delays when document review cannot keep pace with volume.

Solution

H2O.ai's Enterprise h2oGPTe and Document AI ingest, extract, and reason across complex multi-page document packages autonomously. The agentic AI platform performs multi-step compliance review — extracting key data from invoices, timesheets, receipts, and applications; cross-referencing every line item against governing regulations (FAR, DFARS, 2 CFR Part 200, agency-specific policies); flagging discrepancies with verbatim regulatory citations; and producing a defensible approval memo or exception report for human sign-off.

H2O.ai Platform Components Used

- Enterprise h2oGPTe — Agentic reasoning, RAG over policy corpus, multi-step audit logic
- H2O Document AI — OCR, extraction from scanned/native documents (80+ file types)
- H2O Vertical Agents — Autonomous document audit agents with tool calling and human-in-the-loop
- Driverless AI — Anomaly detection for duplicate/fraudulent submissions

Benefits & Value

- Reduce document review time from days to minutes (demonstrated 90-second review of 120–250 page packages)
- Eliminate or sharply reduce Prompt Payment Act interest penalties and audit findings
- Catch anomalies that manual review routinely misses — duplicates, overages, policy violations, math errors
- Ground every finding in verbatim federal authority for defensible, auditable decisions
- Free skilled staff from paperwork to focus on mission-critical judgment calls
- Scale across the entire agency without adding headcount

Applicable Agencies

Civilian: FEMA (PA invoice audit), CMS (claims review), GSA (contract compliance), IRS (return processing), VA (benefits applications). DoD: DCAA (contract audit), DLA (procurement review), Service branches (travel vouchers). IC: Contract compliance review, FOIA processing, source report triage.

Use Case 2: Predictive Maintenance & Supply Chain Optimization

Problem

Federal agencies operate vast fleets of equipment, vehicles, aircraft, ships, and infrastructure — from military weapons systems to civilian facility HVAC systems. Unplanned equipment failures cause mission disruption, safety hazards, and costly emergency repairs. Traditional time-based maintenance schedules either replace parts too early (wasting resources) or too late (causing failures). Meanwhile, supply chains struggle with demand forecasting — leading to stockouts of critical parts, excess inventory of slow-moving items, and billions in waste. Maintenance logs, sensor data, and supply records often exist in disconnected systems with no unified analytics.

Solution

H2O.ai's Driverless AI builds predictive models on equipment sensor data, maintenance history, and operational conditions to forecast failures before they occur — enabling condition-based maintenance that maximizes equipment availability while minimizing cost. For supply chain, the platform forecasts demand by part number, location, and time horizon, optimizing inventory levels and procurement timing. Agentic AI can autonomously

monitor sensor streams, trigger maintenance work orders, and generate procurement recommendations when stock levels approach reorder points.

H2O.ai Platform Components Used

- Driverless AI — Time-series forecasting, anomaly detection, classification (failure prediction), automatic feature engineering from sensor data
- Enterprise h2oGPTe — Natural language querying of maintenance logs; automated report generation
- H2O Vertical Agents — Autonomous monitoring agents that trigger alerts and work orders
- H2O MLOps — Model lifecycle management, drift detection, automated retraining on new sensor data

Benefits & Value

- Reduce unplanned equipment downtime by 25–40% through early failure prediction
- Extend equipment service life by replacing parts based on condition rather than schedule
- Optimize inventory levels — reduce excess stock by 15–30% while eliminating critical stockouts
- Improve demand forecast accuracy by 20–50% over traditional methods
- Reduce emergency procurement costs and expedited shipping expenses
- Provide commanders and facility managers with real-time readiness visibility

Applicable Agencies

DoD: DLA (supply chain forecasting), Army/Navy/Air Force (fleet/weapons system maintenance), TRANSCOM (logistics). Civilian: GSA (federal building systems), FAA (aviation infrastructure), DOE (facility maintenance), VA (medical equipment). IC: Facility infrastructure, SIGINT/GEOINT collection platform maintenance.

Use Case 3: Fraud, Waste, and Abuse Detection

Problem

The federal government loses an estimated \$175–\$250 billion annually to improper payments, fraud, and abuse across benefit programs, grants, contracts, and healthcare. Traditional rule-based detection systems catch only known fraud patterns and generate excessive false positives that overwhelm investigators. Sophisticated fraud schemes evolve faster than rules can be updated. Agencies lack the ability to detect subtle anomalies across millions of transactions, identify emerging fraud networks, or prioritize investigations by likely recovery amount. Inspector General offices are understaffed relative to the volume of suspicious activity.

Solution

H2O.ai's Driverless AI builds machine learning models that learn complex fraud patterns from historical data — detecting anomalies that rule-based systems miss. The platform identifies suspicious transactions, scores them by risk and estimated dollar impact, and clusters related activities to reveal fraud networks. Enterprise h2oGPTe generates investigator-ready case summaries with supporting evidence, while agentic AI autonomously monitors transaction streams in real-time, escalating high-confidence findings to human investigators with full explainability of why each transaction was flagged.

H2O.ai Platform Components Used

- Driverless AI — Anomaly detection, classification (fraud/not fraud), clustering (network detection), automatic feature engineering across transaction data
- Enterprise h2oGPTe — Automated case narrative generation; natural language investigation queries
- H2O Vertical Agents — Real-time transaction monitoring agents with escalation logic
- H2O MLOps — Model governance, bias monitoring, explainability reporting for legal defensibility

Benefits & Value

- Detect 2–5x more fraudulent transactions than rule-based systems alone
- Reduce false positive rates by 40–60%, freeing investigators to focus on real cases
- Identify emerging fraud patterns and networks before they scale
- Prioritize investigations by estimated recovery amount for maximum ROI
- Generate audit-ready case documentation automatically
- Recover millions in improper payments annually (typical federal ROI: \$10–\$50 returned per \$1 invested in detection)

Applicable Agencies

Civilian: CMS/HHS (Medicare/Medicaid fraud), IRS (tax fraud), SBA (loan fraud), SSA (benefits fraud), USDA (SNAP/crop insurance fraud), OPM (benefits abuse). DoD: DCAA (contractor fraud), DHA (TRICARE fraud), Service IGs. IC: Financial intelligence, sanctions evasion detection, counterintelligence anomaly detection.

Use Case 4: Knowledge Management & Institutional Intelligence

Problem

Federal agencies accumulate vast repositories of institutional knowledge — policy documents, SOPs, technical manuals, after-action reports, legal opinions, research papers, and lessons learned — often spanning decades and millions of pages across disconnected systems. When employees need answers, they spend hours searching through SharePoint, shared drives, and legacy systems, often failing to find relevant information or unknowingly duplicating prior work. As experienced staff retire (the "silver tsunami"), critical institutional knowledge is lost. New employees face months-long onboarding curves. Decision-makers lack the ability to quickly synthesize information across the full breadth of agency knowledge.

Solution

H2O.ai's Enterprise h2oGPTe creates a secure, agency-wide AI knowledge assistant that ingests and indexes the full corpus of agency documents — regardless of format (PDFs, Word, Excel, emails, presentations, scanned documents, databases). Users ask questions in natural language and receive accurate, cited answers grounded in authoritative agency sources. The platform supports 50+ languages and 80+ file types, with full retrieval-augmented generation (RAG) ensuring answers are always traceable to source documents. All processing occurs within the agency's authorization boundary with no data leaving sovereign control.

H2O.ai Platform Components Used

- Enterprise h2oGPTe — Enterprise RAG, multi-model orchestration, citation-grounded responses, 50+ languages, 80+ file types

- H2O Document AI — Ingestion and indexing of legacy documents, scanned materials, and diverse formats
- H2O Wave — Custom AI-powered search and Q&A applications for agency end users
- LLM Studio — Fine-tuning on agency-specific terminology, acronyms, and domain knowledge

Benefits & Value

- Reduce time-to-answer from hours of searching to seconds of AI-assisted retrieval
- Preserve institutional knowledge as experienced staff retire — capture expertise in queryable AI systems
- Accelerate new employee onboarding by 50–70% with instant access to institutional knowledge
- Eliminate duplicated effort by surfacing prior work, decisions, and lessons learned
- Ensure all answers are grounded in authoritative sources with full citations (no hallucination)
- Support multilingual operations for agencies with global missions
- Maintain full data sovereignty — all knowledge stays within agency authorization boundary

Applicable Agencies

Civilian: Every agency (policy/SOP retrieval), State Dept (diplomatic cables/precedent), DOJ (legal research), HHS (clinical guidelines), EPA (regulatory history). DoD: All services (doctrine, TTPs, lessons learned, technical orders), Joint Staff (planning references), acquisition commands (contract precedent). IC: All-source intelligence retrieval, analytic tradecraft references, historical case research.

Use Case 5: Risk Scoring & Decision Support

Problem

Federal agencies make thousands of high-stakes decisions daily — approving benefits claims, granting security clearances, prioritizing inspections, allocating disaster response resources, assessing threat levels, and evaluating grant applications. These decisions are often made with incomplete information, inconsistent criteria, and human cognitive biases. Manual scoring processes are slow and cannot scale to meet demand. Agencies struggle to consistently apply complex, multi-factor risk criteria across large populations. The result is delayed decisions, inconsistent outcomes, missed high-risk cases, and wasted resources on low-risk activities.

Solution

H2O.ai's Driverless AI builds transparent, explainable risk scoring models that evaluate hundreds of factors simultaneously to produce consistent, defensible risk scores for any decision domain. The platform automatically engineers features from raw data, selects optimal algorithms, and provides full model explainability — showing exactly which factors drove each score. Enterprise h2oGPTe generates natural language explanations of risk scores for decision-makers, while agentic AI can autonomously process incoming cases, score them, route low-risk cases for expedited processing, and escalate high-risk cases with supporting evidence for human review.

H2O.ai Platform Components Used

- Driverless AI — Classification, regression, and ranking models with full explainability (Shapley values, reason codes)
- Enterprise h2oGPTe — Natural language risk explanations; decision support narratives
- H2O Vertical Agents — Automated case intake, scoring, routing, and escalation

- H2O MLOps — Model fairness monitoring, bias detection, performance tracking, regulatory compliance

Benefits & Value

- Process cases 10–100x faster than manual review with consistent, defensible scoring
- Improve detection of high-risk cases by 30–60% through multi-factor ML models
- Reduce decision inconsistency and human bias with transparent, auditable scoring criteria
- Provide full explainability for every score — meeting OMB M-24-10 AI transparency requirements
- Enable risk-based resource allocation — focus human expertise on highest-impact cases
- Monitor for model fairness and bias across protected classes with automated alerts
- Accelerate low-risk case processing (auto-approve) while strengthening high-risk review

Applicable Agencies

Civilian: VA (benefits claim prioritization), USCIS (application risk scoring), FEMA (disaster resource allocation), SEC (market surveillance), FDA (inspection targeting), SBA (loan risk). DoD: Personnel security/clearance adjudication, insider threat scoring, force protection threat assessment, acquisition risk. IC: Threat prioritization, source reliability scoring, collection priority optimization, counterintelligence risk.

Use Case 6: Automated Reporting, Briefing Generation & Workflow Acceleration

Problem

Federal employees spend an estimated 30–40% of their time on administrative tasks — writing reports, preparing briefings, summarizing meetings, drafting correspondence, compiling status updates, and formatting data for leadership consumption. These tasks are repetitive, time-consuming, and pull skilled professionals away from substantive mission work. Reporting requirements multiply across oversight bodies (Congress, OMB, IGs, GAO), each with different formats and deadlines. Agencies struggle to maintain consistent quality and timeliness across hundreds of recurring reports while also responding to ad-hoc data calls and congressional inquiries on tight timelines.

Solution

H2O.ai's agentic AI platform automates the end-to-end reporting workflow — from data collection and analysis through draft generation, formatting, and quality review. Agents autonomously pull data from agency systems, perform required calculations and trend analysis, generate narrative text in the required format and style, populate templates, and produce review-ready drafts for human approval. Enterprise h2oGPTe handles summarization, translation, and natural language generation, while Driverless AI provides the underlying analytics (trends, forecasts, anomalies) that populate the reports with data-driven insights.

H2O.ai Platform Components Used

- Enterprise h2oGPTe — Report drafting, summarization, translation, correspondence generation, meeting notes
- H2O Vertical Agents — End-to-end workflow automation (data pull → analysis → draft → format → review routing)
- Driverless AI — Trend analysis, forecasting, and anomaly detection for data-driven report content
- H2O Wave — Custom dashboards and self-service reporting applications for agency users

Benefits & Value

- Reduce report preparation time by 60–80% — from days of manual compilation to hours of review
- Free 30–40% of staff time currently spent on administrative tasks for substantive mission work
- Ensure consistent quality, formatting, and compliance across all recurring reports
- Respond to ad-hoc data calls and congressional inquiries in hours instead of days
- Generate executive briefings automatically from underlying data with current metrics
- Support multilingual reporting for agencies with international operations
- Maintain human-in-the-loop for final approval while automating all preparation steps

Applicable Agencies

Civilian: Every agency (congressional reporting, OMB data calls, IG responses, FOIA), State Dept (diplomatic reporting/cables), HHS (public health surveillance reports). DoD: All services (readiness reporting, SITREPS, OPODs, after-action reports), Joint Staff (J-staff briefings), combatant commands (theater assessments). IC: Intelligence briefings, PDBs, current intelligence summaries, analytic product generation, FOIA/declassification review.

Use Case 7: SIEM Cost Optimization & Security Data Triage

Problem

Federal agencies generate massive volumes of telemetry from firewalls, endpoints, identity systems, cloud platforms, applications, OT/IoT devices, and network sensors. Much of this data is forwarded into SIEM platforms at full fidelity, regardless of whether it is high value for threat detection, compliance, or incident response. The result is escalating SIEM ingestion and retention cost, analyst overload, low signal-to-noise ratio, and delayed investigation of the events that matter most.

Agencies also face a second-order problem: data reduction efforts are often implemented with coarse filters that lower cost but can unintentionally degrade detection quality, remove needed context, or create friction with SOC, CISO, audit, and mission stakeholders. What agencies need is not blunt-force log suppression, but governed telemetry optimization that preserves security value while materially reducing downstream SIEM storage and processing burden.

Solution

Cribl serves as the telemetry control plane in front of the SIEM, ingesting logs and telemetry from distributed sources, routing data intelligently, reducing duplicate and low-value records, shaping formats, masking sensitive fields where needed, and sending the right data to the right destination at the right fidelity. H2O.ai adds predictive, generative, and agentic intelligence on top of that pipeline to classify event value, recommend routing and retention policies, identify high-cost/low-value data patterns, summarize telemetry trends, and continuously optimize data flows based on detection efficacy, analyst behavior, and mission priorities.

Together, Cribl and H2O.ai enable agencies to move from static SIEM filtering to adaptive security data optimization. Cribl handles collection, reduction, replay, routing, and enrichment across the telemetry estate, while H2O.ai analyzes usage patterns and security outcomes to determine which logs should be retained hot in

the SIEM, tiered to lower-cost storage, sampled, enriched, or excluded. This creates a governed framework for reducing SIEM cost without sacrificing forensic depth, compliance requirements, or operational visibility.

Joint Architecture Concept

- Cribl ingests telemetry from endpoints, identity providers, cloud services, applications, network tools, and existing log collectors.
- Cribl normalizes, deduplicates, routes, masks, and replays data across SIEM, data lake, archive, and observability destinations.
- H2O.ai Driverless AI scores log sources, event types, and routing patterns for detection value, investigation value, compliance value, and cost impact.
- Enterprise h2oGPTe provides natural-language analysis of telemetry growth, SIEM spend drivers, noisy alert sources, and policy options using retrieval over SOC runbooks, architecture records, and detection content.
- H2O Vertical Agents automate optimization workflows such as weekly tuning recommendations, retention-policy review, exception handling, and analyst-ready reports for SOC and CISO leadership.
- H2O MLOps governs model monitoring, drift detection, approval workflows, and auditable change control for routing and optimization policies.
- H2O.ai and Cribl Components Used
- Cribl Stream / edge telemetry pipeline capabilities — collection, routing, reduction, enrichment, replay, and destination-aware delivery.
- H2O Driverless AI — event-value scoring, anomaly detection, forecast models for ingestion growth, and cost-to-value optimization models.
- Enterprise h2oGPTe — telemetry knowledge assistant, policy analysis, report drafting, and investigation-support narratives.
- H2O Vertical Agents — agentic workflows for policy tuning, exception triage, and optimization recommendations with human approval.
- H2O MLOps — governance, explainability, drift monitoring, and model lifecycle control for production optimization decisions.

Benefits & Value

- Reduce SIEM ingestion and retention cost by filtering, tiering, sampling, and rerouting low-value data before it reaches the most expensive analytics tier.
- Improve analyst productivity by increasing signal-to-noise ratio and reducing duplicate, verbose, or low-action telemetry.
- Preserve forensic and compliance needs by routing full-fidelity data to lower-cost retention tiers while forwarding only the highest-value subset to the SIEM.
- Identify high-cost log sources and event classes that contribute little to detections, investigations, or reporting outcomes.
- Forecast telemetry growth and model the budget impact of onboarding new data sources before they are enabled.
- Support Zero Trust and SOC modernization by aligning telemetry handling to risk, mission priority, and threat-informed detection engineering.

- Maintain human-in-the-loop governance so SOC and CISO teams approve optimization policies before production rollout.

Applicable Agencies

Civilian: DHS component SOCs, Treasury, IRS, HHS, VA, DOJ, and any agency with high-volume hybrid-cloud security logging. DoD: defense enterprise SOCs, base and command cyber operations centers, and platform programs with large sensor estates. IC: mission systems that require selective forwarding, controlled retention, and rapid replay of high-value telemetry.

Proposal Narrative: How H2O.ai and Cribl Come Together

Cribl solves the transport, shaping, routing, and reduction problem for machine data, which is the operational foundation for SIEM cost control. H2O.ai solves the intelligence and optimization problem by learning which data actually contributes to detections, investigations, compliance, and mission decisions, then turning that insight into governed recommendations and automation.

In a joint deployment, Cribl sits in the telemetry path and gives the agency immediate control over what data goes where, in what format, and at what volume. H2O.ai then analyzes historical ingestion patterns, alert outcomes, analyst case activity, detection content performance, and retention obligations to identify the most cost-effective telemetry strategy for each source and event class.

That combination is especially strong for agencies trying to reduce SIEM spend without weakening security posture. Instead of relying on static rules alone, the agency can use explainable machine learning to score data value, generative AI to summarize trends and decision options for leadership, and agentic workflows to operationalize reviews and policy updates with human approval.

Use Case Summary Matrix

Use Case	AI Type	Primary Value	Time to First Value	Applicable Across
1. Intelligent Document Processing & Compliance Review	Agentic + Generative + Predictive	Automate manual document review; eliminate penalties; catch errors	30–60 days	Civilian, DoD, IC, A&D, and State and Local, A&D, and State and Local
2. Predictive Maintenance & Supply Chain Optimization	Predictive + Agentic	Prevent equipment failures; optimize inventory; reduce downtime	60–90 days	Civilian, DoD, IC, A&D, and State and Local, A&D, and State and Local
3. Fraud, Waste, and Abuse Detection	Predictive + Agentic + Generative	Detect fraud at scale; prioritize investigations; recover funds	60–90 days	Civilian, DoD, IC, A&D, and State and Local
4. Knowledge Management & Institutional Intelligence	Generative (RAG)	Instant answers from agency knowledge; preserve expertise; accelerate onboarding	30–45 days	Civilian, DoD, IC, A&D, and State and Local

Use Case	AI Type	Primary Value	Time to First Value	Applicable Across
5. Risk Scoring & Decision Support	Predictive + Agentic	Consistent, explainable scoring; faster decisions; better resource allocation	45–90 days	Civilian, DoD, IC, A&D, and State and Local
6. Automated Reporting & Workflow Acceleration	Agentic + Generative + Predictive	Automate admin tasks; free staff for mission; faster reporting	30–60 days	Civilian, DoD, IC, A&D, and State and Local
7. SIEM Cost Optimization & Security Data Triage	Predictive + Generative + Agentic	Reduce SIEM cost; improve telemetry quality; preserve detection and forensic value	30–60 days	Civilian, DoD, IC, A&D, and State and Local

Use Case Alignment to Package Tiers

Use Case	Small (\$350K)	Medium (\$700K)	Large (\$1.4M)
1. Document Processing	1 document type/workflow	2–3 document types + agents	Enterprise-wide, all document types, custom agents
2. Predictive Maintenance	1 equipment class	2–3 equipment classes + supply chain	Fleet-wide + full supply chain optimization
3. Fraud Detection	1 program/transaction type	2–3 programs + real-time monitoring	Enterprise-wide, all programs, network detection
4. Knowledge Management	Single division/repository	Multi-division, 2–3 repositories	Agency-wide, all repositories, fine-tuned models
5. Risk Scoring	1 decision domain	2–3 decision domains	Enterprise-wide, multiple domains, auto-routing
6. Reporting & Workflows	2–3 report types	5–8 report types + agents	All recurring reports, full workflow automation
7. SIEM Cost Optimization	3–5 priority log sources; one SIEM; basic routing and value scoring	10–20 sources; multiple security tools; policy recommendations and agentic tuning workflows	Enterprise-wide telemetry optimization; multi-SIEM/data-lake architecture; cross-domain policy automation

Bundled Solution Packages (Software + Services)

Each package below is a single Time and Material procurement combining the H2O.ai software license with a fully defined 12-month enablement Statement of Work. The SOW includes embedded Forward Deployed Engineer (FDE) services, defined milestones, measurable deliverables, acceptance criteria, and a structured knowledge transfer program. All packages are available for both On-Premises and FedRAMP High Cloud deployment at identical pricing.

The contracting officer may evaluate each package as a complete, self-contained deliverable with clear performance milestones and measurable outcomes tied to each payment milestone.

Package 1: Small — \$350,000 (12 Months)

Package Summary

Total Package Price	\$350,000
Software License (H2O.ai)	\$250,000
FDE & Enablement Services	\$100,000
Period of Performance	12 months from contract award
FDE Level	Senior FDE (4–8 years experience)
FDE FTE Allocation	0.25 FTE (~440 hours over 12 months)
Use Cases Delivered to Production	2–3
Deployment Option	On-Premises OR FedRAMP High Cloud

Software License Included — \$250,000

- Enterprise h2oGPTe — Full generative AI/RAG platform (50+ languages, 80+ file types, multi-model orchestration, guardrails, PII detection)
- Driverless AI — AutoML (classification, regression, time-series, anomaly detection, NLP, full explainability)
- H2O Document AI — OCR, document extraction, intelligent document processing
- Unlimited users, unlimited queries, no token fees
- Standard support and maintenance (business hours, SLA response)
- Base GPU memory allocation
- 12-month subscription term

FDE & Enablement Services — \$100,000

Staffing

Role	Level	FTE	Hours	Rate Range
Forward Deployed Engineer	Senior (4–8 yrs)	0.25 FTE	440 hrs	\$255–\$310/hr
Solutions Architect (Pre-Sales/Transition)	Senior	As needed	~40 hrs	Included

Milestones, Deliverables, and Hours Allocation

Milestone	Timeline	Hours	Deliverables	Acceptance Criteria	Payment
M1: Platform Deployment	Weeks 1–4 (Month 1)	80 hrs	- Production environment operational - Data source(s) connected - User accounts provisioned	Platform accessible by agency users; data pipeline validated; architecture doc accepted by agency technical lead	25% (\$25,000)

H2O.ai Zero-Dollar Token AI Unsolicited Proposal

Milestone	Timeline	Hours	Deliverables	Acceptance Criteria	Payment
			• Architecture decision record • Initial ATO artifacts		
M2: Use Cases in Production	Months 2–5	180 hrs	• 2–3 AI use cases deployed to production • Model performance baselines documented • Integration with agency systems complete • User acceptance testing passed	Each use case processing live agency data; model accuracy meets agreed thresholds; agency SMEs sign off on outputs	35% (\$35,000)
M3: Knowledge Transfer & Enablement	Months 5–9	100 hrs	• 4 quarterly KT sessions delivered • Agency staff trained on platform operations • Production runbooks delivered • Model cards and governance documentation	Agency staff can independently operate platform for trained use cases; runbooks reviewed and accepted	25% (\$25,000)
M4: Sustainment & Transition	Months 9–12	80 hrs	• Agency-led operations validated • Expansion roadmap delivered • Value realization report • Renewal recommendation • Quarterly executive brief	Agency operating independently with FDE in advisory role; executive brief delivered and accepted	15% (\$15,000)
TOTAL	12 months	440 hrs			100% (\$100,000)

Package 2: Medium — \$700,000 (12 Months)

Package Summary

Total Package Price	\$700,000
Software License (H2O.ai)	\$500,000
FDE & Enablement Services	\$200,000
Period of Performance	12 months from contract award
FDE Level	Senior FDE (4–8 years experience)
FDE FTE Allocation	0.5 FTE (~880 hours over 12 months)
Use Cases Delivered to Production	4–6
Deployment Option	On-Premises OR FedRAMP High Cloud

Software License Included — \$500,000

- All Small tier modules (Enterprise h2oGPTe + Driverless AI + Document AI) with enhanced GPU capacity
- H2O Vertical Agents — Agentic AI with autonomous multi-step reasoning, tool calling, function execution, MCP integration, human-in-the-loop
- H2O MLOps — Full model lifecycle governance (versioning, monitoring, drift detection, audit logging, deployment pipelines)
- Enterprise LLM Studio — Custom model development (fine-tuning, RLHF, model evaluation, prompt engineering workbench)
- Unlimited users, unlimited queries, no token fees
- Premium support and maintenance (24x7, priority SLA, dedicated support engineer)
- Enhanced GPU memory allocation
- 12-month subscription term

FDE & Enablement Services — \$200,000**Staffing**

Role	Level	FTE	Hours	Rate Range
Forward Deployed Engineer	Senior (4–8 yrs)	0.5 FTE	880 hrs	\$255–\$310/hr
Solutions Architect (Pre-Sales/Transition)	Senior	As needed	~60 hrs	Included
Data/ML Engineer (Supporting)	Mid-level	As needed	~80 hrs	Included

Milestones, Deliverables, and Hours Allocation

Milestone	Timeline	Hours	Deliverables	Acceptance Criteria	Payment
M1: Platform Deployment & Configuration	Weeks 1–3 (Month 1)	120 hrs	• Production environment operational • Multiple data sources connected • Agentic AI framework configured • User accounts & RBAC provisioned • Architecture decision record • ATO artifact package (initial)	Platform accessible; all data pipelines validated; agentic framework operational; architecture doc accepted	20% (\$40,000)
M2: Initial Use Cases in Production	Months 2–4	200 hrs	• 2–3 AI use cases deployed to production • 1 agentic workflow operational • Model performance baselines • Integration testing complete	Use cases processing live data; agent workflow executing autonomously with human oversight; accuracy thresholds met	25% (\$50,000)

H2O.ai Zero-Dollar Token AI Unsolicited Proposal

Milestone	Timeline	Hours	Deliverables	Acceptance Criteria	Payment
			• Adoption scorecard (initial)		
M3: Expanded Use Cases & Agents	Months 4–7	220 hrs	• Additional 2–3 use cases in production (total 4–6) • Additional agentic workflows • MLOps governance fully operational • Monthly adoption scorecards	All use cases in production; MLOps pipeline automated; monthly metrics reported	25% (\$50,000)
M4: Knowledge Transfer & Certification	Months 7–10	200 hrs	• Monthly KT sessions (4 sessions) • Agency staff hands-on training • H2O.ai platform certification for agency staff • Production runbooks • Model cards & governance docs • ATO continuous monitoring artifacts	Agency staff certified; can independently build and deploy new use cases; runbooks accepted; ATO artifacts complete	20% (\$40,000)
M5: Sustainment & Transition	Months 10–12	140 hrs	• Agency-led operations validated • Model refresh/retraining cycle executed • Expansion roadmap (next 12 months) • Value realization report • Quarterly executive briefs (3 delivered) • Renewal recommendation	Agency operating independently; model refresh completed successfully; executive brief accepted; roadmap approved	10% (\$20,000)
TOTAL	12 months	880 hrs			100% (\$200,000)

Package 3: Large — \$1,400,000 (12 Months)

Package Summary

Total Package Price	\$1,400,000
Software License (H2O.ai)	\$1,000,000
FDE & Enablement Services	\$400,000

H2O.ai Zero-Dollar Token AI Unsolicited Proposal

Period of Performance	12 months from contract award
FDE Level	Principal FDE (8+ years experience)
FDE FTE Allocation	1.0 FTE (~1,760 hours over 12 months)
Use Cases Delivered to Production	8–12
Deployment Option	On-Premises OR FedRAMP High Cloud

Software License Included — \$1,000,000

- All Medium tier modules (h2oGPTE + Driverless AI + Document AI + Vertical Agents + MLOps + LLM Studio) with maximum GPU capacity
- H2O Wave — AI application development layer for building and deploying custom AI-powered applications
- Custom Agent Development — Bespoke agentic AI agents tailored to agency-specific mission workflows
- Advanced Fine-Tuning & Model Customization — Domain-specific LLM fine-tuning, custom model training, specialized embeddings
- Unlimited users, unlimited queries, no token fees
- Enterprise support (24x7, 1-hour critical SLA, dedicated Customer Success Manager, quarterly business reviews)
- Maximum GPU memory allocation
- 12-month subscription term

FDE & Enablement Services — \$400,000

Staffing

Role	Level	FTE	Hours	Rate Range
Forward Deployed Engineer (Lead)	Principal (8+ yrs)	1.0 FTE	1,760 hrs	\$320–\$385/hr
Solutions Architect (Pre-Sales/Transition)	Senior	As needed	~80 hrs	Included
Data/ML Engineer (Supporting)	Mid-Senior	0.15 FTE	~260 hrs	Included
K3 Practice Lead (Oversight/QA)	Principal	As needed	~80 hrs	Included

Milestones, Deliverables, and Hours Allocation

Milestone	Timeline	Hours	Deliverables	Acceptance Criteria	Payment
M1: Rapid Platform Deployment	Weeks 1–2	160 hrs	• Production environment operational • All data sources connected • Full platform suite configured (predictive + generative + agentic) • RBAC & identity integration • Architecture decision record	Full platform operational; all integrations validated; ATO package accepted by ISSO	10% (\$40,000)

H2O.ai Zero-Dollar Token AI Unsolicited Proposal

Milestone	Timeline	Hours	Deliverables	Acceptance Criteria	Payment
			• Complete ATO artifact package		
M2: Initial Use Cases & Agent Development	Months 1–3	320 hrs	• 3–4 AI use cases in production • 2 custom agentic workflows operational • LLM fine-tuning for agency domain • Performance baselines established • Initial adoption scorecard	Use cases live on agency data; agents executing with human oversight; fine-tuned model outperforms baseline	20% (\$80,000)
M3: Scaled Production Deployment	Months 3–6	360 hrs	• Additional 3–4 use cases (total 6–8) • Additional custom agents deployed • AI application(s) built on H2O Wave • MLOps governance fully automated • Monthly adoption scorecards • Cross-division expansion initiated	All use cases in production; Wave app deployed to end users; MLOps automated; cross-division users onboarded	25% (\$100,000)
M4: Enterprise Expansion	Months 6–8	280 hrs	• Additional 2–4 use cases (total 8–12) • Enterprise-wide agent orchestration • Advanced model customization complete • Multi-division adoption validated • Model performance optimization	Full use-case portfolio in production; enterprise-wide adoption metrics met; optimization targets achieved	20% (\$80,000)
M5: Knowledge Transfer & Certification	Months 8–10	320 hrs	• Full certification program delivered • Agency staff trained on all platform modules • Complete production runbooks • Model cards for all production models • Governance documentation complete • ATO continuous	Agency staff independently building and deploying use cases; all documentation accepted; ATO monitoring active	15% (\$60,000)

Milestone	Timeline	Hours	Deliverables	Acceptance Criteria	Payment
			monitoring fully operational		
M6: Sustainment, Optimization & Transition	Months 10–12	320 hrs	• Agency-led operations validated • Model refresh cycles executed • Performance optimization complete • 12-month expansion roadmap • Value realization report (quantified ROI) • Quarterly executive briefs (4 delivered) • Renewal/expansion recommendation • Lessons learned document	Agency fully self-sufficient; all models refreshed; ROI documented; roadmap approved by agency leadership	10% (\$40,000)
TOTAL	12 months	1,760 hrs			100% (\$400,000)

12-Month Enablement Statement of Work

This section defines the common SOW framework applicable to all three packages. Specific milestone counts, hours, and deliverables scale by package tier as detailed in the preceding sections.

SOW Scope, Milestones, and Deliverables

The 12-month enablement SOW encompasses the following work areas:

- Platform Deployment & Configuration — Installation, configuration, data source integration, security hardening, identity/RBAC integration, and ATO artifact preparation.
- AI Use-Case Development & Production Delivery — Requirements gathering, data validation, model development, training, tuning, validation, integration testing, and production deployment.
- Agentic AI Workflow Development — Design, build, test, and deploy autonomous agent workflows with tool calling, function execution, and human-in-the-loop controls.
- Model Governance & Compliance — Model cards, versioning, audit logging, bias detection, drift monitoring, and alignment with OMB M-24-10 and NIST AI RMF.
- Knowledge Transfer & Capability Building — Structured training sessions, hands-on workshops, H2O.ai certification, production runbooks, and operational documentation.
- Adoption & Value Realization — Use-case backlog management, consumption tracking, adoption scorecards, executive value briefs, and ROI documentation.

- Sustainment & Transition — Agency-led operations validation, model refresh cycles, expansion roadmap, and renewal planning.

Deliverable Summary by Package

Deliverable	Small (\$350K)	Medium (\$700K)	Large (\$1.4M)
Production AI environment	1	1	1
Architecture decision record	1	1	1
AI use cases in production	2–3	4–6	8–12
Agentic workflows deployed	—	1–2	4–6
Custom AI applications (Wave)	—	—	1–2
Fine-tuned domain models	—	1	2–3
Knowledge transfer sessions	4 (quarterly)	8+ (monthly)	12+ (monthly + workshops)
Agency staff certifications	Basic	Full platform	Full platform + advanced
Production runbooks	1 per use case	1 per use case	1 per use case + master runbook
Executive value briefs	2 (semi-annual)	3 (quarterly)	4 (quarterly)
Value realization report	1 (end of term)	1 (end of term)	1 (end of term, quantified ROI)

Hours and FTE Summary

Package	FDE Level	FDE FTE	FDE Hours	Supporting Staff Hours	Total Services Hours
Small (\$350K)	Senior (4–8 yrs)	0.25 FTE	440 hrs	~40 hrs	~480 hrs
Medium (\$700K)	Senior (4–8 yrs)	0.5 FTE	880 hrs	~140 hrs	~1,020 hrs
Large (\$1.4M)	Principal (8+ yrs)	1.0 FTE	1,760 hrs	~420 hrs	~2,180 hrs

FDE Role, Responsibilities, and Qualifications

The H2O.ai Forward Deployed Engineer (FDE) is a senior AI engineer and client-facing technical expert who embeds directly inside federal agency environments to design, deploy, integrate, tune, and sustain AI and machine learning solutions in production. Unlike traditional ML engineers who build and hand off, the FDE owns the full lifecycle from design through adoption and value realization.

FDE vs. Traditional AI/ML Engineer

Dimension	Traditional ML Engineer	Forward Deployed Engineer
Primary focus	Internal model development and research	Customer AI adoption and production outcomes
Work location	Company lab or remote	Inside customer environment
Measure of success	Model accuracy metrics	AI use cases running in production
Collaboration model	Data science team	Customer engineering + agency mission teams

Dimension	Traditional ML Engineer	Forward Deployed Engineer
Scope	One model across many users	Many AI use cases for one customer
Lifecycle ownership	Experiment to model card	Design through adoption and value realization

Core Responsibilities

- **AI Solution Design & Architecture** — Design end-to-end AI/ML solutions using H2O.ai platform; architect agentic AI systems, RAG pipelines, and fine-tuning workflows tailored to federal mission needs; define data architecture and inference infrastructure for air-gapped, FedRAMP High, and IL4/IL5 environments.
- **Platform Deployment & Integration** — Install, configure, and integrate H2O.ai platform with agency data sources, identity systems, and infrastructure; validate data pipelines and model serving endpoints.
- **Use-Case Development & Production Delivery** — Develop, train, tune, validate, and deploy AI models and agentic workflows into production; conduct integration testing and user acceptance testing.
- **Production Support & Troubleshooting** — Root cause analysis; rollback procedures; triage issues spanning platform, infrastructure, data pipeline, and network/security controls.
- **Governance & Compliance** — Build model cards, data lineage documentation, and operational procedures; implement AI governance controls (versioning, audit logging, human-in-the-loop, bias detection); support ATO documentation and POA&M evidence.
- **Knowledge Transfer & Capability Building** — Conduct structured training and certification; develop production runbooks; mentor agency data scientists and IT staff to ensure independent sustainment.
- **Adoption & Value Realization** — Maintain prioritized use-case backlog; track platform consumption and adoption metrics; translate technical progress into executive-ready mission value briefings.

Qualifications

Role Level	Experience	Bill Rate Range	Key Qualifications
FDE — Associate	2–4 years	\$195–\$240/hr	H2O.ai ML certification; Python proficiency; supervised by Senior FDE
FDE — Senior	4–8 years	\$255–\$310/hr	Driverless AI practitioner cert; independent engagement leadership; DoD clearance eligible
FDE — Principal	8+ years	\$320–\$385/hr	Multi-agency AI delivery; agentic AI architecture depth; practice leadership; mentoring capability

Note: Bill rates are labor-only. Travel, clearance-related access costs, GPU infrastructure (for on-premises), and other direct costs are outside the base rate unless separately authorized.

Engagement Operating Rhythm

Activity	Cadence	Small	Medium / Large
Embedded technical work (model dev, integration, tuning)	Daily	✓	✓
Agency stand-up participation	Daily / Weekly	Weekly	Daily
AI use-case backlog review with technical and mission leads	Biweekly	✓	✓
Model performance and platform health review	Biweekly	—	✓
Adoption and consumption scorecard update	Monthly	Quarterly	Monthly
AI governance and ATO artifact updates	Monthly / per sprint	✓	✓
Knowledge transfer and capability-building sessions	Monthly / Quarterly	Quarterly	Monthly
Executive value brief preparation and delivery	Quarterly	✓	✓

Acceptance Criteria

The following general acceptance criteria apply to all milestones across all packages. Specific criteria for each milestone are defined in the package-level milestone tables above.

- **Platform Deployment:** H2O.ai platform is accessible by authorized agency users; all configured data sources return valid data; platform passes agency security scan.
- **Use Case Delivery:** Each use case processes live agency data (or agency-approved test data); model accuracy meets mutually agreed thresholds documented in the use-case backlog; agency subject-matter experts validate outputs and sign off.
- **Agentic Workflows:** Agent executes end-to-end workflow autonomously with human-in-the-loop checkpoints functioning correctly; agent outputs are auditable with full tool-invocation logging.
- **Knowledge Transfer:** Agency staff demonstrate independent ability to perform trained tasks (e.g., deploy a new model, modify an agent workflow, generate a report); certification exam passed where applicable.
- **Documentation:** All deliverable documents (runbooks, model cards, architecture records, ATO artifacts) are reviewed and accepted by designated agency technical lead within 10 business days of submission.
- **Executive Briefs:** Delivered on schedule; contain quantified mission value metrics; accepted by agency program manager.
- **Transition:** Agency operates platform independently for minimum 30 consecutive days with FDE in advisory-only role before final milestone acceptance

Security and Compliance

H2O.ai is FedRAMP High Authorized and holds SOC 2 Type 2, with FIPS-validated cryptography and support for fully on-premises and air-gapped deployment. The platform runs entirely within your agency's authorization boundary — no data leaves your control, and H2O.ai does not train models on customer data.

- FedRAMP High Authorized (not merely "FedRAMP Ready")
- SOC 2 Type 2 certified
- FIPS 140-2 validated cryptography
- Supports DoD SRG IL4/IL5 environments
- NIST 800-53 Rev 5 control mapping for AI components
- NIST AI Risk Management Framework (AI RMF) alignment
- OMB M-24-10 compliance support for AI governance
- Role-based access control (RBAC) with agency identity integration
- Full request/response and tool-invocation audit trails
- Configurable guardrails and PII-detection stack
- All models U.S.-developed; no models of foreign-adversary origin
- Zero Trust architecture principles applied to AI inference endpoints
- CUI, FOUO, PII, PHI data handling compliance

Company Backgrounds

Solution Provider: H2O.ai

H2O.ai is an enterprise AI company headquartered in Mountain View, California and founded by its CEO, Sri Ambati. Its mission is to democratize artificial intelligence — making AI accessible, responsible, and accountable for every organization. H2O.ai delivers to Federal customers both directly and through Carahsoft as its primary Federal distribution partner.

Built on widely adopted open-source foundations, H2O.ai has grown into a complete enterprise AI suite spanning generative and agentic AI with multi-agent orchestration (Enterprise h2oGPTe), document intelligence (H2O Document AI), custom model development (Enterprise LLM Studio), predictive machine learning (Driverless AI), and model lifecycle and governance (H2O MLOps). The platform unifies these capabilities under a common identity, access-control, and audit layer.

H2O.ai is FedRAMP High Authorized and holds SOC 2 Type 2, supports FIPS-validated cryptography, and offers fully on-premises and air-gapped deployment for the most sensitive workloads. The company holds the GSA Multiple Award Schedule (47QSWA18D008F) and NASA SEWP V (NNG15SC03B / NNG15SC27B) contract vehicles and serves customers across the Department of Defense, the Intelligence Community, Federal civilian agencies, and regulated commercial industries.

Distributor and contract vehicle leader: Carahsoft Technology Corp.

Carahsoft Technology Corp. is The Trusted Government IT Solutions Provider®. As the Master Government Aggregator® and Distributor for our vendor partners, we deliver solutions for Cybersecurity, MultiCloud,

DevSecOps, Big Data, Artificial Intelligence, Open Source, Customer Experience and more. Working with resellers, systems integrators, MSPs and consultants, our sales and marketing teams provide industry leading IT products, services, and training through hundreds of contracts.

Founded in 2004, Carahsoft is headquartered in Reston, Virginia and employs more than 3,600 professionals dedicated to serving our Public Sector and Healthcare customers and partners. Carahsoft has demonstrated impressive growth year after year, with annual revenue growing from \$3.4 million in 2004 to over \$18.9 billion in 2024.

Implementation: K3 Clouds (H2O.ai and Cribl)

K3 Clouds delivers reliable IT services to support mission and business objectives, including cloud solutions, artificial intelligence integration, and server and desktop management. Our cleared personnel are equipped to address complex IT challenges, including those in secure environments. We have a proven track record with select software partners for whom we have developed consulting practices.

A K3 H2O.ai **Forward Deployed Engineer (FDE)** works directly with customers to implement, customize, and optimize complex technical systems. FDEs connect technology with business objectives. We integrate with your teams, support operations, and co-develop solutions where they deliver the most value. Accelerating time to value and achieving mission outcomes are central to every FDE engagement.

K3 Clouds is a small business with strategic relationships with larger systems integrators, which we leverage for larger projects as needed.

Reseller: Cygnus

As a value added reseller, the Cygnus team brings deep experience advising Public Sector agencies on all technology aspects of their respective Cloud Adoption journeys and what that means in helping them execute against their Mission Objectives, more efficiently and cost effectively.

We want to earn your trust, while giving you the assurance that you are working with a technology partner focused on meeting your business needs, while driving operational excellence as you embrace the Cloud.

As your government cloud software sales partner, Cygnus is diligently adding to our portfolio of technology solutions that span the functional areas of Cloud Automation, Digital Transformation, Machine Learning & AI, while keeping in lock-step with current trends and best-of-breed solutions across the Industry.

Ordering Guide

H2O.ai solutions and services are available through multiple Governmentwide Acquisition Contracts, including Carahsoft's GSA Multiple Award Schedule and NASA SEWP V Contracts.

- GSA Multiple Award Schedule (MAS) — Contract 47QSWA18D008F
- NASA SEWP V — Contracts NNG15SC03B / NNG15SC27B
- Agency-specific vehicles available through Carahsoft's authorized reseller network

H2O.ai Zero-Dollar Token AI Unsolicited Proposal

Suggested period of performance: 12-month base period with option years. This proposal and its pricing remain valid for 90 days from the date of the cover letter.

Cygnus is an authorized reseller, K3 Clouds is the certified implementation partner for H2O.ai and is supporting this initiative alongside Carahsoft, providing additional acquisition flexibility and hands-on deployment support. H2O.ai is a strategic artificial intelligence software partner.

Please address all questions:

Please feel free to contact us directly at 703-431-1224 / jason.richards@carahsoft.com 571-719-8256 / kent.warren@k3clouds.com / 571-533-7726 / jeffery.phelan@h2o.ai 804-513-1779 doug.cornell@cygtech.us/ with any questions or communications about our proposed solution.

Thank you for your consideration,

Jeff Phelan, Jason Richards, Kent Warren & Doug Cornell
H2O.ai | H2O Team Lead Carahsoft | GM K3 Clouds, LLC. | CRO Cygnus

In Summary

Carahsoft and H2O.ai appreciate the opportunity to offer this solution to your agency.

This proposal provides three complete, bundled packages — each a single Time and Material procurement combining H2O.ai software licensing with a fully defined 12-month enablement SOW. Each package includes defined milestones, measurable deliverables, acceptance criteria, FDE staffing with specific FTE and hours allocations, and a structured payment schedule tied to milestone completion.

Package	Total Price	SW License	Services	FDE FTE	Use Cases
Small	\$350,000	\$250,000	\$100,000	0.25 FTE Senior	2–3
Medium	\$700,000	\$500,000	\$200,000	0.5 FTE Senior	4–6
Large	\$1,400,000	\$1,000,000	\$400,000	1.0 FTE Principal	8–12

Carahsoft is confident that H2O.ai's Enterprise AI Suite will allow your agency to:

- Deploy production AI (predictive, generative, and agentic) within 30–60 days of contract award;
- Deliver measurable mission value with 2–12 AI use cases in production within the first year;

H2O.ai Zero-Dollar Token AI Unsolicited Proposal

- Build lasting internal AI capability through embedded FDE support and structured enablement with defined milestones;
- Maintain full data sovereignty on a FedRAMP High Authorized, U.S.-sovereign platform; and
- Establish a governed AI foundation that scales enterprise-wide adoption in subsequent fiscal years.

Carahsoft understands the importance of speed and efficiency in end-of-year procurement and is confident your agency will benefit from this solution and our expertise. Carahsoft would appreciate the opportunity to discuss the benefits of this solution with you in further detail.