Performance Work Statement

For

Enterprise Data Engineering

Department of Defense
Chief Digital and Artificial Intelligence Office
Solicitation # FA8003-24-R-0017
30 September 2024

Table of Contents

1.0 Introduction

1.1 Background

Advana is the Department of Defense (DoD) Chief Digital and Artificial Intelligence Office's (CDAO) enterprise-wide, multi-domain data, analytics, and artificial intelligence (AI) platform that provides all DoD military and civilian decision makers, analysts, and builders with enterprise data, tools, and capabilities—all in a scalable, reliable, and secure environment. Jointly managed by the DoD CDAO Mission Analytics and Enterprise Platforms and Services Directorates, the mission of Advana is to accelerate decision advantage by making data, analytics, and AI widely accessible, understandable, and actionable across DoD. To do this, the Advana platform hosts curated analytics applications in a total of 13 DoD business domains, to include logistics, people and health, intelligence and security, readiness and global force management, etc.; it provides market-leading, self-service tools to perform data exploration, model development, data visualization, and other analysis activities; and it provides both users and builders on-demand data, analytics, and AI capabilities via a DoD-wide platform.

1.2 Objectives

The objectives for this procurement are as follows:

1.2.1. Obtain program management expertise that drives continuously improving program efficiencies and operational effectiveness in supporting approximately 400 automated data pipelines and approximately 30 AI/machine learning (ML) models.

1.2.2. Provide the DoD with state-of-the-art enterprise data engineering capabilities to amplify the impact of new and existing analytics applications used on Advana.

1.2.3. Prepare the DoD for data productization, by systematically implementing strong data governance and AI governance principles, processes, and practices within its centralized data, analytics, and AI platform, Advana.

2.0 Requirements

2.1 Program Management

The contractor shall:

2.1.1. Organize a kickoff meeting within two weeks of award to identify roles and responsibilities, review expectations, review the proposed software development plan, review the proposed data management plan, and accomplish the coordination necessary for performance. Submit agenda prior to the event, and minutes after the event (A001; A002; A003; A004).

2.1.2. Plan and implement an Agile software management process across Advana utilizing industry best practices; and review completed, in-progress, planned near-term

(e.g., increment planning), and planned long-term (e.g., strategic roadmap) technical activities at a level of detail and cadence as agreed upon with the CDAO (A003; A004).

2.1.3. Prepare and maintain a roadmap and backlog to identify, plan, coordinate, prioritize, and manage execution of all required technical activities (A005; A006).

2.1.4. Report the following (A007):

> 2.1.5.1. Number and overall summary of incidents involving spillage of classified information, personally identifiable information (PII), protected health information (PHI), and other restricted/controlled information.
>
> 2.1.5.2. Summary of template, guide, and playbook development under paragraph 2.2.2.
>
> 2.1.5.3. Summary of engineering support provided under paragraph 2.2.3.
>
> 2.1.5.4. Summary of training plan development and execution under paragraph 2.3.3.
>
> 2.1.5.5. Summary of training plan development under paragraph 2.3.11.

2.1.5. Support SSAE-18 and SOC 1 audit response between Advana management and external independent auditors. This includes quality reviews over all Provided by Client requests, policies and documentation, and preparation for auditor observations and meetings. Coordinate all phases of audit from planning, execution, and remediation (A001; A002; A007).

2.2 Data Engineering

The contractor shall:

2.2.1. Manage a team of specialized data scientists and data engineers focused on optimizing the creation, deployment, testing, and iterative development of all bronze, silver, and gold data tables deployed across unclassified and classified CDAO environments, in accordance with the References. Advana currently uses Python, SQL, and PySpark languages along with standard Python packages in Databricks. Bronze tables are query-able versions of data as they are received including historical copies. Silver tables are refined source data with a schema standardized against guidance from the data source. Gold tables store fit for purpose analytic data transformed, cleansed and aggregated for the use case or application (A008).

2.2.2. Create and maintain templates, guides, and playbooks to inform and enable the DoD to reuse and scale data, analytic, and AI work for decision advantage (A009).

2.2.3. Provide engineering support to all customers working to leverage the Advana platform and other CDAO platforms to (A008):

> 2.2.3.1. Build data products and data mesh services.
>
> 2.2.3.2. Register, ingest, store, search, and retrieve data products.
>
> 2.2.3.3. Develop suites of data quality business logic.

2.2.3.4 Train, test, and deploy production AI/ML models.

2.2.3.5. Integrate and maintain metadata.

2.3 Data and AI Governance

The contractor shall:

2.3.1. Maintain existing and implement more mature data and AI governance processes that allow for finer grained access control and decentralized stewardship of DoD data and AI assets. Existing data and AI governance processes center around four key products, maintaining a metastore, data cataloging, data lineage, and data access control. The decentralized model will enable data stewards across the DoD to govern their data and AI assets promoting flexibility while enforcing minimum requirements.

2.3.2. Enforce asset model structure standards for all asset types available within the DoD Federated Data Catalog. These standards are implemented to maintain consistency with the Federated Data Catalog as the Federated Data Catalog receives metadata from across the DoD (A010).

2.3.3. Systematically develop and execute training plans that enable DoD data stewards to catalog and federate organizational data assets within the DoD Federated Data Catalog (A011).

2.3.4. Develop, maintain, and continually expand the DoD Federated Data Catalog across DoD unclassified and classified environments, while integrating it with other catalogs across the federal Government, e.g., General Services Administration (GSA), Department of Veterans Affairs (VA), etc. Per the Open Government Data Act, share metadata with GSA Catalog (data.gov) (A010).

2.3.5. Identify how to promulgate the DoD Federated Data Catalog as DoD's authoritative metadata interface at all classification levels, by establishing and maintaining standards for connecting to other public sector data and model catalogs, such as that hosted at data.gov (A007).

2.3.6. Ensure all data assets and AI/ML model assets are registered within the DoD Federated Data Catalog and Advana's AI/ML model registry; and kept up to date in an event driven fashion (A010).

2.3.7. Develop and deploy a systematic, automated metadata card and AI/ML model metadata card strategy that is consistent with prevailing data mesh principles; and aligned to DoD and CDAO policy identified in the References. Data mesh principles to adhere to include but are not limited to data as a product, domain-oriented decentralized data ownership and architecture, and self-service data infrastructure as a platform (A010).

2.3.8. Provide a DoD Federated Data Catalog with associated application program interfaces (API) and integration toolsets that lists complete metadata and access

information for DoD datasets and AI models, supporting reuse and sharing for future development (A010).

2.3.9. Make cataloged data cards and model cards discoverable to consumers, so that the DoD Federated Data Catalog provides a comprehensive view of the stewardship and use of data and model assets across the DoD (A010).

2.3.10. Define, implement, and maintain interoperable, policy-as-code data access control processes and capabilities that enable attribute-based access control for data available on the Advana platform's unclassified and classified environments; and which may be exported and shared to inform and enforce access to data in other enterprise platforms. With policy-as-code in place, data shared external to CDAO platforms can include code for policy enforcement (A010).

2.3.11. Develop and deliver training plans that enable federated stewardship over data access control, e.g., data tagging, policy creation, etc. (A011).

2.3.12. Develop, test, deploy, and maintain automated data quality and data lineage management capabilities, which enable tracing data lineage through transformations, and provide the ability to share and consume lineage from other DoD systems and platforms. Enable self-service data quality that allows users to define data quality rules specific to their data (A010).

2.3.13. Integrate all the above into the Advana system and other CDAO and DoD platforms (e.g. Advana Azure, SUNet).

2.4 Models as a Service

The contractor shall:

2.4.1. Create and support AI models for use cases in a specific Advana functional domain (e.g., financial management analytics), that may be inferenced for similar use cases within other DoD platforms via the existing Advana Enterprise API offering as follows (A012):

> 2.4.1.1. Models trained and deployed shall integrate, maintain, and share model specific metadata such as training dataset, model deployment date, model description, appropriate model usage and limitations.

> 2.4.1.2. Support the existing 10+ models in production. Perform retraining, performance monitoring, and updates as requested by domain use-case owners. Examples of the types of models currently in production include time series, anomaly detection, natural language parsing and information extraction, and text summarization via Large Language Model (LLM).

> 2.4.1.3. Collaborate closely with domain experts and stakeholders to understand business requirements for specific model implementations, gather, transform, and prepare data required for model training.

2.4.2. Implement state-of-the-art algorithms within the Advana platform and other CDAO platforms for relevant categories of AI/ML models (e.g. open-source foundational large language models, object detection, object classification, anomaly and outlier detection, financial and event time series analysis) (A012).

3.0 Performance Management

3.1 General

The contractor service requirements are summarized into performance objectives that relate directly to mission essential items. The performance threshold describes the minimum acceptable levels of service required for each performance objective. These thresholds are critical to mission success but do not represent all contract performance requirements. All contract performance requirements in this performance work statement will be monitored.

3.2 Services Summary

Metric #	Performance Objective	PWS Para	Performance Threshold	Surveillance Method
1	User Experience	2.2.3	No more than three (3) substantiated negative feedback from Advana users.	Customer Input
2	Training Product Quality	2.2.2, 2.3.3, 2.3.11	No more than two (2) late submissions of drafted training material corrections.	Periodic Inspection
3	Responsiveness	4.2	Respond to classified work requests within one (1) hour during established work hours.	100% Inspection

4.0 General Requirements

4.1 Period of Performance

The period of performance shall be 12 months from the date of contract award.

4.2 Place of Performance

The primary place of performance shall be the contractor's facilities. Cleared individuals supporting classified work shall be located within two hours of a sensitive compartmented information facility (SCIF).

4.3 Hours of Performance

The Contractor shall maintain work hours Monday through Friday 0800-1700 Eastern Time. In event of emergency situations, the government may temporarily require support outside of these work hours.

4.4 Government-Furnished Property

Up to 50 unclassified laptops may be provided, as mission circumstances dictate. No other Government-furnished Property (GFP) will be provided.

4.5 Government-Furnished Information

Government-furnished information (GFI) means information in the possession of, or directly acquired by, the government and subsequently furnished to the Contractor for performance of a contract. GFI also includes contractor-acquired information if the contractor-acquired information is a deliverable under the contract and is for continued use under the contract. Upon award of the contract, the government will provide the GFI listed under References.

4.6 Security

The contractor shall:

4.6.1. Possess an active Top Secret Facility Clearance Level within the National Industrial Security System (NISS) at the time of proposal submission and maintain an active one for the duration of the contract.

4.6.2. Comply with the security classification guidance specified in the contract DD Form 254 before performing work on any classified Government systems or classified material.

4.6.3. Perform all classified work at a SCIF.

4.6.4. Maintain at least a 30% workforce with adjudicated Top Secret Sensitive Compartmented Information (SCI) clearance for the duration of the contract.

4.7 Travel

4.7.1. Contractor travel may be required under this contract to support the following trips:

Trip Purpose	Projected Dates of Travel	Travel To	Number of Trip Days	Number of Travelers
Quarterly In-Person Working Session	12 weeks after contract award date	Washington, D.C.	5	Max of 15

	24 weeks after contract award date 36 weeks after contract award date 48 weeks after contract award date			
CCMD: INDOPACOM	8 weeks after contract award date	Honolulu, HI	5	Max of 4
DON IT West Conference	8 weeks after contract award date	San Diego, CA	3	Max of 3
NVIDIA Conference	12 weeks after contract award date	San Jose, CA	4	Max of 4
Collibra Data Citizens Conference	16 weeks after contract award date	Orlando, FL	3	Max of 4
CCMD: CENTCOM + SOCOM	16 weeks after contract award date	Tampa, FL	4	Max of 4
Microsoft Ignite Conference	16 weeks after contract award date	San Jose, CA	4	Max of 4
Tableau Conference	16 weeks after contract award date	San Diego, CA	3	Max of 2
CCMD: SOUTHCOM	20 weeks after contract award date	Miami, FL	4	Max of 3
Qlik Connect Conference	20 weeks after contract award date	Orlando, FL	3	Max of 2
Databricks Data + AI Summit	24 weeks after contract award date	San Francisco, CA	4	Max of 4

DON IT East Conference	24 weeks after contract award date	Norfolk, VA	3	Max of 3
CCMD: NORTHCOM + SPACECOM	32 weeks after contract award date	Colorado Springs, CO	4	Max of 3
CCMD: EUCOM + AFRICOM	36 weeks after contract award date	Stuttgart, DE	5	Max of 4
DoDIIS + STRATCOM	40 weeks after contract award date	Omaha, NE	5	Max of 4
NDTA + TRANSCOM	40 weeks after contract award date	St. Louis, MO	4	Max of 4
AWS Re:Invent Conference	44 weeks after contract award date	Las Vegas, NV	4	Max of 4

4.7.2. Upon the Government's notification to proceed with a scheduled trip, the contractor shall arrange all transportation and other travel-related requirements to ensure attendance by their appropriate personnel.

4.7.3. The contractor shall submit a report upon completing each travel event (A013).

4.8 Associate Contractor Agreements

4.8.1. "Associate Contractor Agreements" (ACA) are defined as agreements between contractors that are working on government contracts or projects that specify requirements for them to share information, data, technical knowledge, expertise, or resources. The Government anticipates the need for one or more ACAs to successfully perform all requirements of this contract. Prime contractor to subcontractor relationships do not constitute ACAs and are not subject ACA requirements, as defined herein.

4.8.2. Before executing an ACA, the contractor shall provide a final draft agreement to the CO that identifies the following information:

    4.8.2.1. The associated contractors and their relationships.

    4.8.2.2. The program(s) involved and the relevant Government contracts of the associate contractors.

    4.8.2.3. General description of associate contractor interfaces.

    4.8.2.4. Categories of information to be exchanged or support to be provided.

    4.8.2.5. Expiration date, or completion event, of the ACA.

4.8.2.6. Potential conflicts between relevant Government contracts and the ACA.

4.8.3. The contractor will not be relieved of any contract requirements or entitled to any adjustments to the contract terms because of a failure to resolve a disagreement with an associated contractor. Additionally, liability for the improper disclosure of any proprietary data contained in or referenced by any agreement rests with the parties to the agreement, and not the Government.

4.8.4. All costs associated with the agreements are considered included in the negotiated cost of this contract. Agreements may be amended only as required by the Government during the performance of this contract.

5.0 Data Management

5.1 Data Requirements

The Contractor shall implement data management procedures for preparation, quality control, administration, and delivery of data required by the following Contract Data Requirements Lists (CDRL):

Exhibit Line Item (ELIN)	Title
A001	Meeting Agenda
A002	Meeting Minutes
A003	Software Development Plan
A004	Data Management Plan
A005	Roadmap
A006	Backlog
A007	Status Report
A008	Computer Software Product
A009	Training Materials
A010	DoD Federated Data Catalog
A011	Training Materials
A012	Artificial Intelligence (AI) Model
A013	Trip/Travel Report

5.2 Data Rights

To enable maximum use, reuse, modification, and/or distribution throughout the Advana ecosystem, the Government shall have an unlimited rights license to all technical data and software documentation developed and acquired under this contract.

6.0 References

(a) "Designation of Enterprise Services Supporting Federated Data Cataloguing", 2 Dec 2020

(b) "Federated Data Catalog – Minimum Metadata Requirements", 1 Oct 2021
(c) "DoD Data Stewardship Guidebook", 10 Oct 2021
(d) DoD Federated Data Catalog
(e) Advana Data Acquisition Playbook
(f) Advana Data Operations Playbook
(g) Advana Data Catalog
(h) Advana Data Playbook
(i) Advana Machines Learning Operations Playbook
(j) Advana Builder Operations Playbook